UCSF

UC San Francisco Electronic Theses and Dissertations

Title

A second generation force field for the simulation of proteins and nucleic acids

Permalink

https://escholarship.org/uc/item/8b09h0w0

Author

Cornell, Wendy D.

Publication Date

1995

Peer reviewed|Thesis/dissertation

A SECOND GENERATION FORCE FIELD FOR THE SIMULATION OF PROTEINS AND NUCLEIC ACIDS

by

Wendy D. Cornell

B.S., Case Western Reserve University, 1985

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of
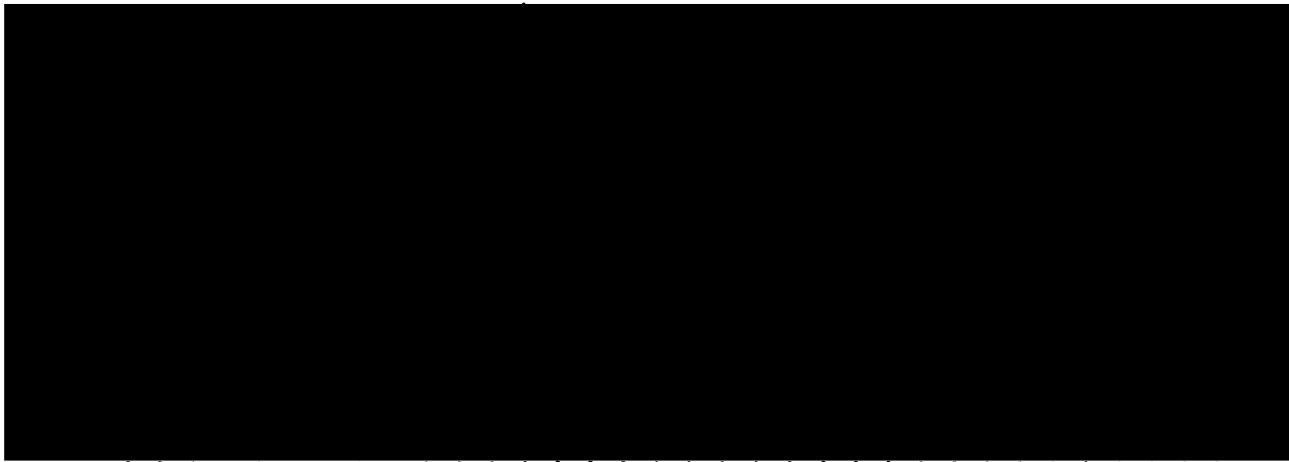
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

. . . .
Date   University Librarian

Degree Conferred: . . . .

Dedication

This thesis is dedicated to my parents, Harry Henderson and Marilyn Keslar Cornell, with appreciation for all of their love and support.

Preface

It is rather difficult to summarize a six year journey in a page or two, but I will make an attempt here. My introduction to UCSF began with meeting Tack Kuntz when he came to Kalamazoo, Michigan to give a seminar at Upjohn. Tack was the head of the Biophysics Program during most of my time at UCSF and I believe that he has built up a fine program of which the students and faculty are all quite proud. If Tack was the father of the Biophysics Program, then Julie Ransom was surely the mother, attending to our day to day needs and making sure that we were adequately housed, fed, and paid. I cannot imagine a better graduate program and I would like to thank everyone associated with the Graduate Group in Biophysics, including my classmates Jennifer Fung, Stephen Rader, and Jonathan Davis.

Eventually every student must leave the security of the first year womb and choose a research adviser and group to call one's own. I chose Peter Kollman's group for the opportunities that it offered to learn about fundamental theoretical chemistry and biochemistry and I was not disappointed. I feel particularly lucky to have been able to work with Peter on the development of the new second generation force field, given that his expertise in this area is world reknowned. Peter attracted many bright and accomplished people to work in his group and I benfitted from interacting with many of these co-workers as well. In particular, I learned much from Jim Caldwell, Ian Gould, David Ferguson, Yax Sun, and Alain St-Amant. Piotr Cieplak was my collaborator on the new force field and working with him was also a pleasure. I would also like to acknowledge the social interactions with group members, in particular the occupants of S-947 and my officemate Meg McCarrick.

Many other "pals" contributed to the UCSF "milieu." Good times were had with Angelika Muscate, Thomas Fox, Simon Friedman, Chris and Melanie Schafmeister, Eric Pettersen, Shauna Farr-Jones, Phyllis Kosen, Debbie Kallick, and many others. I would also like to thank the members of the UCSF Graphics Lab -- Eric P., Al Conte, and Greg Couch -- for their support.

Friends back home were always an important source of encouragement and comfort and I would like to thank Jill D'Angelo, Andrea Miller, Annette Meyer, Jamie Sparks, and Scott Rose for their support over the years. I would like to thank my family in particular -- my parents Harry and Marilyn Cornell, my brother and sister-in-law Keith and Amy Cornell, grandmother June Cornell, and cousins Nevin and Maryjane Voll. My dog Maggie was a fine companion until she decided that she had had enough of San Francisco and flew back home to PA.

Finally, I would like to thank the faculty members who played key roles in my education. Ken Dill was the chair of my orals committee with Dave Agard, Tom James, and Robert Fletterick as the remaining members and Ken and Dave both served as members of my thesis committee. The biggest and final "thanks," of course, goes to my adviser Peter, whose enthusiasm, insight, and generosity I benefitted from greatly, and whose support and encouragement enabled me to develop as a scientist.

A SECOND GENERATION FORCE FIELD FOR THE SIMULATION OF PROTEINS AND NUCLEIC ACIDS

Wendy D. Cornell

Dissertation Abstract

This thesis describes the development and testing of a second generation additive force field for the molecular mechanical simulation of proteins and nucleic acids. Chapter 1 provides an introductory overview of this work. In Chapter 2 the validation of the new RESP (restrained electrostatic potential-fit) model for calculating atom-centered charges is presented. Chapter 3 describes the results of high level *ab initio* calculations on the glycyl and alanyl dipeptides, comparing the results obtained using different theoretical models. In Chapter 4 the derivation of charges for the amino acids is presented. These charges are calculated using multiple molecules, multiple conformations, and restrained electrostatic potential fitting. Chapter 5 presents the remainder of the derivation and testing of the new force field. In Chapter 6 a non-additive model is applied to the calculation of the conformational energies of the glycyl and alanyl dipeptides. The underlying assumptions and approximations inherent in the new additive force field are reviewed in Chapter 7, along with the imperfections and directions for future work in the area. Finally, Appendix I presents the application of the first generation force field to the simulation of the conformational dynamics of cyclopentane.
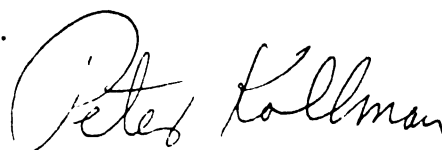
Table of Contents

List of Tables

Chapter 6:

List of Figures

Chapter 1

Introduction

The field of computer simulation occupies a distinct position in the world of science. It is different from both experiemental and theoretical science and ideally is able to explore in detail both experimental observations and theoretical predictions. Dependent as it is on computational resources, this field has grown over the last few decades in conjunction with developments in computer hardware. The application of computer simulation to atomic level systems of chemical and biochemical interest became fairly widespread during the 1980's, when a number of general protein and nucleic acid force fields became available. One of the most widely used of these classical force fields was developed at UCSF in this laboratory and has been used more in less in its original form for nearly ten years.

Although the Weiner *et al.* force field [1] has been quite successful, it was developed primarily for use without explicit solvent, with a distance dependent dielectric. This simplification was necessary in order to treat the relatively large and complex proteins and nucleic acids within the constraints of the avaliable computing power. More recently, advances in computer technology have allowed for the simulation of such biological molecules in explicit solvent, which is a more accurate representation of these systems. We therefore chose to undertake the development of a second generation force field, based on the general philosophy of Weiner *et al.*, which would be appropriate for calculations carried out in solution.

A new charge model has been developed for this second generation force field, and its validation is described in Chapter 2. The new model employs the 6-31G* quantum mechanical basis set, multiple conformations, restrained fitting t oa quantum mechanical electrostatic potential (esp), and a 1-4 electrostatic scale factor of 1/1.2. The 6-31G* basis set was chosen because it results in molecular dipole moments which are about 10-20% greater than the gas phase values and thus implicitly includes

the approximate amount of polarization which would be found in an aqueous environment. Multiple conformation fitting was first suggested by Reynolds *et al.* [2] and it directly addresses the problem of the conformational dependence of esp-fit charges. The use of restraints in the fitting process was developed by Christopher Bayly [3] and tested extensively by the author and Piotr Cieplak. [4] The RESP (restrained esp-fit) charge model employs hyperbolic restraints on the charges on non-hydrogen atoms, in order to attenuate the magnitudes of charges on buried atoms which are not well defined by the shells of potential points.

The RESP charges are more consistent with respect to the values calculated from different conformations and are less sensitive to me 1-4 electrostatic scale factor. Single conformation RESP charges are not as good as multiple conformation fit standard ESP charges at reproducing the electrostatic potential of different conformations of a molecule, but in tests on propylamine they are shown to provide an improvement beyond that achieved through multiple conformation fitting alone. Furthermore, they provide some of the benefit derived from multiple conformation fitting withouth the requirement of carrying out multiple *ab initio* calculations on multiple conformations of a molecule.

High level quantum mechanical calculations on glycyl and alanyl dipeptides are described in Chapter 3. These calculations follow up on work originally carried out by Ian Gould, who calculated relative conformational energies for four low energy conformations of alanyl dipeptide at the MP2/TZP//HF/6-31G* level. [5] Those calculations were followed by a similar set of calculations on glycyl dipeptide which were carried out by the author in collaboration with Ian Gould. [6] The results presented in Chapter 3 address the question of the effect of the particular theoretical treatment (basis set) employed in the calculations as well as the choice of the

molecular model, i.e. the use of methyl or hydrogen blocking groups on the two backbone amide groups. We chose to use the larger methyl-blocked analogs for our model systems, even though they made the *ab initio* calculations much more expensive. In order to obtain molecular mechanical residues of the correct net charge for inclusion in the database, a Lagrange constraint must be applied during the fit which forces the blocking groups plus outer amide atoms to be neutral. This simplification is less severe when the blocking groups consist of methyl groups rather than just hydrogen atoms. Quantum mechanical results on the methyl-blocked analogs can then be compared directly with the molecular mechanical model.

These *ab initio* calculations were motivated by the fact that the conformational energies calculated for glycyl dipeptide (using the same protocol employed by Gould and Kollman for alanyl dipeptide) differed from those determined by Head-Gordon *et al.* at the MP2 level of theory on hydrogen-blocked analogs. [7] The results presented here suggest that the different blocking groups (methyl vs. hydrogen) rather than the different basis sets were the main cause of the disagreement seen between the two sets of calculations. This implies that the use of diffuse functions is not necessarily called for to model hydrogen bonded systems when using a sufficienly large and well-balanced basis set.

In Chapter 4 the RESP model is applied to the derivation of charges for the amino acids. Analogous calculations were carried out by Piotr Cieplak for nucleic acids and reported in a common paper. [8] Chapter 4 represents a major revision of that paper with the nucleic acid results removed and some additional data provided on the amino acids. The final charge model for the amino acids employs two conformations for each amino acid, one with the backbone in an α-helical conformation and the other with it in an extended (β-sheet) conformation. Side chain χ orientations were then

assigned based on a PDB survey [9] so that each of the two conformations had different values for a given χ_n. A multiple molecule fit of gly, ala, val, ser, and asn was carried out in order to obtain a consensus set of charges for the backbone amide atoms. Different sets of consensus amide charges were calculated for the positively and negatively charged amino acids. The effects of constraining the two backbone amide groups to have the same charges and each blocking group to be neutral were explored and found to be not too severe. The alanyl and glycyl dipeptide molecular mechanical conformational energies calculated with even the unconstrained charge sets were found not to agree well with the quantum mechanical energies. In particular, relative conformational energies of the α_R conformations of glycyl and alanyl dipeptides were seen to differ by over 2 kcal/mol, when the quantum mechanical energies were essentially the same. The reason for this disparate behavior is currently not clear.

Chapter 5 describes the development of the new additive force fieold using small molecules and simple liquids. This chapter reflects work carried out by a number of other people in the group, including Ian Gould, Ken Merz, David Spellmeyer, David Ferguson, Christopher Bayly, David Veenstra, Thomas Fox, and Jim Caldwell. For the most part, however, the results presented therein reflect work carried out by the author, Piotr Cieplak, and Peter Kollman. The test cases are similar to those employed in the development of the first generation force field, based on the reproduction of geometries, conformational energies, interaction energies, and vibrational frequencies. For this force field, the calculation of accurate free energies of solvation is also critical, and so this data has been added to the test set. The major challenge posed by this new force field was the need to develop a charge model which performed well at reproducing both inter- and intra-molecular interactions. Charges derived using the 6-31G* basis set are larger than ones derived with the

STO-3G basis set and are very sensitive to the magnitude of the 1-4 interaction. The new 1-4 electrostatic scale factor of 1/1.2 is thus critical for maintaining the proper balance between 1-4 and 1-5 interactions and results in conformational energies which agree well with experiment for many of the small molecules studied. This was not the case with the alanyl and glycyl dipeptides, however, and in that case it was necessary to develop a set of ϕ and ψ dihedral parameters with V_1, V_2, V_3, and V_4 Fourier components.

A new non-additive model employing atom-centered polarizability is applied to the dipeptide conformational energies in Chapter 6. This model uses polarizabilities derived from optical spectroscopy [10] and has been developed by Jim Caldwell by fitting to the density and enthalpy of vaporization of neat liquid methanol and trans-NMA. The model requires scaling the 6-31G* charges by a factor of 0.88 The model is found to improve greatly the conformational energies calculated with no ϕ and ψ dihedral parameters (as compared to the additive model with no ϕ and ψ dihedral parameters), and can be improved even further through the addition of a very simple set of dihedral parameters which are small in magnitude.

In Chapter 7 the assumptions and approximations inherent in the force field are summarized and some of the limitations of both the additive and non-additive model are discussed. Although the non-additive model performs well as calculating the conformational energies of the dipeptides, it significantly underestimates the interaction energies of the DNA base pairs. Work is currently underway in the group by Richard Dixon investigating the improvement afforded by the inclusion of off-center charges (lone pairs) in the charge model. It is possible that the neglect of charge anisotropy about a given atoms is more severe than the neglect of polarization.

Finally, Appendix I describes a study of the dynamical properties of cyclopentane in the gas phase as modelled by our first generation force field. This work was carried out by the author, Yax Sun, and Maria Ha, a high school student who was a summer intern with the Science and Education Partnership (SEP) porgram. This project was motivated by a study carried out using the MM3 force field which was the first application of that force field which used the new molecular dynamics capability of the MM3 program. [11] Our somewhat surprising results showed that a simple harmonic diagonal force field [1] performed nearly as well as the more complex MM3 force field at reproducing the velocity and puckering of the pseudorotation process. Furthermore, our simple force field more accurately represented the barrier to planarity. Results such as these auger well for the potential of the new force field to be extended to application to small molecules.

In conclusion, the work presented in this thesis represents the development and testing of a new second generation force field for the simulation of proteins and nucleic acids. It is hoped that it will prove to be a useful model for carrying out new and interesting chemical and biochemical applications and that it will serve as a solid platform for further development of both additive and non-additive models.

References

1. (a) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. (b) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230.

2. Reynolds, C.A.; Essex, J.W.; Richards, W.G. *J. Amer. Chem. Soc.* **1992**, *114*, 9075.

3. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P.A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

4. Cornell, W.; Cieplak, P.; Bayly, C.; Kollman, P.A. *J. Amer. Chem. Soc.* **1993**, *115*, 9620.

5. Gould, I.R.; Kollman, P.A. *J. Phys. Chem.* **1992**, *96*, 9255-9258.

6. Gould, I.R.; Cornell, W.D.; Hillier, I.H. *J. Amer. Chem. Soc.* **1994**, *116*, 9250-9256.

7. Head-Gordon, T.; Head-Gordon, M.; Frisch, M.; Pople, J.A.; Brooks, C.L. *J. Amer. Chem. Soc.* **1991**, *113*, 5989.

8. Cieplak, P.; Cornell, W.D.; Bayly, C.; Kollman, P.A. "Application of the Mutlimolecule and Multiconformation RESP Methodology to Biopolymers: Derivation for DNA, RNA and Proteins," *J. Comp. Chem.*, submitted.

9. McGregor, M.J.; Islam, S.A.; Sternberg, M.J.E. *J. Mol. Biol.* **1987**, *198*, 295.

10. Applequist, J.B.; Carl, J.R.; Fung, K.-K. *J. Amer. Chem. Soc.* **1972**, *94*, 2952.

11. Cui, W.; Li, F.; Allinger, N.L. *J. Am. Chem. Soc.* **1993**, *115*, 2943.

Chapter 2

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

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

by

Wendy D. Cornell*

Piotr Cieplak †

Christopher I. Bayly §

and

Peter A. Kollman

Department of Pharmaceutical Chemistry

University of California, San Francisco

San Francisco, CA   94143

* Graduate Group in Biophysics

† Permanent Address: Department of Chemistry, University of Warsaw, Pasteur 1, 02-093   Warsaw, Poland

§ Current Address: Merck Frosst Canada Inc., C.P. 1005 Pointe Claire - Dorval, Quebec, H9R 4P8, Canada

Abstract

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

Introduction

It is hard to overestimate the importance of electrostatic effects in the energetics of most intermolecular interactions. The ability to simulate such intermolecular interactions accurately using empirical force fields requires great care in the development of the electrostatic model. The use of *ab initio* electrostatic potential derived (ESP) charges has been a promising start in this pursuit.[1-3] With a suitable basis set for the calculations that is balanced with effective two-body potential water models, e.g. 6-31G*, one expects a very good reproduction of experimental free energies of solvation. This is indeed the case.

One problem with electrostatic potential fit charges, however, is that they are conformationally dependent.[4-6] Furthermore, the conformational energies which are calculated using standard ESP charges are not sufficiently in agreement with experimental results and high level theoretical calculations and therefore require adjustment through the contribution of the torsional energy term. Because charges on common functional groups are not consistent between homologous molecules, one is unable to derive torsional parameters to adjust the conformational energies for certain classes of molecules.[7] That is because any error in the conformational energies resulting from the nonbonded electrostatic contribution is not systematic.

These problems have led to the development of a new charge model which restrains the magnitude of the partial atomic charges that are least well determined by the electrostatic potential -- RESP charges. We show that this model reasonably meets the challenge to restrain the charges on nonpolar groups without greatly reducing the charges on polar groups and thereby having a deleterious effect on important intermolecular interactions such as hydrogen bonding and free energies of solvation. In addition, we address below the issue of whether to attenuate the electrostatic interaction between atoms separated by exactly 3 bonds (1-4 interactions). By comparison with high level *ab initio* calculations on 1,2-

ethanediol, we are able to suggest an optimum 1-4 electrostatic scale factor and evaluate the sensitivity of the model to the exact value of this scale factor.

Below we present the results of studies of conformational energies, hydrogen bonding energies, and free energies of solvation for a model which is a reasonable compromise between the need to have large charges on polar atoms to reproduce intermolecular interaction energies and small charges on nonpolar atoms to reproduce intramolecular conformational energies. The evolution of this model is described in detail in another paper[8] and involves a two-stage fitting. In both stages of the fit, restraints are used only on non-hydrogens atoms. In the first stage, the charges are optimized and any necessary molecular symmetry is imposed by constraining charges on equivalent atoms to have the same value. Two types of equivalent atoms are not constrained to be equivalent in the first stage, however. These are hydrogens within methyl and methylene groups. The carbon and hydrogen atoms in those groups are reoptimized in the second stage of the fit in the presence of frozen charges from the first stage on the other atoms. This two-stage fit was found to be necessary because a one-stage fit which constrained methyl hydrogens to have equivalent charges adversely affected charges on nearby polar atoms.

Methods
A. Charge Models

The derivation of the final charge model is described in detail in another paper.[8] The terminology and notation for describing the charge models are as follows. The term "RESP" is used to refer to any of the restrained ESP models. The models are distinguished by the strength of the restraint used (field 1) and the treatment of the methyl and methylene hydrogens (field 2). Standard ESP charges (un.ap) were calculated according to the method described by Singh and Kollman.[3] The notation then refers to the fact that the charges were unrestrained (un) and that methyl and methylene hydrogen charges were

averaged internally *a posteriori* (ap) to the fit. Even though all three methyl hydrogens are rarely equivalent by formal molecular symmetry, it is necessary for them to have equivalent charges because they will interchange under the conditions of molecular dynamics and should therefore be indistinguishable.

Five other models are examined in this paper. Four of the models resulted from one-stage optimization of the charges with the inclusion of hyperbolic restraints on non-hydrogen atoms. Both a strong restraint of 0.0010 a.u. (st) and a weak restraint of 0.0005 a.u. (wk) were tested. Methyl hydrogen atoms were either averaged *a posteriori* (ap) to the fit or constrained to be equivalent during the fit (eq). The four models arrived at were thus (st.ap), (st.eq), (wk.ap), and (wk.eq).

The fifth and preferred model (wk.fr/st.eq) resulted from a two-stage fitting process where the charges were optimized in the first stage with weak hyperbolic restraints of 0.0005 a.u. on non-hydrogen atoms. In the second stage, charges were frozen on all atoms except those in methyl and methylene groups, and the charges on those atoms were then re-optimized in the presence of strong hyperbolic restraints on the non-hydrogen atoms (i.e. the methyl and methylene carbons). Methyl or methylene hydrogens were thus free (fr) in the first stage, and not constrained to have equivalent charges within each group (eq) until the second stage. When charges on non-methyl or non-methylene atoms needed to be equivalent (such as those on an amino group's two hydrogens or carbons 1 and 4 or 2 and 3 in butane) they were constrained to be so in the first stage. This fifth model is also referred to as the "two-stage" model. The two-stage model presented here is not all-inclusive. Additional issues not addressed in this paper will be examined in a future paper presenting charges for the nucleic acids and amino acids. [9] We summarize our notation in Table I.

Table I. Summary of various charge models examined in this paper. [a,b]

model	hydrogen atoms	non-hydrogen atoms	methyl and methylene hydrogens	other equivalent atoms
un.ap	unrestrained	unrestrained	made equivalent *a posteriori* to fit	made equivalent *a posteriori* to fit
st.ap	unrestrained	strong restraint	made equivalent *a posteriori* to fit	constrained to be equiv. during fit
wk.ap	unrestrained	weak restraint	made equivalent *a posteriori* to fit	constrained to be equiv. during fit
st.eq	unrestrained	strong restraint	constrained to be equiv. during fit	constrained to be equiv. during fit
wk.eq	unrestrained	weak restraint	constrained to be equiv. during fit	constrained to be equiv. during fit
wk.fr/st.eq	unrestrained	stage 1: weak restraint; stage 2: strong restraint; (only methyl and methylene groups, both C's and H's, refit in stage 2 -- other charges frozen)	stage 1: "free" (no constrained equiv.); stage 2: constrained to be equiv. during fit	constrained to be equiv. during stage 1

[a] Atoms made equivalent "*a posteriori* to fit" were made so by averaging their charges. [b] Only results for models employing a *hyperbolic* restraint are presented in this paper. The less satisfactory results obtained using a harmonic restraint are presented in a related paper. 8

B. Bonded and van der Waals Parameters

Bond, angle, and torsion parameters were taken from the Weiner *et al.* [10] all atom force field. Bonded parameters for the aliphatic amino group were adapted from existing parameters in the Weiner *et al.* force field. They are CT-NT: $r = 1.471$ and $K_r = 367.0$; CT-CT-NT: $\theta = 109.7$ and $K_\theta = 80.0$; HC-CT-NT: $\theta = 109.5$ and $K_\theta = 35.0$; CT-NT-H2: $\theta = 109.5$ and $K_\theta = 305$; and X -CT-NT -X: a sixfold degenerate torsion with $V_3 = 1.0$ and a phase of $0°$. Van der Waals parameters used are those adapted or developed for the new force field [11] and are presented in Table II.[10,12-15] The conformational studies employed HC atom types on all of the aliphatic hydrogens.

C. Conformational Energy Analyses

Molecular mechanics minimizations were carried out using the AMBER program.[16] Conformations corresponding to rotational barriers were examined using dihedral constraints imposed in the PARM module. For the 1,2-ethanediol minimizations, starting geometries were defined using canonical trans and gauche dihedral values with no constraints. Conformations which were not minima on the molecular mechanical potential energy surface were minimized with the necessary dihedral constrained to the 6-31G* quantum mechanically optimized value. Quantum mechanics calculations were carried out using the Gaussian 90 suite of programs.[17]

Models for 1,2-ethanediol using different 1-4 electrostatic scale factors were evaluated by three different measures and using three different sets of reference energies: MP3/6-31+G**//HF/6-31G* energies on the four lowest energy conformations [18] and MP2/6-31G*//HF/6-31G* energies on the other six; MP2/6-31G*//HF/6-31G* energies for all conformations; [19] and MM2 [19,20] minimized energies for all conformations. The first of the three measures was simply the sum of the absolute values of the difference between the

Table II. Nonbonded parameters used in the calculations. [a]

AMBER atom type	Description	r* (Å)	ε (kcal/mole)
CT	sp^3 carbon	1.9080	0.1094
C	sp^2 carbon	1.9080	0.0860
NT	sp^3 nitrogen	1.8240	0.1700
N	sp^2 nitrogen	1.8240	0.1700
OH	alcohol oxygen	1.7210	0.2104
OS	ether oxygen	1.6387	0.1700
O	carbonyl oxygen	1.6612	0.2100
OW	TIP3P water oxygen	1.7683	0.1520
HC	hydrocarbon hydrogen	1.4870	0.0157
H1	hydrocarbon hydrogen w/ 1 electronegative neighbor	1.3870	0.0157
HA	aromatic hydrogen	1.4590	0.0150
H	hydrogen on nitrogen	0.6000	0.0157
HW	TIP3P water hydrogen	0.0000	0.0000
S	sulfur	1.9920	0.2500
LP	lone pairs on sulfur	0.0000	0.0000
P	phosphate phosphorous	2.1000	0.2000

[a] CT, C, HC, and HA parameters from unpublished work by Spellmeyer and Kollman, ref. 12. H1 parameters from Veenstra and Ferguson, ref. 13. P parameters from Weiner et al. force field, ref. 10. TIP3P water from Jorgensen ref. 14. All others from OPLS force field of Jorgensen and Tirado-Rives, ref. 15.

relative reference energy and the relative molecular mechanics energy for each conformation. The second measure was a Boltzmann weighted RMS of the difference between the molecular mechanics and reference relative energies. In this case the Boltzmann weight of each conformation was calculated from the reference energy. This procedure then penalized most heavily energy deviations in conformations which were "supposed to" be lower in energy. In the third measure, a Boltzmann weighted RMS was again calculated, but in this case the molecular mechanics energy was used to assign the Boltzmann weight of each conformation. This procedure penalized for conformations which were not supposed to be low in energy but which had low calculated molecular mechanics energies.

One needs to consider both of the Boltzmann weighted RMS values together -- i.e. the one that uses the reference energy as a weight and the one that uses the calculated energy as a weight. This is because each one neglects the problem that the other measure is flagging. The Boltzmann weighted RMS values serve best as a means of eliminating models and will not necessarily directly identify an optimal model.

D.   Hydrogen Bond Energies

DNA base pairs were set up using the computer graphics program MIDAS.[21,22] N-methyl acetamide (NMA) homo-dimer and NMA-water dimer configurations were set up according to Jorgensen and Swensen.[23] Methanol homo-dimers and methanol-water dimers were set up according to Tse and Newton.[24] Each system was then minimized using the AMBER program with conjugate gradient minimization with a constant dielectric of 1 and 1-4 van der Waals and electrostatic scale factors of 1/2.

E.   Free Energy Perturbation Calculations

All simulations were run using the AMBER program with the all-atom type force field.

Each initial system contained the solute trans-NMA with 259 TIP3P water molecules or methanol with 208 TIP3P water molecules. Each system was initially minimized using 1000 cycles of conjugate gradient minimization followed by 20 psec of molecular dynamics equilibration.

The perturbations of methanol to methane or ethane were carried out with over 202 psec of molecular dynamics simulation using the slow growth approach.[25] Van der Waals parameters and charges were perturbed simultaneously. The potential of mean force (pmf) correction, necessary because of the manner in which AMBER defines the topologies of the perturbed groups or molecules, was calculated for perturbed bonds and added to the total free energy change.[26]

The above protocol was carried out using the standard ESP charges. Results for free energy differences based on the other three charge sets were obtained by performing shorter perturbations for methanol, involving only changing the standard ESP charges into the new set. These simulations were carried out using the windows approach[27] with 21 windows and 500 steps of molecular dynamics equilibration and 500 steps of data collection for a total of 42 psec.

The perturbation of NMA to methane was carried out with decoupling of the electrostatic and van der Waals components of the perturbation. NMA standard ESP charges were first perturbed to zero during 404 psec of molecular dynamics using the windows approach with 101 windows consisting of 1000 steps of equilibration and 1000 steps of data collection. The van der Waals perturbation was similarly carried out over 404 psec. For the second part of the electrostatic perturbation, that of perturbing methane's neutral atomic charges to standard ESP charges, results were taken from previously published calculations by Sun *et al.* [28]

The solvation free energies reported for the three RESP models were obtained by carrying out electrostatic perturbations for NMA where only the standard ESP charges were perturbed to RESP charges. These simulations were carried out using the windows approach with 21 windows and 500 steps of molecular dynamics equilibration and 500 steps of data collection for a total of 42 psec. The results were then combined with those from the van der Waal's and electrostatic perturbations described in the preceding paragraph. The necessary pmf correction was included as was the Born correction, [29] needed to account for the long range electrostatic effects of perturbing a dipolar species into a nonpolar one.

All simulations were carried out at a constant pressure of 1 atm and a constant temperature of 300K using a time step of 2 fsec with SHAKE [30] applied to constrain bond lengths to equilibrium values. A constant dielectric of 1 was employed with an 8 A cutoff for nonbonded interactions. Periodic boundary conditions were used. All perturbations were performed in the forward and reverse directions. The values of the free energies and errors reported are the mean values and standard deviations of the forward and reverse runs.

Results

A. Conformational Energies in Butane, Methyl Ethyl Thioether, and Simple Alcohols and Amines

Charges were derived for butane using the standard ESP model, the two one-stage restrained ESP models with *a posteriori* averaging on methyl hydrogens, and the two-stage restrained models. The charge on each atom derived from the trans conformation is four to six times greater in the standard ESP model compared to the two-stage restrained ESP model. The smaller restrained ESP charges are more consistent with the notion of a nonpolar alkane, but it is important to note that both sets of charges reproduce the

molecular electrostatic potential quite well, with the standard ESP charges actually having a slightly smaller relative RMS value (RRMS=0.89 *vs.* RRMS=0.90). This underscores the point that one should not rely too heavily on "chemical intuition" when evaluating charge models.

Table III [31-33] presents the results of molecular mechanics minimization on butane, with the conformational energies and geometries presented as a function of charge model. We should note that the Weiner *et al.* force field used a 1-4 scale factor of 1/2 for electrostatics. This scaling, however, as noted by Billeter *et al.* [34] and Smith and Karplus,[35] can lead to artifacts in the conformational energies if relatively large (e.g. 6-31G* electrostatic potential derived) charges are used.

As one can see from Table III, the conformational energies calculated with standard ESP charges are very sensitive to scale factor and also somewhat sensitive to the conformation from which they were derived. Reynolds *et al.* [36] have previously noted the problems inherent in the conformational dependence of electrostatic potential derived charges. The weak one-stage restrained charge model (wk.ap) is less sensitive; and the two-stage restrained model (wk.fr/st.eq), the model of choice here, is rather insensitive to the the 1-4 scale factor and has much less dependence on the conformation from which the charges were derived. In addition, the most important properties of butane, the relative energies of the trans and gauche conformations, the dihedral angle of the gauche conformation, and the relative energies of the skew and eclipsed conformations are represented in quite good agreement with experiment using only threefold torsions on the C-C dihedrals. In principle, the relative energy of the eclipsed conformation of butane could be adjusted with an additional torsional potential. This is not the purpose of this study, however. Torsion parameters will be fine tuned at a later date.

Table III. Relative molecular mechanics conformational energies of butane (kcal/mol).

model[a]	CCCC	exp.	standard ESP (un.ap)			one-stage RESP (wk.ap)			one-stage RESP (st.ap)			two-stage RESP (wk.fr/st.eq)			mm2[31]	mm3[31]
			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor				
			1.0	1.2	2.0	1.0	1.2	2.0	1.0	1.2	2.0	1.0	1.2	2.0		
trans	trans	**0.00**	**0.00**	**0.00**	0.18	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	gauche	0.75 [32]	0.70	0.42	**0.00**	0.71	0.63	0.45	0.69	0.67	0.61	0.70	0.67	0.61	0.86	0.81
		(71°)	(68°)	(67°)	(65°)	(68°)	(68°)	(67°)	(68°)	(68°)	(68°)	(68°)	(68°)	(68°)	(65°)	(65°)
	ecl	4.56 [33]	5.06	4.69	4.11	5.20	5.08	4.84	5.19	5.16	5.07	5.20	5.16	5.08	4.73	4.83
	skew		3.32	3.25	3.27	3.35	3.33	3.27	3.35	3.34	3.33	3.35	3.34	3.33	3.34	3.30
gauche	trans	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	gauche	0.75	0.66	0.57	0.40	0.68	0.65	0.59	0.68	0.67	0.64	0.66	0.65	0.61	0.86	0.81
		(71°)	(68°)	(68°)	(67°)	(68°)	(68°)	(68°)	(68°)	(68°)	(68°)	(68°)	(68°)	(68°)	(65°)	(65°)
	ecl	4.56	5.10	4.98	4.76	5.17	5.13	5.05	5.17	5.16	5.12	5.14	5.12	5.07	4.73	4.83
	skew		3.33	3.31	3.26	3.35	3.34	3.32	3.35	3.35	3.34	3.34	3.33	3.32	3.34	3.30

a Conformation used to generate the charges.

Charges were calculated for methyl ethyl thioether using the standard, one-stage restrained, and two-stage restrained ESP models. In Table IV [38] we present the relative conformational energies of trans and gauche methyl ethyl thioether. Again, the two-stage model has the least conformational and 1-4 scale factor dependence. Here, the wrong conformer is lower in energy, but this could easily be corrected with a small V_1 or V_2 torsional potential.

In Table V [39-42] we present the conformational energies of methanol, ethanol, and propanol. We use the results of high level *ab initio* calculations as an experimental reference for ethanol and propanol. The experiemental energies for methanol are from microwave spectroscopy. [40] The methanol results are essentially insensitive to both the charge model used and the 1-4 electrostatic scale factor. The scale factor independence is a result of the fact that the 1-4 electrostatic energy is the same for both the staggered and eclipsed conformations. For ethanol, the two-stage model is superior in both its small dependence on 1-4 scale factor and (with 1-4 scale factor of 1/1.2) its excellent agreement with high level *ab initio* theory.

In propanol, the relative *ab initio* energies are very small for all conformations. Neither of the models agrees well with the highest level *ab initio* calculations either in the magnitudes of the relative energies found or the identity of the global minimum conformation. However, while a different set of high level *ab initio* calculations yields similarly small values for the relative conformational energies, yet another conformation is identified as the global minimum. Furthermore, the MM2 calculated energies also vary more in relative magnitude than the *ab initio* energies (although not as much as our molecular mechanics models) and also find the same global minimum conformation as do our models. Because of the disagreement seen with the higher levels of theory over the minimum energy conformation, we decided that we would be satisfied with a model that gave fairly small

Table IV. Relative molecular mechanics conformational energies of methyl ethyl thioether (kcal/mol).

model [a]	CCSC	exp. H [b]	standard ESP (un.ap)			one-stage RESP (wk.ap)			one-stage RESP (st.ap)			two-stage RESP (wk.fr/st.eq)		
			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor			1 / 1-4 elect. scale factor		
			1.0	1.2	2.0	1.0	1.2	2.0	1.0	1.2	2.0	1.0	1.2	2.0
trans	trans	0.05-0.20	**0.00**	**0.00**	0.63	**0.00**	**0.00**	0.14	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	gauche	**0.00**	0.40	0.07	**0.00**	0.35	0.19	**0.00**	0.33	0.24	0.06	0.38	0.35	0.28
		(66°)	(66°)	(65°)	(63°)	(66°)	(65°)	(64°)	(65°)	(65°)	(65°)	(65°)	(65°)	(65°)
gauche	trans	0.05-0.20	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	gauche	**0.00**	0.30	0.17	0.09	0.29	0.19	0.03	0.26	0.21	0.08	0.26	0.23	0.19
		(66°)	(65°)	(65°)	(64°)	(65°)	(65°)	(64°)	(65°)	(65°)	(64°)	(65°)	(65°)	(65°)

[a] Conformation used to generate the charges. [b] Experimental data given in ref. 37.

Table V. Relative molecular mechanics conformational energies for three simple alcohols as a function of charge model and 1-4 electrostatic scale factor (kcal/mol).

molecule	-CCO[a]	-COH[a]	exp.	standard ESP (un.ap) 1 / 1-4 electrost. scale factor			two-stage RESP (wk.fr/st.eq) 1 / 1-4 electrost. scale factor			mm2	mm3 38
				1.0	1.2	2.0	1.0	1.2	2.0		
methanol		stag	**0.00**	0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		ecl	1.07 39	1.01	1.00	1.00	1.02	1.02	1.02	0.87	0.78
ethanol	stag	trans	**0.00**	0.30	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	stag	g+	0.40	**0.00**	0.22	1.04	0.36	0.46	0.65	0.60	0.40
			(54°) 40	(55°)	(62°)	(73°)	(64°)	(65°)	(67°)	(65°)	
	stag	ecl	1.66 b	0.65	1.19	2.65	1.50	1.67	2.00		
	ecl	trans	3.08 41	3.49	3.16	3.09	3.10	3.09	3.07	2.73	2.80
propanol	trans	g+	0.19 b	1.41	1.14	3.92	0.99	0.94	1.74	0.62	
	g+	g-	0.33	1.78	0.22	**0.00**	1.32	0.60	**0.00**	0.97	
	g+	g+	0.17	3.01	1.54	1.55	2.36	1.64	1.06	1.03	
	trans	trans	0.18	**0.00**	**0.00**	3.40	**0.00**	**0.00**	0.90	**0.00**	
	g+	trans	**0.00**	1.78	0.67	1.55	1.56	0.93	0.50	0.34	

[a] First dihedral atom refers to a carbon, when present. [b] MP3/6-31+G**//HF/6-31G* energies for ethanol CCOH barrier and propanol -- unpublished results from T.A. Halgren.

conformational energies and chose not to focus on the minimum energy conformation.

For propanol, the two-stage restrained model is both less dependent on the 1-4 scale factor, and with a scale factor of 1/1.2, in respectable agreement with the MM2 model. The fact that the two-stage RESP charges result in lower 1-4 electrostatic energies also means that minimized geometries are less sensitive to the choice of 1-4 electrostatic scale factor. For example, with the standard ESP charges, the minimized value of the CCOH dihedral in the Gg conformation ranges from 67° to 49°, depending on whether a 1-4 electrostatic scale factor of 1 or 1/2 was used, respectively. When two-stage RESP charges are used, this range is reduced to 63° to 60°, and agrees well with the MM2 value of 62°.

Finally, we turn to the amines. In Table VI [43-46], the conformational energies are presented as a function of charge model and 1-4 electrostatic scale factor. Methylamine is shown to be sensitive to the 1-4 scale factor and its barrier to rotation is calculated as being in excellent agreement with the experimental value. For ethylamine using the standard ESP charges, the minimum energy conformation and conformational energy difference is quite sensitive to the scale factor. This sensitivity is reduced with the restrained ESP charges, and using a scale factor of 1/1.2, the conformation having the lone pair gauche to the beta carbon is found to bo 0.59 kcal/mole higher in energy than the trans conformation -- about 0.3 kcal/mole too high. Again, this difference could possibly be adjusted with torsion parameters, depending on the error seen with this particular torsion in other contexts (such as propylamine). Our error is on the order of that given by MM3 which finds the trans conformation to be higher in energy by 0.1 kcal/mole.

The propylamine energies are evaluated against high level *ab initio* calculations. As withpropanol, the conformational energy differences are quite small. Here both the standard ESP and restrained ESP charges yield good results when a scale factor of 1/1.2 is used. All energies are about 1 kcal/mole or less and the minimum energy conformation

Table VI. Relative molecular mechanics conformational energies for three simple amines as a function of charge model and 1-4 electrostatic scale factor (kcal/mol).

molecule	-CCN[a]	-CNlp[a]	exp.	standard ESP (un.ap) 1/1-4 electrost. scale factor			two-stage RESP (wk.fr/st.eq) 1/1-4 electrost. scale factor			mm2[42]	mm3[43]
				1.0	1.2	2.0	1.0	1.2	2.0		
methylamine	stag		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
	ecl		1.98[44]	1.97	1.98	1.99	2.01	2.01	2.01	1.90	1.45
ethylamine	stag	trans	**0.00**	**0.00**	**0.00**	0.38	**0.00**	**0.00**	**0.00**	0.13	0.10
	stag	g	0.3[45]	1.59	0.89	**0.00**	0.84	0.59	0.10	**0.00**	**0.00**
propylamine	trans	g+	0.30[b]	0.01	0.61	3.16	0.20	0.20	0.94	**0.00**	(**0.00**)[c]
	g+	trans	0.32	0.55	**0.00**	**0.00**	0.53	0.11	**0.00**	1.00	0.42
	trans	trans	**0.00**	**0.00**	0.43	2.63	**0.00**	**0.00**	0.72	0.13	**0.00**
	g+	g-	0.25	1.32	1.01	1.50	1.35	0.96	0.90	0.65	(0.48)
	g+	g+	0.61	0.73	0.37	0.75	0.83	0.44	0.35	0.88	(0.88)

a First dihedral atom refers to a carbon, when present. b MP3/6-31+G*//HF/6-31G* energies for propylamine -- unpublished results from T.A. Halgren. c Schmitz and Allinger report two sets of relative energies for the five propylamine conformers. One set is relative to the ground state with the CCCN torsion in the trans conformation and the other set is for a ground state with CCCN in the gauche conformation.

found by our molecular mechanics models is in agreement with the high level *ab initio* calculations.

B. 1-4 Electrostatic Scale Factor Calculations on 1,2-Ethanediol

We chose 1,2-ethanediol as a particularly sensitive model system for examining conformational energies as a function of charge model and 1-4 electrostatic scale factor. This sensitivity arises from the fact that 1,2-ethanediol has three major dihedrals with two polar atoms in a 1-4 configuration. One can define ten unique conformations where each of the three dihedrals is in either a trans or gauche conformation.

Table VII presents the results obtained using only the two-stage fit charges (obtained from the all-trans conformation), showing the relative conformational energies as a function of 1-4 scale factor. These energies are compared with two *ab initio* quantum mechanical models: the first is MP3/6-31+G**//HF/6-31G* energies for the four lowest energy conformations [18] and MP2/6-31G*//HF/6-31G* energies for the rest; the second is MP2/6-31G*//HF/6-31G* energies for all ten conformations. A third set of energies which is used for comparison were calculated using MM2.

This data show that based on the absolute errors, a scale factor of 1/1.1 performs best with the restrained ESP charges. However, the lowest energy conformations are arguably the most important, and the Boltzmann weighted RMS values show that a slighly smaller scale factor results in better agreement for the four lowest energy conformations. We feel that a scale factor of 1/1.2 gives the best agreement with the *ab initio* energies and is significantly superior to a scale factor of 1 or 1/2 in ths regard. This model performs even better than MM2 on this molecule, even though no reoptimization of torsional potentials has been done - -only the standard threefold parameters from the Weiner *et al.* [10] force field were used. The choice of a scale factor of 1/1.2 is supported by the conformational

Table VII. Relative molecular mechanics conformational energies of 1,2-ethane diol as a function of conformation and 1-4 electrostatic scale factor using two-stage RESP (wk.fr/st.eq) charges (kcal/mol). [a]

conf. [b]	(MP3/MP2) [c]	(MP2) [d]	(MM2) [e]	1 / 1-4 electrostatic scale factor					
				1.0	1.1	1.2	1.3	1.5	2.0
t T t	2.69	3.37	1.51	0.00	1.16	2.58	3.91	6.24	10.53
t G t	4.45	4.45	2.19	3.75	4.51	5.16	5.81	7.03	9.40
t T g	3.41	3.41	2.52	1.50	2.53	3.83	5.04	7.18	11.13
g G t	4.65	4.65	3.21	5.96	5.94	6.37	6.82	7.69	9.45
g T g	3.49	3.49	3.75	3.32	4.17	5.31	6.39	8.30	11.81
g G g	3.56	3.56	2.63	4.16	3.57	3.47	3.46	3.57	4.01
g T g-	3.15	3.15	3.05	2.20	3.05	4.19	5.27	7.19	10.73
g G g-	0.65	0.24	1.11	1.86	1.30	1.13	0.99	0.75	0.40
t G g-	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00
g G- g	1.22	1.25	1.16	1.56	1.22	1.26	1.33	1.46	1.69

Table VII. Cont'd.

		1/1-4 electrostatic scale factor				
	1.0	1.1	1.2	1.3	1.5	2.0
Sum of absolute errors						
reference = MP3/MP2	7.58	10.24	5.20	6.43	11.95	22.14	42.38
reference = MP2	8.70	11.30	6.32	7.49	11.65	21.84	41.58
reference = MM2	0.00	11.16	7.02	12.17	18.13	29.00	49.44
Boltzmann weighted RMS (wt from ref E)						
reference = MP3/MP2	0.25	0.70	0.34	0.25	0.28	0.49	0.99
reference = MP2	0.54	1.04	0.65	0.55	0.49	0.48	0.73
reference = MM2	0.00	0.63	.0.37	0.54	0.84	1.45	2.61
Boltzmann weighted RMS (wt from calc E)						
reference = MP3/MP2	0.46	2.08	0.53	0.17	0.14	0.08	0.17
reference = MP2	0.62	2.58	0.77	0.32	0.29	0.24	0.13
reference = MM2	0.00	1.18	0.14	0.14	0.12	0.19	0.42

[a] Charges derived from 6-31G* optimized all-trans conformation. [b] Conformations described using a capital letter for the central torsion (OCCO) and small letters for the CCOH torsions. [c] This set uses MP3/6-31+G**//HF/6-31G* energies for the tTt, gGg-, tGg-, and gG-g conformations (unpublished results from T.A. Halgren) and MP2/6-31G*//HF/6-31G* energies for remaining conformations. [d] MP2/6-31G*//HF/6-31G* energies for all conformations.

energies seen for butane, methyl ethyl thioether, and the simple alcohols and amines discussed above.

C. Hydrogen Bond Energies

In Table VIII [47-48] we present the results of calculations on the hydrogen bond energies of $1\text{-}CH_3$ Thymine : $9\text{-}CH_3$ Adenine in Watson-Crick and Hoogsteen geometries and $1\text{-}CH_3$ Cytosine : $9\text{-}CH_3$ Guanine in the Watson-Crick geometry. The results using the standard ESP charges are in fairly good agreement with the *ab initio* results, except that the Hoogsteen and Watson-Crick relative energies are reversed for the adenine-thymine base pairs. The use of the two-stage RESP charges leads to lower hydrogen bond energies which are in better agreement with experiment [47] for GC, but poorer for AT. Finally, the two-stage RESP charges restore the greater stability of A-T Hoogsteen over Watson-Crick hydrogen bonds found in the quantum mechanical results.

In Tables IX and X, the hydrogen bond energies and distances are presented for methanol and trans-NMA complexes with water as well as their homo-dimers. Both standard ESP and two-stage RESP charges lead to nearly identical H-bond energies for methanol, whereas the hydrogen bond energies are about 0.3-0.5 kcal/mole weaker for NMA using the two-stage RESP model.

D. Solvation Free Energies

We have carried out free energy calculations on the aqueous solvation of methanol and NMA to evaluate the effect of changing the charges from standard ESP (un.ap) to three RESP models: the weak hyperbolic (wk.ap) and (wk.eq) models and the two-stage (wk.fr/st.eq) models. The free energies reported for the RESP models were calculated by considering the effect of perturbing the standard ESP charges (st.ap) for methanol or NMA into the two-stage RESP (wk.fr/st.eq) model. These results were then added to the results

Table VIII. DNA base pairing energies and distances.

base pair [a]	model [b]	-ΔE [c] (kcal/mol)	-ΔH$_{298}$ [d] (kcal/mol)	R (H-bonds) [e] (Å)		
A-T Watson-Crick	exp.	—	—	2.95	2.82	
	QM	12.5	10.5	3.06	2.88	
	standard ESP	14.0	11.9	2.93	2.89	
	two-stage RESP	12.6	10.7	2.90	2.91	
A-T Hoogsteen	exp.	—	13.0	2.86	2.93	
	QM	13.5	11.3	3.09	2.85	
	standard ESP	13.8	11.9	2.89	2.95	
	two-stage RESP	13.0	11.2	2.91	2.90	
G-C Watson-Crick	exp.	—	21.0	2.91	2.95	2.86
	QM	26.0	23.2	2.87	3.01	2.96
	standard ESP	28.1	25.6	2.88	2.92	2.87
	two-stage RESP	27.2	24.8	2.86	2.92	2.84

[a] Base pair types -- see Saenger, ref. 22. [b] exp = experimental enthalpies from ref. 46. QM = *ab initio*, MP2/6-31G*//HF/6-31G* with basis set superposition error correction, as described in ref. 47. "standard esp" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) charge models. [c] Minimized energies. [d] After normal mode and thermal corrections for theoretical energies, see ref. 47. [e] Experimental H-bond distances from Saenger.

Table IX. Hydrogen bonding energies and distances for NMA-water and methanol-water interactions with different charge models. [a]

molecule	charge model		H_2O as proton acceptor [b,c]		H_2O as proton donor [b,c]	
			distance (Å)	ΔE(kcal/mole)	distance (Å)	ΔE(kcal/mole)
NMA	standard ESP	(un.ap)	1.92	-7.1	1.70	-9.7
	one-stage RESP	(wk.eq)	1.95	-6.3	1.73	-8.8
	two-stage RESP	(wk.fr/st.eq)	1.94	-6.7	1.71	-9.4
methanol	standard ESP	(un.ap)	1.75	-7.0	1.80	-6.3
	one-stage RESP	(wk.eq)	1.78	-6.1	1.83	-5.8
	two-stage RESP	(wk.fr/st.eq)	1.75	-7.0	1.81	-6.3

[a] Water model used is TIP3P, ref. 14. [b] Configuration definitions for NMA-water according to Jorgensen and Swensen, ref. 23. [c] Configuration definitions for methanol-water according to Tse and Newton, ref. 24.

Table X. Hydrogen bonding energies and distances for NMA and methanol homo-dimers with different charge models.

dimer	configuration [a]	charge model		ΔE(kcal/mole)	distance (Å)		
					$r_{O...H}$	$r_{N...O}$	$r_{O...O}$
NMA	parallel	standard ESP	(un.ap)	-10.2	1.82	2.84	
		one-stage RESP	(wk.eq)	-8.8	1.87	2.88	
		two-stage RESP	(wk.fr/st.eq)	-9.8	1.88	2.87	
NMA	antiparallel	standard ESP	(un.ap)	-10.4	1.84	2.85	
		one-stage RESP	(wk.eq)	-8.9	1.88	2.89	
		two-stage RESP	(wk.fr/st.eq)	-9.7	1.86	2.87	
NMA	stack	standard ESP	(un.ap)	-8.6			
		one-stage RESP	(wk.eq)	-7.9			
		two-stage RESP	(wk.fr/st.eq)	-8.2			
methanol		standard ESP	(un.ap)	-6.9	1.78		2.75
		one-stage RESP	(wk.eq)	-5.6	1.83		2.80
		two-stage RESP	(wk.fr/st.eq)	-6.8	1.76		2.76

[a] Configuration definition according to Jorgensen and Swensen, ref. 23.

from the molecular perturbations using standard ESP charges. The effects of perturbing methane or ethane using standard ESP charges (un.ap) into the two-stage RESP charges (wk.fr/st.eq) was within the noise of the calculations, so we did not carry out those calculations for the other two RESP models (wk.ap and wk.eq).

The free energies of solvation of methanol relative to both ethane and methane are presented in Table XI. [49] As one can see, the standard ESP (un.ap) charges as well as the wk.ap and two-stage RESP (wk.fr/st.eq) models lead to solvation free energies very close to experiment. [49] Forcing equivalence on the methyl hydrogens in a one-stage fit (wk.eq) leads to a significantly less favorable solvation free energy. That is why we do not favor the use of this model over the more elaborate two-stage approach.

The relative free energies of solvation of NMA and methane with three charge models are presented in Table XII. [49] The standard ESP (un.ap) and weak hyperbolic (wk.ap) models both lead to a solvation free energy (ΔG=-12.4 kcal/mole and ΔG=-12.1 kcal/mole) in good agreement with experiment (ΔG=-12.2 kcal/mole). [49,50] The two-stage model (wk.fr/st.eq) is less accurate but still good, resulting in ΔG=-11.6 kcal/mole.

E. Conformational Dependence of ESP Charge Models: Intermolecular Effects

There are two issues which can be defined with respect to the conformational dependence of electrostatic potential fit charges. The first issue is how well the charges derived from one particular conformation of a molecule reproduce the electrostatic potential of another conformation of the molecule. Tables XIIIA and XIIIB present the results of calculations on five low energy conformations of propylamine examining the conformational dependence of intermolecular properties. We describe each conformation using a capital letter for the C-C-C-N torsion and a small letter for the C-C-N-lp torsion. Lone pairs are

Table XI. Relative free energies of solvation for the perturbations of methanol --> methane and methanol --> ethane with different charge models.

perturbation	charge model [a]		$\Delta\Delta G_{solv}$ (kcal/mole)
methanol --> methane	experiment [b]		7.0
	standard ESP	(un.ap)	6.91±0.01
	one-stage RESP	(wk.eq)	5.71±0.02
	one-stage RESP	(wk.ap)	6.83±0.01
	two-stage RESP	(wk.fr/st.eq)	6.86±0.01
methanol --> ethane	experiment [b]		6.8
	standard ESP	(un.ap)	7.02±0.13
	one-stage RESP	(wk.eq)	5.82±0.14
	one-stage RESP	(wk.ap)	6.90±0.13
	two-stage RESP	(wk.fr/st.eq)	6.93±0.13

[a] Charges for methane are C=-0.464 and H = 0.116 and for ethane C = -0.027 and H=0.009 with the standard ESP model. Using the two-stage RESP model, methane charges are C = -0.390 and H = 0.098 and ethane charges are C = 0.009 and H = -0.003.
[b] Experimental numbers from Ben-Naim and Marcus, ref. 48.

Table XII. Relative free energies of solvation for the perturbation of NMA --> methane with different charge models.

perturbation	charge model		$\Delta\Delta G_{solv}$ (kcal/mole)
1. NMA electrostatic (q --> 0)	standard ESP	(un.ap)	11.23±0.01
2. methane electrostatic (q --> 0)	standard ESP	(un.ap)	0.04±0.01
3. NMA --> methane (VDW)			0.90±0.14
4. Born correction [a]			0.3
NMA --> methane (1. + 3. + 4. - 2.)	standard ESP	(un.ap)	12.4±0.2
NMA --> methane [b]	one-stage RESP	(wk.eq)	10.2±0.2
NMA --> methane [b]	one-stage RESP	(wk.ap)	12.1±0.2
NMA --> methane [b]	two-stage RESP	(wk.fr/st.eq)	11.6±0.2
NMA --> methane	experiment [c]		12.2

[a] The Born correction (ref. 29) accounts for the long range electrostatic effects resulting from perturbing a dipolar species (NMA) to a nonpolar one. [b] Results with non-standard esp charge models obtained by adding $\Delta\Delta G_{solv}$ for perturbation of standard esp charges to given non-standard model to $\Delta\Delta G_{solv}$ for NMA --> methane using standard esp charges. [c] Experiment from refs. 48 and 49.

Table XIII A. Effect of conformational dependence of propylamine standard and two-stage restrained ESP charges on calculated dipole moments (D).

test conf[a]	E (MP3)[b]	q.m.[c]	raw std esp[d]	charge model[e,f]									
				Tt		Tg		Gg-		Gt		Gg	
				std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP
Tt	0.00	1.55	1.53	**1.38**[g]	**1.48**	2.12	2.09	1.92	1.86	1.44	1.48	2.05	2.03
Tg	0.30	1.43	1.42	2.09	2.01	**1.78**	**1.73**	1.83	1.88	2.17	2.12	1.79	1.63
Gg-	0.25	1.49	1.47	2.48	2.44	2.48	2.39	**1.79**	**1.55**	2.23	2.11	1.79	1.91
Gt	0.32	1.56	1.50	1.44	1.52	2.13	2.09	1.93	1.92	**1.43**	**1.48**	2.05	1.97
Gg	0.61	1.41	1.40	2.16	2.09	1.86	1.80	1.84	1.92	2.18	2.13	**1.80**	**1.59**
sum of abs. errors:		0.12	0.13	2.69	2.32	2.93	2.66	1.87	1.69	2.49	2.18	2.04	1.69
range of dips.:		0.15	0.13	1.10	0.96	0.70	0.66	0.14	0.37	0.80	0.65	0.26	0.44

[a] Conformations described by C-C-C-N and c-c-n-lp torsions. [b] MP3/6-31+G**//HF/6-31G* energies in kcal/mol -- unpublished results from T.A. Halgren. [c] Dipole moment calculated from quantum mechanical wavefunction. [d] Dipole moment calculated from unaveraged esp charges derived from and tested against the potential of the same conformation. [e] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) models. [f] Model conformation is the one used to generate the charges. [g] Numbers in bold correspond to situations where charges were derived from and tested on the same conformation.

Table XIII B. Effect of conformational dependence of propylamine standard and two-stage restrained ESP charges on the relative RMS (rrms) of the fit of the classical electrostatic potential to the quantum mechanical potential.

test conf[a]	E (MP3)[b]	raw std esp[c]	charge model[d,e]									
			Tt std ESP	Tt two-stage RESP	Tg std ESP	Tg two-stage RESP	Gg- std ESP	Gg- two stage RESP	Gt std ESP	Gt two-stage RESP	Gg std ESP	Gg two-stage RESP
Tt	0.00	0.20	**0.21**[f]	**0.21**	0.36	0.35	0.31	0.33	0.23	0.24	0.34	0.38
Tg	0.30	0.17	0.42	0.38	**0.27**	**0.26**	0.32	0.37	0.45	0.43	0.31	0.29
Gg-	0.25	0.20	0.57	0.53	0.59	0.54	**0.31**	**0.27**	0.45	0.40	0.33	0.35
Gt	0.32	0.21	0.33	0.32	0.44	0.41	0.30	0.32	**0.23**	**0.23**	0.33	0.35
Gg	0.61	0.20	0.51	0.47	0.42	0.39	0.31	0.35	0.46	0.43	**0.30**	**0.26**
sum of RRMS's:		0.98	2.04	1.91	2.08	1.95	1.55	1.64	1.82	1.73	1.61	1.63

[a] Conformations described by C-C-C-N and c-c-n-lp torsions. [b] MP3/6-31+G**//HF/6-31G* energies -- unpublished results from T.A. Halgren. [c] Rrms's calculated from unaveraged esp charges derived from and tested against the potential of the same conformation. [d] Charge model conformation is the one used to generate the charges. [e] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) models. [f] Numbers in bold correspond to situations where charges were derived from and tested on the same conformation.

not actually used on nitrogen atoms in our force field model, but the "virtual" lone pair is used to define the conformation since it is unique, whereas the two amino hydrogens are not.

Table XIIIA gives the dipole moment calculated for each of the five conformations using the quantum mechanical wavefunction, unaveraged standard ESP charges (un.fr), and standard (un.ap) and two-stage restrained (wk.fr/st.eq) ESP charges calculated from each of the five conformers. The unaveraged standard ESP charges naturally give the best agreement, but cannot be used in simulations without averaging (vide infra).

The numbers shown in bold represent dipole moments calculated when charges were derived from and tested on the same conformation. Not surprisingly, they show the best agreement with the quantum mechanical dipole moments. The five dipole moments shown in bold which pertain to the standard ESP model show that in three of the five cases, the dipole moment is increased by a.p. averaging, while in two of the cases it is decreased. These deviations result from a.p. averaging and not conformational dependence. The dipole moments shown in standard type reflect the conformational dependence of the standard and two-stage restrained ESP models. Of the twenty non-bold dipole moments relating to the standard ESP model, in eighteen of the cases the dipole moment given by the charge set overestimates the quantum mechanical dipole by up to 67%. This could result in the overstabilization in solution of conformations which were not used in the ESP charge derivation, due to their spuriously large dipole moments. It is encouraging that in 22 of the 25 combinations of model conformation/test conformation examined, the two-stage restrained ESP charges gave dipole moments which were closer to the quantum mechanical values than the standard ESP charges did.

Another way of evaluating the calculated dipole moments given by the different charge

models is to calculate the range of dipole moments for the five conformations. This data is also presented in Table XIIIA. The quantum mechanical dipole moments span a range of only 0.15 D. A higher sum of absolute errors might be acceptable if the range of numbers was appropriately small, since no one of the conformations would then be overstabilized in solution by a *relatively* large dipole moment. By this measure neither the standard or two-stage restrained ESP models is better in general. Interestingly, both the Gg- and Gg conformations provide charges which perform quite well as judged by the dipole moment error, range of dipole moments and sum of RRMS's.

F. Conformational Dependence of ESP Charge Models: Intramolecular Effects

The second issue which can be defined with respect to conformational dependence is to compare conformational energies calculated with charges derived from different conformations of a molecule. Table XIV presents the conformational energies calculated for five conformers of propylamine as a function of charge model and the conformation used for the charge calculation. The MP3/6-31+G**//HF/6-31G* energies are shown as a reference. The Tt conformation yields standard ESP charges which do not reproduce the correct global minimum energy conformation. In fact, only one of the five standard ESP charge sets gives the proper global minimum. The Tt two-stage restrained ESP charges do give the proper global minimum conformation, as do two of the other four sets of two-stage restrained charges.

For all of the conformations except for Gg, the two-stage restrained ESP charges have better agreement with the quantum mechanical conformational energies than the standard ESP charges do. It is interesting to note that the Gg- conformation produces very good agreement with the quantum mechanical conformational energies using either the two-stage restrained or standard ESP models. Urban and Famini recently reported the results of a

Table XIV. Effect of conformational dependence of propylamine standard and two-stage restrained ESP charges on relative conformational energies (kcal/mol). [a]

test conf[b]	E MP3[c]	charge model [d,e] Tt std ESP	Tt two-stage RESP	Tg std ESP	Tg two-stage RESP	Gg- std ESP	Gg- two-stage RESP	Gt std ESP	Gt two-stage RESP	Gg std ESP	Gg two-stage RESP
Tt	**0.00**	0.43	**0.00**	1.36	0.51	0.05	**0.00**	**0.00**	**0.00**	0.19	0.43
Tg	0.30	0.61	0.20	0.60	**0.00**	**0.00**	0.15	0.55	0.54	**0.00**	**0.00**
Gg-	0.25	1.01	0.96	1.18	1.01	0.32	0.25	0.96	0.93	0.29	0.63
Gt	0.32	**0.00**	0.11	0.31	0.35	0.25	0.59	0.15	0.36	0.28	0.39
Gg	0.61	0.37	0.44	**0.00**	0.09	0.24	0.63	0.75	0.90	0.14	0.13
sum of abs. errors:		2.06	1.19	3.21	2.12	0.86	0.44	1.27	1.25	1.04	1.66

[a] Energies calculated using 1-4 electrostatic scale factor of 1/1.2. [b] Conformations described by C-C-C-N and c-c-n-lp torsions. [c] MP3/6-31+G**//HF/6-31G* energies -- unpublished results from T.A. Halgren. [d] Charge model conformation is the one used to generate the charges. [e] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) charge models.

study on the conformational dependence of ESP charges calculated for dopamine. [51] They found that of the six conformations examined, the standard ESP charges calculated from the highest energy conformation (5.5 and 15.8 kcal/mole above the global minimum on the STO-3G and 6-31G* potential energy surfaces, respectively) did the best job of reproducing conformational energies. Both of these results are somewhat surprising and intriguing and merit further study. In particular, it would be of interest to examine the conformational behavior of two-stage restrained ESP charges calculated for dopamine to determine if the superior performance of the charges derived from the higher energy conformation was retained.

G. Multiple Molecule Fit Charges

Reynolds *et al.* [36] have shown that it is possible to derive ESP charges from more than one conformation of a molecule, and using such a procedure they obtained a set of charges for propanol which reproduced the dipole moments of different conformers better than any set of charges derived from a single conformation. Given recent advances in quantum chemistry software due to the implementation of direct SCF methods and more efficient integral routines, [17] the computational burden associated with carrying out the requisite 6-31G* level SCF calculation on a molecule has been greatly reduced. It is therefore feasible to consider carrying out multiple molecule fitting in order to obtain the highest quality charges possible.

Standard and two-stage restrained ESP charges were calculated for propylamine using all five conformers, the Tt and Gt conformers, the Tt and Tg conformers, and the Tg and Gt conformers. As was the case with propanol,[36] these multiple conformation fit charges consistently result in good agreement with the quantum mechanical dipole moments and overall potentials (Tables XVA and XVB). All of the single and multiple conformation charge sets result in similar rrms values for the five conformers. The five-conformer

Table XVA. Effect of RESP model and multiple conformation fitting of propylamine charges on calculated dipole moments (D).

test conf [a]	E (MP3) [b]	q.m. [c]	Tt std ESP	Tt two-stage RESP	multi(5) std ESP	multi(5) two-stage RESP	multi(2) Tt,Gt std ESP	multi(2) Tt,Gt two-stage RESP	multi(2) Tt,Tg std ESP	multi(2) Tt,Tg two-stage RESP	multi(2) Tg,Gt std ESP	multi(2) Tg,Gt two-stage RESP
Tt	0.00	1.55	**1.38** [f]	**1.48**	**1.79**	**1.78**	**1.41**	**1.48**	**1.71**	**1.76**	1.74	1.77
Tg	0.30	1.43	2.09	2.01	**1.90**	**1.85**	2.12	2.04	**1.92**	**1.86**	**1.94**	**1.85**
Gg-	0.25	1.49	2.48	2.44	**2.11**	**1.92**	2.35	2.13	2.45	2.40	2.29	2.09
Gt	0.32	1.56	1.44	1.52	**1.78**	**1.77**	**1.41**	**1.47**	1.75	1.78	**1.72**	**1.73**
Gg	0.61	1.41	2.16	2.09	**1.92**	**1.85**	2.15	2.05	2.00	1.94	1.96	1.85
sum of abs. errors:			2.69	2.32	2.06	1.73	2.58	2.05	2.39	2.30	2.21	1.85
					(0.78) g		(1.73)		(1.47)		(1.00)	
range of dip. moms: 0.15			1.10	0.96	0.33	0.15	0.94	0.66	0.74	0.64	0.57	0.36

[a] Conformation described by C-C-C-N and c-c-n-lp torsions. [b] MP3/6-31+G**//HF/6-31G* energies -- unpublished results from T.A. Halgren (kcal/mol). [c] Dipole moment from 6-31G* wavefunction. [d] Charge model conformation is the one used to generate the charges. [e] "std" and "two-stage" refer to (un.ap) and (wk.fr/st.eq) models. [f] Numbers in bold refer to situations where the charges are being tested on (one of) the conformation(s) used to derive the charges. [g] Errors given in parentheses are those obtained when equivalencing of hydrogens is forced during fit.

Table XVB. Effect of RESP model and multiple conformation fitting of propylamine charges on the relative RMS (rrms) of the fit of the classical electrostatic potential to the quantum mechanical potential.

test conf[a]	E (MP3)[b]	q.m.[c]	Tt std ESP	Tt two-stage RESP	multi(5) std ESP	multi(5) two-stage RESP	multi(2) Tt,Gt std ESP	multi(2) Tt,Gt two-stage RESP	multi(2) Tt,Tg std ESP	multi(2) Tt,Tg two-stage RESP	multi(2) Tg,Gt std ESP	multi(2) Tg,Gt two-stage RESP
Tt	0.00	0.20	**0.21** [f]	**0.21**	**0.25**	**0.28**	**0.21**	**0.22**	**0.23**	**0.25**	0.25	0.27
Tg	0.30	0.17	0.42	0.38	**0.31**	**0.32**	0.42	0.39	**0.32**	**0.30**	**0.32**	**0.30**
Gg-	0.25	0.20	0.57	0.53	**0.40**	**0.32**	0.50	0.40	0.55	0.51	0.47	0.38
Gt	0.32	0.21	0.33	0.32	**0.27**	**0.27**	**0.26**	**0.23**	0.35	0.34	**0.28**	**0.27**
Gg	0.61	0.20	0.51	0.47	**0.35**	**0.32**	0.47	0.41	0.45	0.41	0.39	0.33
sum of RRMS's:		0.98	2.04	1.91	1.58 (1.37) [g]	1.51	1.86 (1.68)	1.65	1.90 (1.72)	1.81	1.71 (1.40)	1.55

[a] Conformation described by C-C-C-N and c-c-n-lp torsions. [b] MP3/6-31+G**//HF/6-31G* energies -- unpublished results from T.A. Halgren (kcal/mol). [c] Unaveraged charges derived from and tested against the potential of the same conformation. [d] Charge model conformation is the one used to generate the charges. [e] "std" and "two-stage" refer to (un.ap) and (wk.fr/st.eq) models. [f] Numbers in bold refer to situations where the charges are being tested on (one of) the conformation(s) used to derive the charges. [g] Errors given in parentheses are those obtained when equivalencing of hydrogens is forced during fit.

standard ESP model is superior to all but one of the single molecule standard ESP charge sets and the five-conformer two-stage restrained model is superior to all of the single molecule standard or two-stage restrained ESP charge sets.

The range of dipole moments calculated from each charge set are also presented in Table XVA. By this measure the five-conformer models do particularly well, especially the two-stage restrained model. Of the three two-conformer models, the Tg/Gt models clearly exhibits the best behavior, performing nearly as well as the five-conformer models. Most importantly, based on the three measures of intermolecular behavior -- sum of dipole moment errors, range of dipole moment, and sum of rrms's -- the multiple conformation two-stage restrained ESP charges outperform the corresponding multiple conformation standard ESP charges in every case. Restraining the charges thus achieves improvement beyond that available through multiple molecule fitting.

The standard ESP multiple molecule charges were obtained by constraining corresponding heavy atoms to be equivalent between the different conformations, while all hydrogens were left free. Equivalent hydrogens were then averaged *a posteriori*. In parentheses we present the results obtained when all equivalent atoms were constrained to have equivalent charges during the fit (i.e. methyl, α—methylene, and β—methylene hydrogens were constrained to be equivalent within each group and between conformations). While these charge sets result in much better agreement between the calculated classical and quantum mechanical dipole moments for the five conformations, in three of the four sets this forced equivalence of hydrogens results in a significantly reduced charge on the nitrogen atom. The nitrogen charge changes from -1.046 to -0.914 (five-conformer model), from -1.080 to -0.950 (Tt/Tg model), and from -1.063 to -0.934 (Tg/Gt model). It is therefore likely that such charge sets would result in unacceptably low solvation free energies, as was the case with methanol and NMA (vide infra).

The intramolecular effects of propylamine multiple conformation fit charges are examined in Table XVI. Considering the standard and two-stage restrained RESP charges derived using all five conformers, both models result in very small relative conformational energies, in good agreement with the high level quantum mechanical results. The two-stage restrained model comes closer to finding the proper global minimum conformation, and has an overall sum of absolute errors equal to 0.66 kcal/mol as compared to 1.46 kcal/mol for the five-conformation standard ESP charge model. The single conformation Tt standard and two-stage restrained ESP charges yielded conformational energies with absolute errors of 2.06 and 1.19 kcal/mol, respectively. Since the relative quantum mechanical energies for all five conformations are close to zero, it was thought unnecessary to Boltzmann weight the errors. Of the models examined in Table XVI, three of the two-stage RESP models identified the proper global minimum conformation and none of the standard ESP models did.

Of the two-conformation multiple molecule models studied, the standard and two-stage restrained ESP charge models which employed the Tt and Gt conformations performed best of all, with absolute errors of 0.69 and 1.50 kcal/mol, respectively. The Tg/Gt model also did quite well.

Since the number of conformations of a molecule increases exponentially with the number of rotatable dihedrals, it will be important to identify the dihedral types which most affect electrostatic potential derived charges. In the case of propylamine restrained ESP charges, it appears that a more robust set of charges is obtained by varying the dihedral which positions the N (i.e. CCCN) than by varying the CCNlp dihedral. The most important result is that in all of the multiple conformation models examined, the two-stage restrained ESP models consistently outperformed the corresponding standard ESP models by from 29% to 54% in reproducing conformational energies.

Table XVI. Effect of RESP model and multiple conformation fitting of propylamine charges on relative conformational energies (kcal/mol). [a]

		charge model [d,e]									
		Tt		multi (5)		multi (2) Tt,Gt		multi (2) Tt,Tg		multi (2) Tg,Gt	
test conf [b]	E (MP3) [c]	std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP	std ESP	two-stage RESP
T t	**0.00**	0.43	**0.00**	0.17	**0.00**	0.16	**0.00**	0.68	0.11	0.40	0.03
T g	0.30	0.61	0.20	0.12	**0.00**	0.52	0.35	0.46	**0.00**	0.34	**0.00**
G g-	0.25	1.01	0.96	0.51	0.44	0.93	0.80	0.97	0.87	0.85	0.61
G t	0.32	**0.00**	0.11	**0.00**	0.32	**0.00**	0.33	**0.00**	0.09	**0.00**	0.20
G g	0.61	0.37	0.44	0.08	0.34	0.49	0.69	0.08	0.16	0.17	0.26
sum of abs.		2.06	1.19	1.46	0.76	1.50	0.69	2.41	1.71	1.80	1.16
errors:				(1.05) [f]		(0.71)		(1.50)		(1.45)	

[a] Energies calculated using 1-4 electrostatic scale factor of 1/1.2. [b] Conformations described by C-C-C-N and c-c-n-lp torsions. [c] MP3/6-31+G**/6-31G* energies in kcal/mol -- unpublished results from T.A. Halgren. [d] Charge model conformation is the one used to generate the charges. [e] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) models. [f] Errors given in parentheses are those obtained when equivalencing of hydrogens is forced during fit.

Discussion

The study of conformational energies of butane, methyl ethyl thioether, methanol, ethanol, propanol, methylamine, ethylamine, propylamine, and 1,2-ethanediol makes clear that two-stage RESP charges exhibit less conformational and 1-4 electrostatic scale factor dependence than do the standard ESP charges. Based on these calculations, we suggest the two-stage RESP fit charges with an electrostatic scale factor of 1/1.2 to be a particularly promising model.

The choice of a 1-4 scale factor (<1) for van der Waals interactions has had considerable justification given the known overestimate of short range repulsion by a 6-12 form of potential. Thus, the choice of a van der Waals scale factor of 1/2 for 1-4 interactions only, as used in the Weiner et al. [10] force field, seems reasonable and justifiable here as well. Weiner et al. justified the 1-4 electrostatic scale factor of 1/2 mainly on empirical results with the alanine dipeptide. Billeter et al. and Smith and Karplus have shown that such scaling can cause artifacts in conformational energies. We suggest that scaling 1-4 electrostatic interactions by 1/2 and leaving 1-5 interactions intact can unbalance the electrostatics of the system, leading to results such as the gauche conformation of butane being more stable than the trans. (On the other hand, none of the valence force fields include 1-3 interactions at all, but these interactions should remain fairly constant with rotation about a dihedral angle and thus should roughly cancel out between different conformations.)

As Table VII shows, using an electrostatic scale factor for 1/2 for 1,2-ethanediol and simple three-fold torsions results in conformational energies which are up to 7.5 kcal/mole in error for the ten minimum energy conformations. One could, in theory, fix those energies with a contribution from the torsional energy, but the torsional parameters would

have to be so large as to be physically quite unreasonable. Such torsional parameters would also not likely be transferable to that torsion in other molecules. Our study of 1,2-ethanediol suggests that a 1-4 electrostatic scale factor of 1/1.2 performs optimally and is a good compromise between full inclusion of 1-4 electrostatics and the overly severe scale factor of 1/2. The suitability of this scale factor is supported by the results given for the simple alcohols and amines.

Another difficulty associated with full inclusion of 1-4 electrostatic interactions using standard ESP charges (derived using the rather polar 6-31G* basis set) is that it leads to very large angle distortion energies in the exocyclic NH_2 groups in the nucleic acid bases. This angle distortion energy can be as great as 8 kcal/mole in the case of adenine! Using two-stage restrained ESP charges and the 1-4 electrostatic scale factor of 1/1.2 significantly reduces the problem, such that the angle distortion energy is ~1-2 kcal/mole. This is still ;arger than the ~0.5 kcal/mole angle distortion energy found with the Weiner *et al.* force field and a 1-4 electrostatic scale factor of 1/2, but further work on these systems will be needed to find the best approach to reduce the angle distortion energy.

The conformational dependence of ESP charges as revealed in Tables XIII and XIV is an issue which will be studied in more depth as we are developing our new force field. The intermolecular problem, that of reproducing dipole moments accurately for more than one conformation of a molecule, is one which is likely to be present in any effective two-body force field which does not allow for polarization. It is encouraging that the two-stage restrained ESP charges exhibit less inter- and intramolecular conformational dependence than do the standard ESP charges. It is clear from the results presented above that both equivalencing and restraining the charges is beneficial for the derivation of an optimum charge model. Furthermore, the use of multiple conformations of a molecule to derive RESP charges allows for further refinement of the model.

Why do we employ the two-stage RESP fit, rather than just using a one-stage calculation? As described in more detail in Bayly *et al.* there are two issues: first, the need to reduce spuriously large charges which are statistically poorly determined, and second, the desire to make equivalent those charges on atoms which are not necessarily equivalent during the SCF calculation, but which can interconvert during a molecular dynamics simulation (e.g. the methyl hydrogens in methanol.) Also of relevance is the use of the 6-31G* basis set for determining the electrostatic potential fit charges. The motivation for using this basis set is has been that it consistently overestimates the dipole moment by an amount (5-20%) consistent with the TIP3P/SPC "effective two-body" models for water. Thus, it fortuitously contains about the amount of "polarization" contained in such water models and should therefore be "balanced" with respect to those water models. The 6-31G* basis set is then expected to enhance the solute dipole moment over the actual gas phase value to about the same extent as seen in water models.

How does this work in practice? Methanol is a good example. The 6-31G* calculation on this molecule gives a dipole moment of 1.9 D, compared to the experimental gas phase moment of 1.7 D. The "raw" unrestrained and restrained ESP fit charges lead to a dipole moment of ~1.9 D, but when one averages the standard ESP methyl hydrogen charges after the fitting, the dipole moment increases to 2.15 D. The two-stage restrained ESP fitting of the CH_3 group only reduces the dipole moment to 2.14 D. If one forces equivalent charges on the CH_3 hydrogens in the initial electrostatic potential fit, one retains the 6-31G* dipole moment, but the charges on the hydroxyl oxygen and hydrogen, critical for hydrogen bonding, are significantly reduced. Trans NMA provides another example of the effects of making methyl hydrogens equivalent by a.p. averaging or forced equivalence during the fit. Whereas the 6-31G* quantum mechanical calculation gives a dipole moment of 4.2 D, compared to the gas phase experimental value of 3.7 D, *a posteriori* averaging of each of

the two methyl group hydrogens increases this to 4.6 D. Refitting in the second stage to make each set of methyl hydrogen charges internally equivalent decreases the dipole moment to 4.4 D.

It is therefore clear that the two-stage fit, by keeping the heteroatoms and hydrogen bonding hydrogen charges fixed, may lead to a dipole moment for the molecule which is larger than the dipole moment determined by the quantum mechanical wavefunction. In the case of methanol and NMA, it increases the enhancement in dipole moment to 10-20% over the gas phase value, more in line with the ~20% enhancement of TIP3P water over gas phase water. In the case of the nucleic acid bases, this enhancement is modest in percentage (increase of ~0.2D for adenine, cytosine, and thymine and no change for guanine) and these lead to essentially no change in hydrogen bond energies for the base pairs. Obviously, as noted before, the dipole moment is a useful first estimate for what the hydrogen bonding or solvation free energy of a model will be, but the larger the molecule, the larger role higher moments must play. Furthermore, although the dipole moment serves as a useful predictor of experimental solvation energies, it may not perform as well in a theoretical model using nonadditive potentials.

Why bother with electrostatic potential fit charges at all -- why not just use the empirical approaches embodied in TIP3P/OPLS models? In our opinion, the use of electrostatic potential fit charges allows us a general, unbiased, and more accurate representation of electrostatic charge distribution. This method is less subject to arbitrariness than empirically derived charges [52] and can easily be generalized to any molecule or functional group. Given current and ever increasing computer power, 6-31G* electrostatic potential charges can be derived for virtually any molecule, possibly in multiple conformations. Electrostatic potential fit charges or those based on distributed multipole analyses will be even more critical if one hopes to go beyond the atom centered monopole or empirical bond

dipole models for charge distribution, where even more degrees of freedom are being fit.

Conclusion

In this paper and a related one,[8] we have presented some new approaches to deriving electrostatic potential fit charges and have used these new approaches to study conformational energies, conformational dependencies, hydrogen bonding, and solvation free energies. It is clear that restraining these electrostatic potential charges has rather little effect on the quality of the fit to the potential and the calculated molecular properties and provides a somewhat better representation of conformational properties of molecules compared to the standard ESP model. The set of two-stage RESP charges thus derived give an excellent fit to the solvation free energy of methanol and an adequate fit to the solvation free energy of NMA. We have also further evaluated the multiple conformation ESP fitting studied by Reynolds *et al.* [36]. We have confirmed and extended their findings that such an approach is useful and suggest that the use of restraints and multiple molecule fitting will lead to an optimal set of charges for the broadest range of molecular systems.

The value of electrostatic potential derived (ESP) charges in modelling the important electrostatic interactions in biological systems has been known for some time. We have shown that two-stage RESP charges retain this excellent intermolecular behavior while exhibiting intramolecular behavior which makes them suitable for conformational analysis. Two-stage RESP charges thus reproduce both intermolecular and intramolecular energies and structures quite well, making this charge model a critical advancement in the development of a general force field for modelling biological macromolecules and their ligands, both in the gas phase and in solution.

Acknowledgements

References

(1) Momany, F.A. *J. Phys. Chem.* **1978**, *82*, 592.

(2) Cox, S.R.; Williams, D.E. *J. Comput. Chem.* **1981**, *2*, 304.

(3) Singh, U.C.; Kollman, P.A. *J. Comput. Chem.* **1984**, *5*, 129.

(4) Williams, D.E. *Biopolymers* **1990**, *29*, 1367.

(5) Stouch, T.R.; Williams, D.E. *J. Comput. Chem.* **1992**, *13*, 622.

(6) Reynolds, C.A.; Essex, J.W.; Richards, W.G. *Chem. Phys. Lett.* **1992**, *199*, 257.

(7) Cornell, W.D.; Kollman, P.A. unpublished results.

(8) Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. *J. Phys. Chem.*, submitted.

(9) Cieplak, P.; Cornell, W.D.; Bayly, C.I.; Kollman, P.A., work in progress.

(10) (a) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Jr.; Weiner, P. *J. Am. Chem.Soc.* **1984**, *106*, 765. (b) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230.

(11) Cornell, W.D., *et al.*, work in progress.

(12) Spellmeyer, D.C.; Kollman, P.A. unpublished results.

(13) Veenstra, D.L.; Ferguson, D.M.; Kollman, P.A. *J. Comput. Chem.* **1992**, *13*, 971.

(14) Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. *J. Chem. Phys.* **1983**, *79*, 926.

(15) Jorgensen, W.L.; Tirado-Rives, J. *J. Am. Chem.Soc.* **1988**, *110*, 1657.

(16) Pearlman, D.A.; Case, D.A.; Caldwell, J.C.; Seibel, G.L.; Singh, U.C.; Weiner, P.A.; Kollman, P.A. AMBER 4.0 (UCSF), 1991, Department of Pharmaceutical Chemistry, University of California, San Francisco.

(17) Frisch, M.J.; Head-Gordon, M.; Trucks, G.W.; Foresman, J.B.; Schlegel, H.B.; Raghavachari, K.; Robb, M.A.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seger. R.; Melius, C.F.; Baker, J.; Martin, L.R.; Kahn, L.R.; Stewart, J.J.P.; Topiol, S.; Pople, J.A. Gaussian, Inc., 1990, Pittsburgh, PA.

(18) Halgren, T.A., personal communication.

(19) Gaussian 90 and MM2 calculations carried out for this work.

(20) (a) Allinger, N.L. *J. Am. Chem.Soc.* **1977**, *99*, 8127. (b) Allinger, N.L.; Burkert, U. *Molecular Mechanics*, American Chemical Society Monograph 177, American Chemical Society: Washington, D.C., 1982. (c) MM2(87), QCPE, Bloomington, IN.

(21) Ferrin, T.E.; Huang, C.C.; Jarvis, L.E.; Langridge, R. *J. Mol. Graph.* **1988**, *6*, 13.

(22) Saenger, W. *Principles of Nucleic Acid Structure*, Springer-Verlag: Tokyo, 1984.

(23) (a) Jorgensen, W.L.; Swensen, C.J. *J. Am. Chem. Soc.* **1985**, *107*, 569. (b) Jorgensen, W.L.; Swensen, C.J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.

(24) Tse, Y.-C.; Newton, M.D. *J. Am. Chem. Soc.* **1976**, *99*, 611.

(25) van Gunsteren, W.F.; Berendsen, H.J.C. *J. Comp. Aided Mol. Des.* **1987**, *1*, 171.

(26) Pearlman, D.A.; Kollman, P.A., *J. Chem. Phys.* **1991**, *94*, 4532.

(27) Pearlman, D.A.; Kollman, P.A., *J. Chem. Phys.* **1989**, *90*, 2460.

(28) Sun, Y.X.; Spellmeyer, D.; Pearlman, D.A.; Kollman, P.A. *J. Am. Chem. Soc.* **1992**, *114*, 6798.

(29) (a) Beveridge, D.L.; Schnuelle, G.W. *J. Phys. Chem.* **1975**, *79*, 2562. (b) Aue, D.H.; Webb, H.M.; Bowers, M.T. *J. Am. Chem.Soc.* **1976**, *98*, 318.

(30) (a) van Gunsteren, W.F.; Berendsen, H.J.C. *Mol. Phys.* **1977**, *34*, 1311. (b) Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. *J. Comput. Phys.* **1977**, *23*, 327.

(31) Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(32) (a) Heenan, R.K.; Bartell, L.S. *J. Chem. Phys.* **1983**, *78*, 1270. (b) Bradford, W.F.; Fitzwater, S.; Bartell, L.S. *J. Mol. Struct.* **1977**, *38*, 185

(33) (a) Compton, D.A.C.; Montero, S.; Murphy, W.S. *J. Phys. Chem.* **1980**, *84*, 3587. (b) Allinger, N.L.; Grev, R.S.; Yates, B.F.; Schaefer, H.F., III *J. Am. Chem. Soc.* **1990**, *112*, 114.

(34) Billeter, M.; Howard, A.E.; Kuntz, I.D.; Kollman, P.A. *J. Am. Chem. Soc.* **1988**, *110*, 8385.

(35) Smith, J.C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *114*, 801.

(36) Reynolds, C.A.; Essex, J.W.; Richards, W.G *J. Am. Chem.Soc.* **1992**, *114*, 9075.

(37) (a) Sakakibara, M.; Matsuura, H.; Harada, I.; Shimanoguchi, T. *Bull. Chem. Soc. Jpn.* **1977**, *50*, 111. (b) Oyanagi, K.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 2243. (c) Durig, J.R.; Compton, D.A.C.; Jalilian, M.R. *J. Phys. Chem.* **1979**, *83*, 511.

(38) Allinger, N.L.; Rahman, M.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293.

(39) Lees, R.M.; Baker, J.G. *J. Chem. Phys.* **1968**, *48*, 5299.

(40) (a) Durig, J.R.; Bucy, W.E.; Wurrey, C.J.; Carriera, L.A. *J. Phys. Chem.* **1975**, *79*, 988. (b) Schaefer, L. *Theochem.* **1982**, *86*, 349; **1982**, *86*, 365.

(41) Durig, J.R.; Bucy, W.E.; Wurrey, C.J.; Carriera, L.A. *J. Phys. Chem.* **1975**, *79*, 988.

(42) Profeta, S.; Jr.; Allinger, N.L. *J. Am. Chem. Soc.* **1985**, *107*, 1907.

(43) Schmitz, L.R.; Allinger, N.L. *J. Am. Chem. Soc.* **1990**, *112*, 8307.

(44) Durig, J.R.; Craven, S.M.; Harris, W.C. *Vibrational Spectra and Structure* , Vol. 1, J.R. Durig, Ed., Marcel-Dekker, Inc: New York, 1973.

(45) (a) Fischer, E.; Botskor, I. *J. Mol. Spectrosc.* **1982**, *91*, 116. (b) Fischer, E.; Botskor, I. *J. Mol. Spectrosc.* **1984**, *104*, 226.

(46) Yanson, I.; Teplitsky, A.; Sukhodub, L. *Biopolymers* **1979**, *18*, 1149.

(47) Gould, I.R.; Kollman, P.A. submitted for publication.

(48) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016

(49) Wolfenden, R. *Biochemistry* **1978**, *17*, 201.

(50) Urban, J.J.; Famini, G.R. *J. Comput. Chem.* **1993**, *14*, 353.

(51) Cieplak, P.; Kollman, P.A. *J. Comput. Chem.* **1991**, *12*, 1232.

Supplementary Table I. Fitted standard and restrained ESP charges for butane. [a,b]

atom	standard ESP (un.ap)		one-stage RESP (wk.ap)		one-stage RESP (st.ap)		two-stage RESP (wk.fr/st.eq)	
	trans	gauche	trans	gauche	trans	gauche	trans	gauche
C (CH3)	-0.344	-0.197	-0.172	-0.100	-0.092	-0.061	-0.089	-0.093
H (CH3)	0.078	0.046	0.038	0.024	0.019	0.014	0.019	0.021
C (CH2)	0.161	0.094	0.063	0.034	0.029	0.016	0.025	0.055
H (CH2)	-0.025	-0.018	-0.003	-0.002	0.003	0.001	0.004	-0.012
RRMS	0.790	0.892	0.812	0.920	0.864	0.943	0.868	0.903
μ (D) [c]	0.000	0.097	0.000	0.103	0.000	0.103	0.000	0.056

[a] Charges given in atomic units. [b] 6-31G* optimized geometries used for charge calculations. [c] Dipole moments from 6-31G*//6-31G* wave functions are $\mu = 0.0$ D (trans) and $\mu = 0.08$ D (gauche).

Supplementary Table II. Fitted standard and restrained ESP charges for methyl ethyl thioether. [a,b]

atom	standard ESP (un.ap)		one-stage RESP (wk.ap)		one-stage RESP (st.ap)		two-stage RESP (wk.fr/st.eq)	
	trans	gauche	trans	gauche	trans	gauche	trans	gauche
C (CH3)	-0.256	-0.051	-0.139	-0.019	-0.083	-0.014	-0.042	-0.011
H (CH3)	0.086	0.024	0.057	0.017	0.042	0.017	0.031	0.011
C (CH2)	0.000	0.061	-0.034	0.015	-0.039	0.002	-0.017	0.043
H (CH2)	0.085	0.058	0.086	0.070	0.083	0.074	0.057	0.050
S	-0.270	-0.295	-0.283	-0.301	-0.294	-0.311	-0.283	-0.301
C(CH3)	-0.255	-0.311	-0.176	-0.228	-0.122	-0.159	-0.074	-0.090
H(CH3)	0.117	0.137	0.097	0.114	0.082	0.095	0.070	0.075
RRMS	0.355	0.367	0.343	0.366	0.343	0.367	0.311	0.345
μ(D) [c]	2.194	2.232	2.182	2.229	2.175	2.229	1.892	2.023

[a] Charges given in atomic units. [b] 6-31G* optimized geometries used for charge calculations. [c] Dipole moments from 6-31G*//6-31G* wave functions are μ = 1.80 D (trans) and μ = 1.84 D (gauche). Experimental dipole moment μ = 1.56 D (McClellan, A.L. *Tables of Experimental Dipole Moments*, Vols. 1-3; W.H. Freeman: San Francisco, 1963).

Supplementary Table III. Fitted standard and two-stage restrained ESP charges for the alcohols. [a,b]

atom	methanol [c] standard ESP (un.ap)	methanol [c] two-stage RESP (wk.fr/st.eq)	ethanol [d] standard ESP (un.ap)	ethanol [d] two-stage RESP (wk.fr/st.eq)	propanol [e] standard ESP (un.ap)	propanol [e] two-stage RESP (wk.fr/st.eq)
O	-0.672	-0.654	-0.702	-0.675	-0.746	-0.709
H	0.426	0.423	0.414	0.414	0.4439	0.432
$C\alpha$	0.195	0.126	0.408	0.326	0.281	0.211
$H\alpha$	0.017	0.035	-0.039	-0.035	-0.018	0.006
$C\beta$			-0.283	-0.092	0.172	0.036
$H\beta$			0.080	0.032	0.012	0.042
$C\gamma$					-0.457	-0.208
$H\gamma$					0.106	0.047
RRMS	0.212	0.210	0.165	0.153	0.112	0.120
$\mu(D)$ [f]	2.197	2.184	1.968	1.841	1.768	1.712

[a] Charges given in atomic units. [b] MM2 minimized geometries used for charge calculations. [c] Methanol - staggered. [d] Ethanol - all staggered with hydroxyl H trans to $C\beta$. [e] Propylamine - all staggered with O trans to $C\gamma$ and hydroxyl H trans to $C\beta$. [f] Dipole moments from 6-31G*/MM2 wave functions are $\mu = 1.94$ D for methanol, $\mu = 1.79$ D for ethanol, and $\mu = 1.72$ D for propanol. Experimental dipole moments are $\mu = 1.66$ D for methanol, $\mu = 1.44$ D for trans ethanol, and $\mu = 1.55$ D for propanol (McClellan, A.L. *Tables of Experimental Dipole Moments*, Vols. 1-3; W.H. Freeman: San Francisco, 1963).

Supplementary Table IV. Fitted standard and two-stage restrained ESP charges for the amines. [a,b]

atom	methylamine [c]		ethylamine [d]		propylamine [e]	
	standard ESP (un.ap)	two-stage RESP (wk.fr/st.eq)	standard ESP (un.ap)	two-stage RESP (wk.fr/st.eq)	standard ESP (un.ap)	two-stage RESP (wk.fr/st.eq)
N	-0.999	-0.969	-1.080	-1.038	-1.071	-1.044
H	0.371	0.367	0.383	0.376	0.370	0.370
Cα	0.314	0.211	0.499	0.400	0.431	0.336
Hα	-0.019	0.008	-0.021	-0.013	-0.019	0.008
Cβ			-0.370	-0.154	0.088	0.011
Hβ			0.075	0.022	-0.022	0.0010
Cγ					-0.292	-0.107
Hγ					0.061	0.016
RRMS	0.293	0.292	0.224	0.223	0.211	0.207
μ (D) [f]	1.819	1.803	1.476	1.457	1.382	1.478

[a] Charges given in atomic units. [b] MM2 minimized geometries used for charge calculations. [c] Methylamine - staggered. [d] Ethylamine - all staggered with lp trans to Cβ. [e] Propyamine - all staggered with N trans to Cγ and lp trans to Cβ. [f] Dipole moments from 6-31G*//MM2 wave functions are μ = 1.57 D for methylamine, μ = 1.48 D for ethylamine, and μ = 1.43 D for propylamine. Experimental dipole moments are μ = 1.30 D for methylamine, μ = 1.22 D for ethylamine, and μ = 1.35 D for propylamine (McClellan, A.L. *Tables of Experimental Dipole Moments*, Vols. 1-3; W.H. Freeman: San Francisco, 1963).

Supplementary Table V. Fitted standard and restrained two-stage ESP charges for 1,2-ethane diol. [a,b]

atom	standard ESP (un.ap)	two-stage RESP (wk.fr/st.eq)
C	0.311	0.242
H (CH2)	-0.002	0.023
O	-0.750	-0.724
H	0.443	0.435
RRMS	0.090	0.094
μ(D) [c]	0.034	0.034

[a] Charges given in atomic units. [b] Charges derived from 6-31G* optimized all-trans conformation. [c] Dipole moment from 6-31G*//6-31G* wave function is $\mu = 0.00$ D.

Supplementary Table VI. Fitted standard and two-stage restrained ESP charges for the DNA bases. [a,b,c]

Adenine			Guanine			Thymine			Cytosine		
atom	standard RESP (un.ap)	two-stage ESP (wk.fr/st.eq)	atom	standard RESP (un.ap)	two-stage ESP (wk.fr/st.eq)	atom	standard RESP (un.ap)	two-stage ESP (wk.fr/st.eq)	atom	standard RESP (un.ap)	two-stage ESP (wk.fr/st.eq)
C1'	-0.196	-0.130	C1'	-0.363	-0.203	C1'	-0.352	-0.083	C1'	-0.338	-0.279
H1'	0.113	0.091	H1'	0.151	0.105	H1'	0.151	0.074	H1'	0.143	0.119
N9	-0.132	-0.057	N9	-0.010	-0.010	N1	-0.107	-0.003	N1	-0.218	-0.041
C8	0.198	0.091	C8	0.081	0.017	C6	-0.118	-0.198	C6	0.187	-0.020
H8	0.151	0.178	H8	0.170	0.191	H6	0.213	0.201	H6	0.164	0.192
N7	-0.583	-0.564	N7	-0.498	-0.513	C5	-0.099	-0.008	C5	-0.802	-0.536
C5	-0.096	0.076	C5	-0.161	0.104	C7	-0.370	-0.157	H5	0.257	0.197
C6	0.892	0.645	C6	0.812	0.472	H7	0.122	0.063	C4	1.177	0.890
N6	-1.062	-0.857	O6	-0.604	-0.540	C4	0.791	0.500	N4	-1.187	-1.025
HN6	0.456	0.395	N1	-0.874	-0.429	O4	-0.594	-0.530	HN4	0.481	0.440
N1	-0.853	-0.749	H1	0.426	0.328	N3	-0.794	-0.418	N3	-0.953	-0.792
C2	0.607	0.548	C2	1.086	0.619	H3	0.421	0.334	C2	0.978	0.808
H2	0.073	0.072	N2	-1.144	-0.905	C2	0.812	0.505	O2	-0.656	-0.631
N3	-0.792	-0.717	HN2	0.480	0.424	O2	-0.622	-0.554			
C4	0.542	0.401	N3	-0.845	-0.635						
			C4	0.511	0.341						
RRMS	0.116	0.118		0.048	0.058		0.079	0.097		0.060	0.066
μ (D) [d]	2.685	2.652		7.589	7.586		5.291	5.246		7.116	7.064

[a] Charges given in atomic units. [b] Standard AMBER database geometries used for charge calculations. [c] Atom numbering as in Saenger, ref. 22.
[d] Dipole moments from 6-31G* wave functions are μ = 2.48 D (adenine), μ = 7.63 D (guanine), μ = 4.98 D (thymine), and μ = 6.82 D (cytosine).

Supplementary Table VII A. Fitted standard and restrained ESP charges for trans-NMA. [a,b]

Atom	standard ESP (un.ap)	one-stage RESP (wk.eq)	one-stage RESP (wk.ap)	two-stage RESP (wk.fr/st.eq)
C (CO)	0.761	0.632	0.587	0.587
O	-0.626	-0.568	-0.591	-0.591
C (CH3CO)	-0.490	-0.318	-0.236	-0.041
H (CH3CO)	0.132	0.092	0.073	0.017
N	-0.537	-0.468	-0.419	-0.419
H (NH)	0.321	0.278	0.282	0.282
C (CH3NH)	-0.049	-0.018	-0.042	-0.208
H (CH3NH)	0.075	0.062	0.067	0.113
RRMS	0.122	0.091	0.131	0.117
μ (D) [c,d]	4.572	4.198	4.589	4.424

[a] Charges given in atomic units. [b] AMBER (Weiner, all atom) optimized geometry used for charge calculations. [c] Dipole moment from 6-31G*//AMBER wave function is μ = 4.17 D. [d] Experimental dipole moment μ = 3.68 D (McClellan, A.L. *Tables of Experimental Dipole Moments*, Vols. 1-3; W.H. Freeman: San Francisco, 1963).

Supplementary Table VII B. Fitted standard and restrained ESP charges for methanol. [a,b]

Atom	standard ESP (un.ap)	one-stage RESP (wk.eq)	one-stage RESP (wk.ap)	two-stage RESP (wk.fr/st.eq)
C (CH3)	0.196	0.055	0.125	0.117
H (CH3)	0.016	0.050	0.034	0.037
O	-0.668	-0.592	-0.650	-0.650
H (HO)	0.423	0.386	0.422	0.422
RRMS	0.172	0.149	0.167	0.167
μ (D) [c]	2.151	1.992	2.132	2.139

[a] Charges given in atomic units. [b] Experimental geometry of Lees and Baker, ref. 40, used for charge calculation. [c] Dipole moment from 6-31G* wave function is $\mu = 1.92$ D.

Supplementary Table VIII. Fitted standard and two-stage restrained ESP charges for five different conformations of propylamine.[a,b]

	T t	T t	T g	T g	G g-	G g-	G t	G t	G g	G g
	std	two-stage	std	two-stage	std	two-stage	std	two-stage	std	two-stage
atom	ESP	RESP	ESP	RESP	ESP	RESP	ESP	RESP	ESP	RESP
N	-1.071	-1.044	-1.099	-1.045	-1.011	-0.977	-1.056	-1.016	-1.007	-0.970
H	0.370	0.370	0.388	0.380	0.373	0.369	0.389	0.381	0.372	0.366
Cα	0.431	0.336	0.362	0.247	0.301	0.229	0.343	0.218	0.257	0.276
Hα	-0.019	0.008	-0.022	0.012	-0.014	-0.003	0.002	0.032	-0.005	-0.019
Cβ	0.088	0.011	0.114	0.004	-0.014	0.055	-0.046	-0.000	0.024	-0.035
Hβ	-0.022	0.001	0.016	0.039	0.006	-0.029	0.002	-0.010	0.005	0.033
Cγ	-0.292	-0.107	-0.452	-0.185	-0.073	0.046	-0.095	-0.022	-0.112	-0.128
Hγ	0.061	0.016	0.104	0.039	0.022	-0.009	0.023	0.004	0.031	0.032

(charge model [c,d,e])

[a] Charges given in atomic units. [b] MM2 minimized geometries used for charge calculations. [c] Charge model conformation is the one used to generate the charges. [d] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) charge models. [e] Conformations described by C-C-C-N and C-C-N-lp torsions.

Supplementary Table IX. Fitted standard and two-stage restrained ESP charges for multiple conformation fits of propylamine.[a,b]

	charge model [c,d,e]									
	T t	T t	multi (5)	multi (5)	multi (2) Tt,Gt	multi (2) Tt,Gt	multi (2) Tt,Tg	multi (2) Tt,Tg	multi (2) Tg,Gt	multi (2) Tg,Gt
	std	two-stage	std	two-stage	std	two-stage	std	two-stage	std	two-stage
atom	ESP	RESP	ESP	RESP	ESP	RESP	ESP	RESP	ESP	RESP
N	-1.071	-1.044	-1.046	-1.010	-1.062	-1.022	-1.080	-1.044	-1.063	-1.014
H	0.370	0.370	0.378	0.374	0.379	0.372	0.378	0.375	0.385	0.375
Cα	0.431	0.336	0.335	0.303	0.386	0.316	0.394	0.293	0.344	0.281
Hα	-0.019	0.008	-0.012	-0.013	-0.008	0.001	-0.018	0.010	-0.008	-0.001
Cβ	0.088	0.011	0.034	0.015	0.016	0.016	0.084	-0.002	0.018	0.004
Hβ	-0.022	0.001	-0.001	-0.006	-0.008	-0.016	0.000	0.022	0.012	0.007
Cγ	-0.292	-0.107	-0.195	-0.031	-0.193	-0.034	-0.357	-0.147	-0.254	-0.074
Hγ	0.061	0.016	0.046	0.005	0.042	0.003	0.080	0.029	0.059	0.014

[a] Charges given in atomic units. [b] MM2 minimized geometries used for charge calculations. [c] Charge model conformation is the one used to generate the charges. [d] "std ESP" and "two-stage RESP" refer to (un.ap) and (wk.fr/st.eq) charge models. [e] Conformations described by C-C-C-N and C-C-C-N-lp torsions.

Chapter 3

The Effects of Basis Set and Methyl Blocking Groups on the
Conformational Energies of Glycyl and Alanyl Dipeptides:
A Hartree-Fock and MP2 Study

The Effects of Basis Set and Methyl Blocking Groups on the Conformational Energies of Glycyl and Alanyl Dipeptides: A Hartree-Fock and MP2 Study

by

Wendy D. Cornell [1]

Ian R. Gould [2]

and

Peter A. Kollman [1]

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

[1] Graduate Group in Biophysics

[2] Department of Chemistry, University of Manchester, Manchester, M13 9PL, UK

Abstract

We present the results of high level ab initio molecular orbital calculations on glycyl and alanyl dipeptides. We have previously reported the results of calculations on the low energy conformers of the methyl-blocked analogs at the MP2/TZVP//HF/6-31G** level of theory. In this paper, we examine the effect of carrying out the geometry optimizations of the three methyl-blocked glycyl dipeptide conformers using the larger TZVP basis set followed by an MP2 calculation with that basis set. The resulting geometries and energies were essentially the same as those obtained from optimization with the smaller 6-31G** basis set followed by an MP2 calculation with the TZVP basis set. We also carried out MP2/TZVP//HF/6-31G** calculations on the hydrogen-blocked analogs of both dipeptides, so that we might make a more direct comparison with energies calculated by Head-Gordon et al. (Head-Gordon, T.; Head-Gordon, M.; Frisch, M.J.; Brooks, C.L.; Pople, J.A. J. Am. Chem. Soc. 1991, 113, 5989.) The results reveal that the source of the different energies seen for the alanyl dipeptide C5 conformation in the two original studies was the presence or absence of the blocking methyl groups, rather than the quantum mechanical protocols.

Introduction

This laboratory has previously published the results of high level quantum mechanical calculations on four low energy conformations of the alanyl dipeptide.[1] Two of us (IRG and WDC) have also carried out the corresponding calculations on three low energy conformations of the glycyl dipeptide.[2] As the conformational energies so obtained are currently inaccessible by experiment, such theoretical calculations are critical to the development and parameterization of a molecular mechanical force field for proteins.[3]

We have chosen to carry out our quantum mechanical calculations on analogs of the glycyl and alanyl dipeptides which have a methyl blocking group attached to each of the two

amide groups in the molecule. These methyl groups can be thought to mimic the presence of the alpha-carbons in the preceding and following residues in a larger peptide or protein. Another high level quantum mechanical study of these two dipeptides by Head-Gordon *et al.*[4] sought to map out the entire phi-psi map of each at the Hartree-Fock level and also included correlated calculations of the stationary points. This study differed from ours in the choice of both the molecular model used and the particular theoretical treatment. The Head-Gordon study used a simpler analog of the dipeptides which had hydrogens rather than methyls for blocking groups. Also, their study involved geometry optimization using the 6-31+G* basis set[5] followed by an MP2[6] calculation with the 6-31+G** basis set.[5] Our calculations, however, involved geometry optimization using the 6-31G** basis set[5] followed by an MP2 calculation with Dunning's triple ζ plus valence polarization (TZVP) [7] basis set.

A further difference between the two studies was that Head-Gordon *et al.*[4] limited their correlated calculations to stationary points whereas we limited our calculations to the three internally hydrogen bonded conformations for alanyl dipeptide, the two hydrogen bonded conformations for glycyl dipeptide, and the alpha-helical conformation for each. While the alpha-helical conformation is not hydrogen bonded at the dipeptide level or particularly low in energy, its presence is favored in proteins where it can hydrogen bond with the residue which is four residues down the chain.

We are thus restricted to comparing our results to those of Head-Gordon *et al.*[4] for the three internally hydrogen bonded conformations of alanyl dipeptide and the two of glycyl dipeptide. For glycyl dipeptide, these conformations are referred to as C7 and C5, respectively, depending on whether the hydrogen bonding results in a five- or seven-membered ring (Figure 1). The seven-membered ring results from the N-terminal carboxyl oxygen hydrogen bonding with the C-terminal amide hydrogen. The five-membered ring

results from the N-terminal amide hydrogen bonding with the C-terminal carboxyl oxygen. In the case of alanyl dipeptide, the C7 conformation can be formed with the methyl sidechain either in the axial or equatorial position. The hydrogen bonded alanyl dipeptide conformations are thus described with the designations C7eq, C7ax, and C5 (Figure 2).

The results obtained by both our group and Head-Gordon *et al.* for the relative conformational energies of C7eq, C7ax, and C5 alanyl dipeptide were quite similar. The C5 conformations differed by 0.34 kcal/mole, but this is well within the accuracy of such calculations. The glycyl dipeptide results, however, differed by 0.9 kcal/mole in their prediction of the C7/C5 energy difference. Furthermore, Head-Gordon *et al.*[4] found the C5 conformations for both alanyl and glycyl dipeptides to be 1.1 kcal/mole above the minimum energy C7 conformations. In our study, however, the alanyl dipeptide C5 conformation was 1.5 kcal/mole above the minimum whereas the glycyl dipeptide C5 was 2.0 kcal/mole above the mimimum.

These differences were of some concern, so we decided to carry out further calculations investigating the effects of different quantum mechanical protocols and model systems. We first present the results of carrying out geometry optimization on the three methyl-blocked glycyl dipeptide conformers using the TZVP basis set followed by an MP2 calculation using that same basis set. Next, we present the results of carrying out geometry optimizations at the HF/6-31G** level followed by single point MP2/TZVP calculations on the hydrogen-blocked analog of both glycyl and alanyl dipeptides. The results reveal that the choice of chemical model, rather than quantum mechanical protocol, was the source of the differences seen between the two original sets of calculations by Head-Gordon *et al.*and this group.
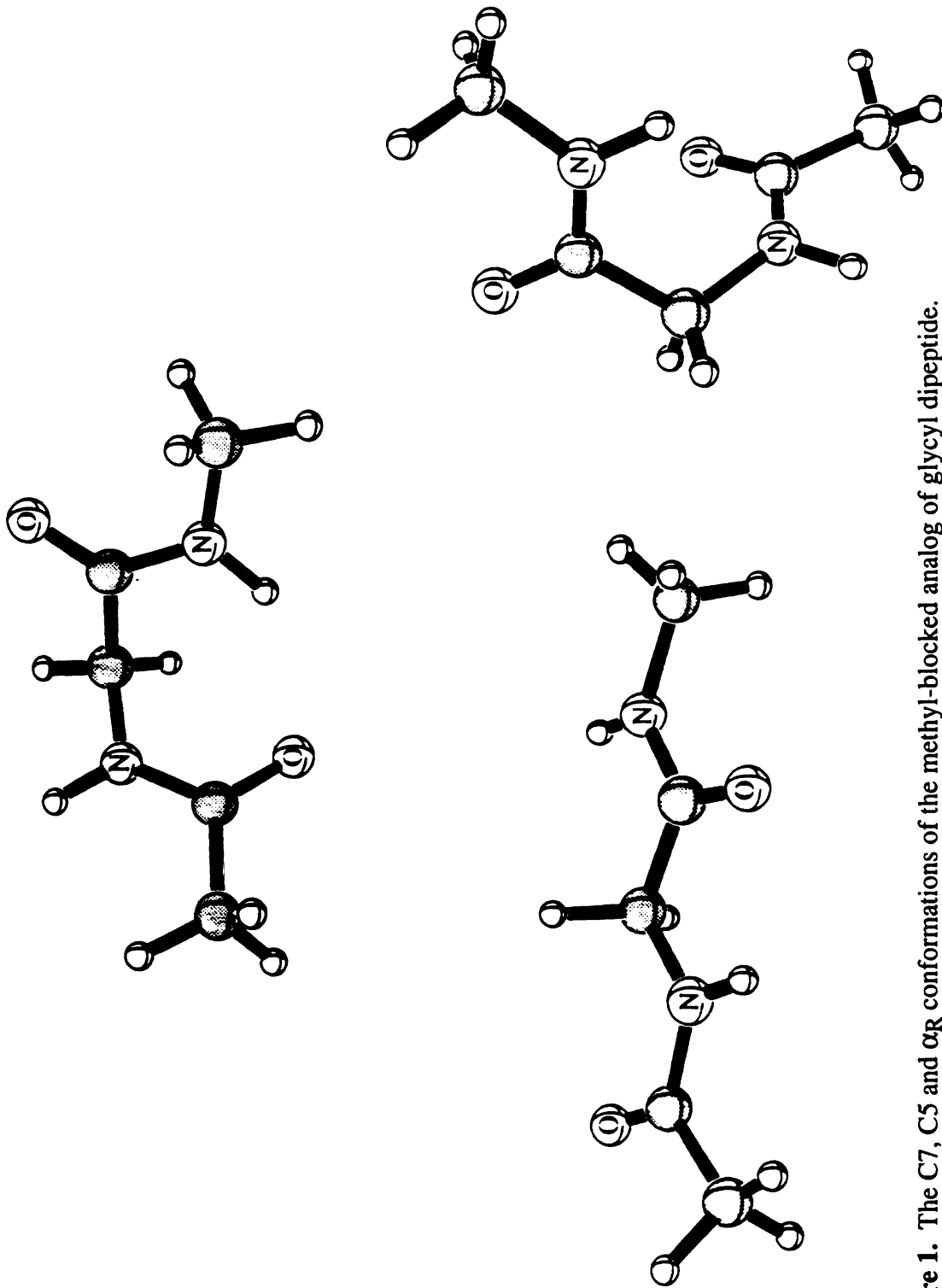
Figure 1. The C7, C5 and α_R conformations of the methyl-blocked analog of glycyl dipeptide.
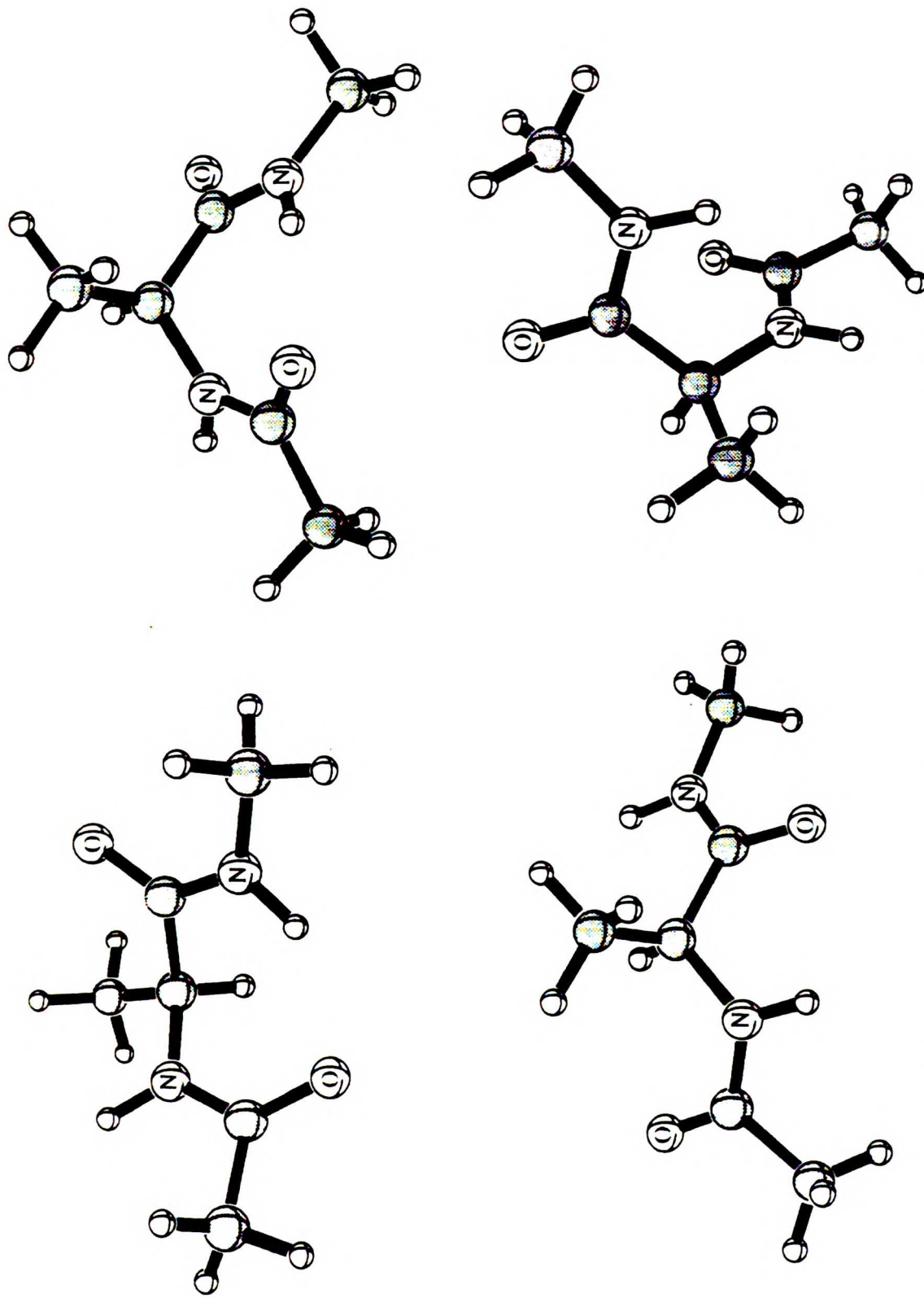
Figure 2. The C7eq, C7ax, C5 and α_R conformations of the methyl-blocked analog of alanyl dipeptide.

Methods

We have performed all of our calculations using either the Gaussian 90 or Gaussian 92 packages.[8] The geometry optimizations were carried out on a VAX 8600 and the MP2 and frequency calculations were carried out on a CRAY YMP and a CRAY C90. The HF/TZVP geometry optimizations reported here for methyl-blocked glycyl dipeptide were started from the 6-31G** optimized geometries. The HF/6-31G** optimizations of the hydrogen-blocked glycyl and alanyl dipeptides were started from the HF/6-31G** geometries of the methyl-blocked analogs by substituting a hydrogen for each blocking methyl group and changing only the bond distance and not the internal valence or dihedral angles. As in previous studies, the alpha-helical optimizations were carried out with the Φ and Ψ dihedrals constrained to be -60.7° and -40.7°, respectively. Frequency calculations were carried out on each optimized structure to determine its character (minimum or stationary point).

Results

I. Effect of Basis Set used in Geometry Optimization

The first aspect of our previous calculations which we chose to investigate was the use of a basis set from one family (6-31G**) for the geometry optimization and a basis set from a different family (Dunning's TZVP) for the single point MP2 calculations. We tested the validity of this particular protocol by carrying out geometry optimizations on the methyl-blocked glycyl dipeptide conformers using the TZVP basis set, which was used for the MP2 calculations in our original studies. Table I presents the optimized Φ and Ψ values and energies obtained from optimizations of the methyl-blocked glycyl dipeptide conformers using the 6-31G** and the TZVP basis sets. The deviations in optimized dihedral values range from 0.5° to 2.1°. More importantly, the HF/TZVP and MP2/TZVP energies of the two sets of optimized structures are nearly identical. From this we conclude

Table I. Comparison of energies obtained for Hartree-Fock optimized methyl-blocked glycyl dipeptide (N-acetyl-N'-methylglycinamide) conformers using 1) the 6-31G** and 2) Dunning TZVP basis sets for the geometry optimizations. [a]

	Geometry Optimization					
	HF/6-31G**			HF/TZVP		
	C7	C5	α_R	C7	C5	α_R
Φ	-85.5°	-179.1°	-60.7°	-86.0°	180.0°	-60.7°
Ψ	72.0°	-179.5°	-40.7°	74.1°	180.0°	-40.7°
E(HF/6-31G**//HF/6-31G**)[b]	0.27	0.00	4.30			
E(HF/TZVP//HF/6-31G**)[c]	0.49	0.00	4.60			
E(HF/TZVP//HF/TZVP)[d]				0.50	0.00	4.59
E(MP2/TZVP//HF/6-31G**)[e]	0.00	1.99	3.95			
E(MP2/TZVP//TZVP)[f]				0.00	1.91	3.96

[a] Energies in kcal/mol. [b] Zero of energy is -453.8447709 hartrees. [c] Zero of energy is -453.9780313 hartrees. [d] Zero of energy is -453.9774196 hartrees. [e] Zero of energy is -455.4347247 hartrees. [f] Zero of energy is -455.4336892 hartrees.

that it is sufficient to carry out the geometry optimizations with the smaller 6-31G** basis set rather than the TZVP one (210 vs. 272 basis functions for alanyl dipeptide). This is useful information as such optimizations can be very time consuming. For example, the frequency calculations for the C5 conformation of the methyl-blocked glycyl dipeptide required 62 minutes of Cray C90 CPU time using the 6-31G** basis set and 267 minutes using the TZVP basis set.

In Table II we present the structural parameters from the 6-31G** and TZVP optimizations. The numbering of the atoms is given in ref. 1. As noted above, the Φ and Ψ dihedral values are quite similar. The distance between the hydrogen bonding atoms (H---O) varies from 2.20 Å for the 6-31G** optimized structure to 2.24 Å for the TZVP optimized structure in the C7 conformation. However, the distance between the heavy atoms (N---O) is also reduced in the 6-31G** optimized structure, apparently offsetting the stronger attractive interaction. The dihedral angles for the peptide bonds ($\omega 1$ and $\omega 2$) were within 1.5° between corresponding structures, with the exception of the C-terminal peptide bond ($\omega 2$) in the C5 conformation. In that case the 6-31G** optimized structure had a value of 175.7° whereas the TZVP optimized structure was 180.0°. Most of the valence bond and angle structural parameters are quite similar between the two sets of structures, with the greatest variation seen with carbonyl bond lengths.

It is worth noting that the TZVP optimized C5 geometry has higher symmetry than the 6-31G** optimized one. The TZVP optimized C5 structure has a plane of symmetry that includes the backbone atoms and one of the hydrogens in each of the blocking groups, thus presenting C_s symmetry. Because this geometry optimization was started from the 6-31G** optimized structure, which has lower symmetry, the higher symmetry seen in the TZVP optimized structure is a result of the basis set and was not "built into" the optimization by the starting geometry.

Table II. Comparison of structural parameters obtained for Hartree-Fock optimized methyl-blocked glycyl dipeptide (N-acetyl-N'-methylglycinamide) conformers using 1) the 6-31G** and 2) Dunning TZVP basis sets.

Parameter[a]	C7 Geometry Optimization		C5 Geometry Optimization		α_R[b] Geometry Optimization	
	HF/6-31G**	HF/TZVP	HF/6-31G**	HF/TZVP	HF/6-31G**	HF/TZVP
Φ (C6-N7-C9-C12)	-85.5	-86.0	-179.1	180.0	-60.7	-60.7
Ψ (N7-C9-C12-N17)	72.0	74.1	-179.5	180.0	-40.7	-40.7
R_H---O	2.20	2.24	2.20	2.22	----	----
$R_{N(H)}$---O	3.04	3.06	2.65	2.66	----	----
ω1 (O5-C6-N7-H8)	185.3	184.6	180.3	180.0	171.4	172.9
ω2 (O13-C12-N17-H18)	175.2	174.8	175.7	180.0	193.0	193.1
H(1-3)-C4-C6-O5	68/189/311	54/176/297	69/190/312	58/180/302	23/144/264	32/154/275
H(14-16)-C19-N17-H18	17/137/257	12/132/252	21/140/260	0/120/240	83/203/322	88/208/327
R_{C4-C6}	1.511	1.510	1.512	1.512	1.512	1.512
R_{C6-N7}	1.348	1.348	1.347	1.346	1.362	1.360
R_{C6-O5}	1.206	1.202	1.203	1.200	1.197	1.193
R_{N7-C9}	1.450	1.450	1.433	1.435	1.449	1.449
R_{C9-C12}	1.527	1.528	1.520	1.520	1.525	1.527
$R_{C12-O13}$	1.203	1.199	1.203	1.199	1.198	1.193
$R_{C12-N17}$	1.343	1.341	1.345	1.343	1.351	1.350
$R_{N17-C19}$	1.445	1.447	1.448	1.449	1.448	1.449
R_{N7-H8}	0.992	0.990	0.994	0.992	0.993	0.990
$R_{N17-H18}$	0.996	0.993	0.993	0.990	0.993	0.991
$\Theta_{C4-C6-N7}$	116.4	116.4	116.1	116.1	115.5	115.6

Table II. Comparison of structural parameters obtained for Hartree-Fock optimized methyl-blocked glycyl dipeptide (N-acetyl-N'-methylglycinamide) conformers using 1) the 6-31G** and 2) Dunning TZVP basis sets, cont'd

Parameter [a]	C7		C5		α_R [b]	
	Geometry Optimization		Geometry Optimization		Geometry Optimization	
	HF/6-31G**	HF/TZVP	HF/6-31G**	HF/TZVP	HF/6-31G**	HF/TZVP
$\Theta_{C6-N7-C9}$	122.5	122.7	121.2	121.4	121.3	122.0
$\Theta_{N7-C9-C12}$	113.0	112.8	109.2	109.4	115.4	115.8
$\Theta_{C9-C12-N17}$	115.2	115.3	114.9	114.9	116.1	116.1
$\Theta_{C12-N17-C19}$	121.3	121.6	121.7	121.8	120.4	120.6

[a] Bond lengths in Å. Angles in degrees. [b] Φ and Ψ constrained.

II.   Effect of Chemical Model

Having confirmed that our quantum mechanical protocol was sound, we next set out to investigate the effect of the chemical model on the calculated conformational energies. That is, how much of an effect did the methyl blocking groups have compared to using only hydrogen atoms? We created H-blocked analogs for both glycyl and alanyl dipeptides by taking the 6-31G** optimized conformers and substituting a hydrogen for the methyl group and making the new C-H bond suitably smaller. All seven conformers were then optimized using the 6-31G** basis set. The optimized Φ and Ψ values and energies obtained for glycyl dipeptide are presented in Table III The corresponding data is also shown for the methyl-blocked analogs using the same quantum mechanical protocol as well as Head-Gordon's[4] data on hydrogen-blocked analogs using their different quantum mechanical protocol. We are limited to comparing the C7/C5 energy difference, however it is striking that using the H-blocked analog and our QM protocol, we get nearly the same energy difference as was obtained by Head-Gordon et al.[4] with their protocol. Also, while all three sets of calculations report similar Φ values for the C7 conformations, the Ψ values are similar between the two H-analog calculations and differ from the value obtained from the methyl-analog calculation.

Table IV compares the structural parameters obtained with our QM protocol for the methyl- and hydrogen-blocked glycyl dipeptide conformers. Even though the same basis set was used, there is an even greater variation in carbonyl bond length than was seen with the same analog but different basis sets. The hydrogen bond distances are nearly identical, but the ω values differ by up to 5°.

In Tables V and VI we present the corresponding data for the alanyl dipeptide analogs. The MP2 energies for the C7eq, C7ax, and C5 conformations are essentially identical for the

Table III. Comparison of energies obtained for Hartree-Fock optimized H- and methyl-blocked glycyl dipeptides (N-formyl-glycinamide and N-acetyl-N'-methylglycinamide) conformers using the same quantum mechanical protocol. [a]

	methyl-analog			H-analog (this work)				H-analog (Head-Gordon et al.)[k]	
	C7	C5	α_R	C7	C5	α_R		C7	C5
Φ	-85.5°	-179.1°	-60.7°	-85.2°	180.1°	-60.7°	Φ	-85.2°	-180.0°
Ψ	72.0°	-179.5°	-40.7°	65.1°	179.9°	-40.7°	Ψ	67.4°	180.0°
E(HF/6-31G**// HF/6-31G**)[b,c]	0.27	0.00	4.30	0.51	0.00	4.40	E(HF/6-31+G*// HF/6-31+G*)[h]	0.58	0.00
E(HF/TZVP// HF/6-31G**)[d,e]	0.49	0.00	4.60	0.77	0.00	4.63	E(HF/6-31+G**// HF/6-31+G*)[i]	0.60	0.00
E(MP2/TZVP// HF/6-31G**)[f,g]	0.00	1.99	3.95	0.00	1.29	3.44	E(MP2/6-31+G**// HF/6-31+G*)[j]	0.00	1.11

[a] Energies in kcal/mol. [b] Zero of energy is -453.8443341 hartrees (methyl-analog). [c] Zero of energy is -375.7656187 hartrees (H-analog). [d] Zero of energy is -453.9780313 hartrees (methyl-analog). [e] Zero of energy is -375.8826322 hartrees (H-analog). [f] Zero of energy is -455.4347247 hartrees (methyl-analog). [g] Zero of energy is -377.0397833 hartrees (H-analog). [h] Zero of energy is -375.7622992 hartrees. [i] Zero of energy is -375.7790576 hartrees. [j] Zero of energy is -376.8782777 hartrees. [k] Reference 4.

Table IV. Comparison of structural parameters obtained for HF/6-31G** optimized conformers of glycyl dipeptide analogs: N-acetyl-N'-methylglycinamide (methyl-analog) vs. N-formyl-glycinamide (H-analog)

Parameter[a]	C7		C5		α_R[b]	
	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog
Φ (C6-N7-C9-C12)	-85.5	-85.2	-179.1	180.1	-60.7	-60.7
Ψ (N7-C9-C12-N17)	72.0	65.1	-179.5	179.9	-40.7	-40.7
R_{H---O}	2.20	2.21	2.20	2.21	---	---
R_{N---O}	3.04	3.04	2.65	2.65	---	---
ω1 (O5-C6-N7-C9)	185.3	180.2	180.3	180.0	171.4	172.3
ω2 (O13-C12-N17-H18)	175.2	180.5	175.7	180.0	193.0	197.1
H(1-3)-C4-C6-O5	68/193/311	---	69/190/312	---	23/144/264	---
H(14-16)-C19-N17-H18	17/137/257	---	21/140/260	---	83/203/322	---
R_{C4-C6}	1.511	---	1.512	---	1.512	---
R_{C6-N7}	1.348	1.342	1.347	1.342	1.362	1.354
R_{C6-O5}	1.206	1.200	1.203	1.198	1.197	1.192
R_{N7-C9}	1.450	1.451	1.433	1.434	1.449	1.449
R_{C9-C12}	1.527	1.527	1.520	1.518	1.525	1.524
$R_{C12-O13}$	1.203	1.200	1.203	1.200	1.198	1.195
$R_{C12-N17}$	1.343	1.346	1.345	1.347	1.351	1.357
$R_{N17-C19}$	1.445	---	1.448	---	1.448	---
R_{N7-H8}	0.992	0.993	0.994	0.995	0.993	0.993
$R_{N17-H18}$	0.996	0.995	0.993	0.992	0.993	0.992
$\Theta_{C4-C6-N7}$	116.4	---	116.1	---	115.5	---

Table IV. Comparison of structural parameters obtained for HF/6-31G** optimized conformers of glycyl dipeptide analogs: N-acetyl-N'-methylglycinamide (methyl-analog) vs. N-formyl-glycinamide (H-analog), cont'd

Parameter [a]	C7		C5		α_R [b]	
	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog
$\Theta_{C6-N7-C9}$	122.5	122.9	121.2	121.5	121.3	121.7
$\Theta_{N7-C9-C12}$	113.0	113.5	109.2	109.2	115.4	115.3
$\Theta_{C9-C12-N17}$	115.2	115.1	114.9	114.9	116.1	116.1
$\Theta_{C12-N17-C19}$	121.3	---	121.7	---	120.4	---

[a] Bond lengths in Å. Angles in degrees. [b] Φ and Ψ constrained.

Table V. Comparison of energies obtained for Hartree-Fock optimized H- and methyl-blocked alanyl dipeptides (N-formyl-alaninamide and N-acetyl-N'-methylalaninamide) conformers using the same quantum mechanical protocol. [a]

	methyl-analog				H-analog (this work)					H-analog (Head-Gordon et al.)[k]		
	C7eq	C7ax	C5	αR	C7eq	C7ax	C5	αR		C7eq	C7ax	C5
Φ	-86.1°	76.0°	-157.1°	-60.7°	-85.6°	75.0°	-158.0°	-60.7°	Φ	-85.8°	75.1°	-155.6°
Ψ	78.8°	-55.4°	159.8°	-40.7°	75.7°	-53.9°	162.2°	-40.7°	Ψ	78.1°	-54.1°	160.2°
E(HF/6-31G**// HF/6-31G**) [b,c]	0.00	2.82	0.40	4.35	0.00	2.52	0.30	4.27	E(HF/6-31+G*// HF/6-31+G*) [h]	0.00	2.56	0.19
E(HF/TZVP// HF/6-31G**) [d,e]	0.00	2.95	0.21	4.21	0.00	2.63	0.02	4.00	E(HF/6-31+G**// HF/6-31+G*) [i]	0.00	2.53	0.14
E(MP2/TZVP// HF/6-31G**) [f,g]	0.00	2.05	1.47	3.91	0.00	2.03	1.12	3.61	E(MP2/6-31+G*// HF/6-31+G*) [j]	0.00	2.19	1.13

[a] Energies in kcal/mol. [b] Zero of energy is -492.8853015 hartrees (methyl-analog). [c] Zero of energy is -414.8059892 hartrees (H-analog). [d] Zero of energy is -493.0263184 hartrees (methyl-analog). [e] Zero of energy is -414.9307183 hartrees (H-analog). [f] Zero of energy is -494.6374648 hartrees (methyl-analog). [g] Zero of energy is -416.2424631 hartrees (H-analog). [h] Zero of energy is -414.7990973 hartrees. [i] Zero of energy is -414.8188004 hartrees. [j] Zero of energy is -416.0674595 hartrees. [k] Reference 4.

Table VI. Comparison of structural parameters obtained for HF/6-31G** optimized conformers of alanyl dipeptide analogs: N-acetyl-N'-methylalaninamide (methyl-analog) vs. N-formyl-alaninamide (H-analog).

Parameter [a]	C7eq		C7ax		C5		αR [b]	
	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog
Φ (C6-N7-C9-C12)	-86.1	-85.6	76.0	75.0	-157.1	-158.0	-60.7	-60.7
Ψ (N7-C9-C12-N17)	78.8	75.7	-55.4	-53.9	159.8	162.2	-40.7	-40.7
RH---O	2.22	2.23	2.04	2.07	2.21	2.07	--	--
RN(H)---O	3.05	3.05	2.92	2.93	2.65	2.93	--	--
ω1 (O5-C6-N7-H9)	185.6	184.9	176.0	175.8	182.5	182.6	171.6	172.5
ω2 (O13-C12-N17-H18)	169.8	170.3	177.6	176.6	172.9	174.8	193.7	197.8
H(1-3)-C4-C6-O5	53/175/297	--	80/200/323	--	42/164/285	--	23/145/264	--
H(14-16)-C19-N17-H18	28/147/268	--	6/126/246	--	23/142/262	--	83/203/322	--
RC4-C6	1.511	--	1.513	--	1.512	--	1.513	--
RC6-N7	1.349	1.342	1.348	1.342	1.348	1.342	1.361	1.353
RC6-O5	1.207	1.201	1.207	1.201	1.204	1.198	1.198	1.193
RN7-C9	1.457	1.458	1.463	1.463	1.442	1.442	1.454	1.455
RC9-C11	1.521	1.521	1.531	1.531	1.535	1.535	1.528	1.528
RC9-C12	1.535	1.535	1.535	1.534	1.526	1.525	1.530	1.529
RC12-O13	1.203	1.200	1.204	1.201	1.204	1.201	1.200	1.196
RC12-N17	1.345	1.347	1.340	1.344	1.345	1.347	1.352	1.357
RN17-C19	1.446	--	1.446	--	1.448	--	1.448	--
RN7-H8	0.993	0.995	0.992	0.995	0.994	0.991	0.994	0.992
RN17-H18	0.996	0.994	0.996	0.993	0.992	0.994	0.993	0.995
ΘC4-C6-N7	116.3	--	115.7	--	115.9	--	115.5	--

Table VI. Comparison of structural parameters obtained for HF/6-31G** optimized conformers of alanyl dipeptide analogs: N-acetyl-N'-methylalaninamide (methyl-analog) vs. N-formyl-alaninamide (H-analog), cont'd

Parameter [a]	C7eq		C7ax		C5		α_R [b]	
	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog	methyl-analog	H-analog
ΘC6-N7-C9	122.9	123.2	127.1	127.2	122.0	122.2	121.8	122.2
ΘN7-C9-C12	109.8	109.8	114.3	114.0	107.4	107.4	113.8	113.7
ΘC9-C12-N17	114.6	114.5	117.4	117.1	115.6	115.5	116.5	116.5
ΘC12-N17-C19	121.2	---	120.9	---	121.7	---	120.4	---

[a] Bond lengths in Å. Angles in degrees. [b] Φ and Ψ constrained.

two sets of calculations carried out on hydrogen-blocked analogs but using different QM protocols. Again the Φ values are fairly consistent for a given conformer with more variation seen in the Ψ values. Also, the distance between the heavy atoms in the hydrogen bonding interaction differs considerably (3.05 Å vs. 2.90 Å) between the two C7eq conformers, even though the H---O distances are nearly the same. In the C5 conformer the hydrogen bond H---O atom distance is 0.14 Å longer in the methyl-blocked analog as compared to the hydrogen-blocked analog, while the heavy atom distance is 0.28 Å shorter! Based on this data we could rationalize the higher energy of the methyl-blocked C5 conformer using simple electrostatic arguments. In the methyl-blocked analog the hydrogen bonding atoms, the H and the O, are farther apart than in the hydrogen-blocked analog, and the two heavy atoms (O and N) are closer together, repelling each other with their partial positive charges. However, this difference in the distance between the hydrogen bonding atoms is not seen for the two C5 glycine analogs, and the energy difference is even greater in that case. The influence exerted by the methyl groups must therefore be more subtle than a simple steric perturbation of the geometry.

Discussion

The conformational energies of the glycyl and alanyl dipeptides have traditionally been of interest to the developers and the users of protein molecular mechanical force fields. The backbones of proteins exhibit distinct conformational preferences[9], and one might expect that a force field which can model conformational preferences at the dipeptide level would also reproduce the conformational preferences seen in a larger polypeptide or protein.

Unfortunately for the developers of such force fields, until recently the size of the glycyl and alanyl dipeptides was such as to make them inaccessible to study with high level ab initio methods. When the Weiner et al. force field[10] was developed in 1984, the highest quality data available was at the HF/4-31G level. Weiner et al. compensated for the lack of

correlation in these calculations by adding an empirical dispersion correction[11], which was more recently shown by Gould and Kollman[1] to be a reasonable estimate of the correlation effects.

With the advent of fast integral routines[12] and direct SCF methods[13-16], calculations such as these are now more feasible. Brooks and Case have reviewed the recent activity reported in the area of alanyl dipeptide calculations.[17] Head-Gordon et al.[4] mapped the entire phi-psi surface at the 3-21G level, carrying out further geometry optimization at the 6-31+G* level on stationary points from the 3-21G surface. A different study by Frey et al.[19] examined the effect of optimizing at the MP2 level of theory. This study was limited to non-ionic glycine and the C7eq and C5 conformers of H-blocked alanyl dipeptide. Their results showed that optimization of alanyl dipeptide at the MP2/6-31G** level results in a final energy difference of 1.65 kcal/mole versus 1.32 kcal/mole obtained from a single point MP2 calculation on the HF/6-311G** optimized geometry. This difference must result from a difference in optimized geometry, since the same basis set was used for the MP2 calculation in each case. When geometry optimization was carried out at the MP2 level, the distance between the hydrogen bonding atoms was reduced for the both the alanyl dipeptide C7eq and C5 conformations, from 2.27 Å to 2.10 Å and from 2.22 Å to 2.18 Å, respectively. While a difference of 0.3 kcal/mole is not enormous, it is troubling that the additional results presented for glycine revealed an opposite trend in that the MP2 optimized structures resulted in a *smaller* energy difference between the C7 and C5 structures as compared to the HF optimized structures. This difference in the trends of the conformational energies occurred despite the fact that the distances between the hydrogen bonding atoms in glycine were reduced by the same amount as the ones in alanyl dipeptide when the MP2 optimized structures were compared to the Hartree-Fock optimized ones.

Halgren has also carried out MP2/6-31+G** optimizations of both H-blocked glycyl and alanyl dipeptides.[19] He found the C5 conformation to be 1.22 kcal/mole higher in energy than the C7 conformation at the MP2//MP2 level. MP4 single point calculations (which included single, double, and quadruple excitations) on the MP2 optimized structures yielded an energy difference of 0.91 kcal/mole. This compares with the energy difference of 1.11 kcal/mole seen with single point MP2 calculations on HF optimized geometries. For alanyl dipeptide, the relative conformational energies at the MP2//MP2 level of theory put C5 at 1.18 kcal/mole and C7ax at 2.17 kcal/mole above the C7eq conformation. The MP4SDQ single point on the MP2 optimized geometries yielded relative energies of 1.08 kcal/mole for C5 and 2.21 kcal/mole for C7ax. These energies are nearly identical to those obtained from MP2 calculations on the HF optimized geometries -- 1.13 kcal/mole for C5 and 2.19 kcal/mole for C7ax. Thus, Halgren's data shows less dependence on the inclusion of the MP2 correction during the geometry optimization than was seen in the study by Frey et al.[19] Geometry optimization at the MP2 level is currently not practical with the methyl-blocked analogs.

Conclusion

We have presented the results of calculations which examined the effects of the basis set used in the geometry optimization and also of the chemical model chosen for calculating the conformational energies of glycyl and alanyl dipeptides. Our results suggest that geometry optimization at the HF/6-31G** level followed by a single point MP2/TZVP calculation is a reasonable protocol, as evaluated by comparison with results obtained at the MP2/TZVP//HF/TZVP level. Furthermore, calculations on the H-blocked analogs of the two dipeptides revealed that the methyl groups did exert a significant effect on the conformational energies, raising the energy of the C5 conformation by 0.35 kcal/mole for alanyl dipeptide and by 0.70 kcal/mole for glycyl dipeptide. We are thus satisfied with our calculations, which have employed the methyl-blocked analogs and a fairly high level of

theory (MP2/TZVP//HF/6-31G**). We have proceeded using the energies of the methyl-blocked analogs in the development of ϕ and ψ parameters for a new force field.[3]

References

1. Gould, I.R.; Kollman, P.A. *J. Phys. Chem.* **1992**, *96*, 9255.

2. Gould, I.R.; Cornell, W.D.; Hillier, I.H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.

3. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M., Jr.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A., submitted.

4. Head-Gordon, T.; Head-Gordon, M.; Frisch, M.; Pople, J.A.; Brooks, C.L. *J. Am. Chem. Soc.* **1991**, *113*, 5989.

5. Hehre, W.J.; Radom, L.; Schleyer, P.v.R.; Pople, J.A. *Ab Initio Molecular Orbital Theory*; Wiley, New York, 1986.

6. Head-Gordon, M.; Frisch, M.J.; Pople, J.A. *Chem. Phys. Lett.* **1988**, *153*, 503.

7. Dunning, T.H. *J. Chem. Phys.* **1971**, *55*, 716.

8. (a) Frisch, M.J.; Head-Gordon, M.; Trucks, G.W.; Foresman, J.B.; Schlegel, H.B.; Raghavachari, K.; Robb, M.A.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seger, R.; Melius, C.F.; Baker, J.; Martin, L.R.; Kahn, L.R.; Stewart, J.J.P.; Topiol, S.; Pople, J.A. *Gaussian 90*, Gaussian, Inc., Pittsburgh, PA 1990. (b) Frisch, M.J.; Trucks, G.W.; Head-Gordon, M.; Gill, P.M.W.; Wong, M.W.; Foresman, J.B.; Johnson, B.G.; Schlegel, H.B.; Robb, M.A.; Replogle, E.S.; Gomperts, R.; Andres, J.L.; Raghavachari, K.; Binkley, J.S.; Gonzalez, C.; Martin, R.L.; Fox, D.J.; Defrees, D.J.; Baker, J.; Stewart, J.J.P.; Pople, J.A. *Gaussian 92*, Revision A, Gaussian, Inc., Pittsburgh, PA 1992.

9. Richardson, J.S., *Adv. Protein Chem.* **1981**, *34*, 167.

10. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem* **1986**, *7*, 230.

11. Weiner, S.J.; Singh, U.C.; O'Donnell, T.J.; Kollman, P.A. *J. Am. Chem. Soc.* **1984,** *106,* 6243.

12. Head-Gordon, M.; Pople, J.A. *J. Chem. Phys.* **1988,** *89,* 5777.

13. Almlof, J.; Faegri, K., Jr.; Korsell, K. *J. Comput. Chem* **1982,** *3,* 385.

14. Frisch, M.J.; Head-Gordon, M.; Pople, J.A. *Chem. Phys.* **1990,** *141,* 189.

15. Head-Gordon, M.; Frisch, M.J.; Pople, J.A. *Chem. Phys. Lett.* **1988,** *153,* 503.

16. Frisch, M.J.; Head-Gordon, M.; Pople, J.A. *Chem. Phys. Lett.* **1990,** *166,* 275.

17. Brooks, C.L. III; Case, D.A. *Chemical Reviews* **1993,** *93,* 2487.

18. Frey, R.F.; Coffin, J.; Newton, S.Q.; Ramek, M.; Cheng, V.K.W.; Momany, F.A.; Schafer, L. *J. Am. Chem. Soc.* **1992,** *114,* 5369.

19. St-Amant, A.; Cornell, W.D.; Halgren, T.A.; Kollman, P.A. *J. Comp. Chem.,* accepted.

Chapter 4

Application of the Multimolecule and Multiconformation

RESP Methodology to Biopolymers:

Charge Derivation for Proteins

[1] Per
093,

[2] Gra

[3] Cur
Cana

Application of the Multimolecule and Multiconformation

RESP Methodology to Biopolymers:

Charge Derivation for DNA, RNA, and Proteins

by

Piotr Cieplak,[1]

Wendy D. Cornell,[2]

Christopher Bayly,[3]

and

Peter Kollman[2]

Department of Pharmaceutical Chemistry, University of California,

San Francisco, CA 94143, USA

[1] Permanent address: Department of Chemistry, University of Warsaw, Pasteura 1, 02-093, Warsaw, Poland.

[2] Graduate Group in Biophysics

[3] Current address: Merck Frosst Inc., C.P. 1005 Pointe Claire - Dorval, Quebec, H9R 4P8 Canada

Abstract

We present the derivation of charges for the amino acids using electrostatic potentials obtained from *ab initio* calculations with the 6-31G* basis set. We have combined multiple conformation fitting, previously employed by Williams [Williams, *Biopolymers*, **29**, **1367**, (1990)] and Reynolds *et al.* [Reynolds *et al.*, *J. Am. Chem. Soc.*, **114**, 9075, (**1992**)] with the RESP approach to derive charges for blocked dipeptide versions of each of the 20 naturally occuring amino acids. Based on our earlier results for propyl amine [Cornell *et al.*, *J. Am. Chem. Soc.*, **115**, 9620, (1993)], we suggest that the use of two conformations for each peptide suffices to give charges that well represent the conformationally dependent electrostatic properties of molecules, provided that these two conformations represent different rotamers of the dihedral angles that terminate in heteroatoms or hydrogens attached to heteroatoms. In these blocked dipeptide models, it is useful to require equivalent N-H and C=O charges for all amino acids with a given net charge (except proline), and this is accomplished in straightforward fashion with multiple molecule fitting. Finally, the application of multiple Lagrange constraints allows for the derivation of monomeric residues with the appropriate net charge from chemically blocked versions of the residues. The multiple Lagrange constraints also enable charges from two or more molecules to be spliced together in a well defined fashion. Thus, the combined use of multiple molecules, multiple conformations, multiple Lagrangian constraints, and RESP fitting is shown to be a very powerful approach to derive electrostatic potential based charges for biopolymers.

I. Introduction

There are three desirable properties of atomic charges to be used in molecular mechanical studies of complex molecules -- accuracy, consistency, and transferability. The accuracy of a set of charges is defined by its ability to reproduce pre-defined physical properties. When the charges are applied to the simulation of biological and organic molecules, the properties of interest are interaction energies, free energies of solvation, and conformational energies. The consistency of a charge model is defined by the extent to which similar charges are determined from different conformations of a molecule. Consistency is important because charges derived from one conformation of a molecule should be able to model the physical properties of the molecule in other conformations as well. The transferability of a charge set refers to how similar the charges are on a given functional group across a series of homologous molecules. Transferability of charges is important because other parameters in the force field, such as the dihedral parameters, are transferable by definition and all of the parameters should form a self-consistent set.

There are a variety of ways to achieve these three desired properties, but two different approaches are highlighted here. The first approach is to derive charges empirically; the most elaborated application of this approach to peptides and proteins is the OPLS model [1]. By carrying out Monte Carlo calculations on representative neat liquids, partial charges on atom types can be derived which optimize the agreement between calculation and experiment. Transferability is assumed, which, based on Monte Carlo calculations on a number of related liquids, is often a reasonable assumption. The main disadvantage of this approach is the requirement of multiple and computationally expensive Monte Carlo simulations on requisite liquids, the fact that such methods cannot be easily extended to excited states [2], the difficulty of assessing when the charge

"transferability" breaks down, and the subjective judgements that must be made in that regard in charge derivation.

The other main approach to deriving partial charges is based on the use of quantum mechanical calculations. The actual use of intermolecular interactions in this derivation [3] is impractical in general for deriving charges but the molecular charge distribution has been useful in this regard. It is also clear that the use of the quantum mechanical electrostatic potential or field is often a useful element in the derivation of charges that accurately represent the molecular multipole moments [4-7]. Thus, the partial charge models most often involve a least squares fit between the model and the quantum mechanical potential. This method has the advantage that, with current computer power, charges can be derived for many molecules of significant size in a reasonable amount of computer time. The charges derived can be quite dependent on the *ab initio* basis set or semi-empirical methodology, but a reasonable model of choice is the use of *ab initio* derived charges using a 6-31G* basis set [2], which uniformly overestimates molecular polarity. This overestimate makes such models relatively well balanced with solvent models such as TIP3P [8] or SPC [9] water, which include polarization effects implicitly because they have been empirically calibrated to reproduce the density and enthalpy of vaporization of the liquid.

These electrostatic potential (esp) derived charges have suffered from two main disadvantages. First, they have not been very consistent, with different conformations of a given molecule giving rise to dissimilar charge sets. Although there are real dependencies of partial charges on molecular conformation, these cannot be easily handled within the current framework of two-body additive molecular mechanical potentials of biopolymers.

Second, derivation of charges for large polymers becomes impractical, even with powerful computers. A way out of the first problem has been offered by the recent development and implementation of multiple conformation fitting [10,11] and RESP charges [11,12]. The use of these two techniques offer a very significant improvement in the quality and applicability of electrostatically determined charges.

How, then, should one derive charges for biopolymers? That is the focus of this chapter. Weiner *et al.* derived electrostatic potential based charges for monomers of proteins and nucleic acids and then pieced these together, adjusting charges on junction atoms to ensure unit charges [6,7]. This is a reasonable approach given that one chooses appropriate atoms (i.e. non-polar and non-conjugated) for these adjustments. However, it suffers from being non-algorithmic and not easily generalizable.

Recently, Bayly *et al.* has developed new software to allow simultaneously: multiple conformations, multiple molecules, charge restraints, and Lagrangian constraints in the derivation of the charge model [11,12]. This allows the derivation of a set of charges for all of the naturally occurring amino acid dipeptides and, thus, all the charges necessary for the simulations of proteins. Although there are still some subjective decisions that need to be made in applying these algorithms, they can be clearly defined at the beginning and consistently followed throughout. Thus, we feel this work offers a new, more powerful, and general approach for the derivation of charges for organic molecules and biopolymers.

II. Methodology

A. General remarks

We begin the derivation of the charges by calculating electrostatic potentials at a grid of points [5] around appropriate components of the amino acids. The Hartree-Fock method with the 6-31G* [13] basis set and the Gaussian 90 program [14] were used to calculate the wave function and electrostatic potentials of all of the amino acids with appropriate CH_3-CO- and -NH-CH_3 blocking groups (i.e. "dipeptides"). The dipeptides were optimized using molecular mechanics [15] with the Weiner *et al.* [6,7] force field. Potentials were calculated for two different conformations of each amino acid (or four in the case of proline) representing different backbone and sidechain conformations.

Faster computers and the direct Hartree-Fock approach enabled calculations on larger systems than previously considered, e.g. nucleosides and dipeptides of each of the naturally occuring amino acids. This reduced the number of components for which *ab initio* calculations needed to be done; also, using larger fragments, we decrease the possibility of force field inaccuracies arising from building larger residues from smaller ones. The electrostatic potentials were subsequently used in our RESP fitting procedure [11,12].

The RESP charges for the amino acids were fit using two different conformations for each amino acid (or four in the case of proline). With the exception of the proline residue, each amino acid was represented in both its α-helical and its extended (β-sheet) forms. Side chain χ values were chosen based on the Protein Data Bank [16] analysis of McGregor *et al.* [17], which correlates backbone and sidechain conformations for each of the amino acids. Each molecule was first minimized with backbone and sidechain restraints. The restraints were then removed (except for the α-helix backbone, where restraints of ϕ= -60° and ψ= -40° were used throughout) and the molecule allowed to minimize freely. Molecules which did not remain in the desired local minimum were

either reminimized with an intermedidate step employing a smaller subset of the original restraints followed by a final free minimization, or in the case that that strategy also failed, the final minimization employed restraints on the necessary dihedrals. All peptide bonds were in the trans conformations, with the exception of proline, as described below.

We assigned χ values for each amino acid as follows. First, using the data from Table I in McGregor et al. [17] for residues in the center of α-helices and β-sheets, we assigned a different χ_1 (t,g+,g-) for the α-helix and the β-sheet conformations for a given residue. We follow the convention used in McGregor et al. [17] that g+ corresponds to 300°. The χ_1's were chosen so as to maximize the total number of occurences of these backbone-χ_1 combinations, where the χ_1 for the α-helix differs from the χ_1 for the β-sheet. Specifically, one calculates the percentage given for that χ_1 within either the α-helix (center) or β-sheet (center) category, multiplied by the total number of occurences of residues in that secondary structure category.

The χ_2 value was then assigned according to the data presented in Tables III and IV of McGregor et al. [17]. Once again, we chose the most common χ_2 for each backbone-χ_1 pair as long as that yielded a different χ_2 for each conformation. When the same χ_2 was preferred by each of the two backbone-χ_1 combinations, then different χ_2's were assigned so as to maximize the total number of occurences of the two backbone-χ_1-χ_2 combinations.

Sidechain hydrogens attached to oxygen or sulfur (Thr, Ser, Cys, Tyr) were placed according to their minimum energy minimized conformation of t/g+/g- or syn/anti. When this preference was the same for the two conformations of an amino acid, then the hydrogen was placed uniquely on each conformation so as to yield the lowest overall

energ

which

This

when

confo

S

confo

minir

pared

with n

T

mation

sidech

most

where

came

altern

robust

F

confor

ferent

trans pe

energy for the two conformations added together.

There were five exceptions to the preceding rules. Firstly, minimized conformations which had hydrogen bonds between the sidechain and backbone atoms were eliminated. This was accomplished either by minimizing with restraints on some of the dihedrals, or when that was not sufficient to eliminate the hydrogen bond, by choosing an alternative conformation.

Secondly, in the case of cysteine, the second most common pair of backbone-χ_1 conformations was used since the extended conformation did not stay in its local minimum when a χ_1 of -60° was used. The second pair occured 42% of the time as compared to 46% of the time for the most common pair, so they were considered to occur with nearly equal frequency.

Thirdly, McGregor et al. [17] tabulated data for the most common overall conformations of methionine, arginine, and lysine. For these three molecules, we chose sidechain conformations for the α-helix and β-sheet backbones that 1) were among the most commonly observed and 2) had different χ_1 values, and 3) had different χ_n values, where χ_n had a heteroatom in the first or fourth positions. The rationale for this choice came from our work on propylamine [11], where it was shown to be beneficial to allow alternative conformations around the central N-C-C-C bond, in order to derive the most robust set of charges for this molecule.

Fourthly, in proline the peptide bond is found to be in the cis rather than the trans conformation approximately 20% of the time (Creighton [18]). For this reason, four different conformations were used for the proline residue, representing both the cis and trans peptide bonds as well as two different backbone conformations. The backbone con-

formatt

[19]. T

minim

mean

confor

for the

= -76°

minim

the cis

F

represe

backbc

tions b

right-h

equal r

than di

with χ

left-han

cystine

quite f

data, w

helical

dihedral

-27°. Th

formations were assigned based on data in the PDB survey by MacArthur and Thornton [19]. They found that phi-psi plots of both trans and cis proline exhibited two distinct minima, corresponding to conformations labelled "α" and "β." For trans proline the mean ϕ and ψ values for the α conformation were ϕ = -61° and ψ = -35° and for the β conformation they were ϕ = -65° and ψ = 150°. For cis proline the mean ϕ and ψ values for the α conformation were ϕ = -86° and ψ = -1° and for the β conformation they were ϕ = -76° and ψ = 159°. These were the conformations used for the backbones. The minimum energy ring pucker was chosen for each backbone conformation (within either the cis or trans set) since it was different in each case.

Finally, the cystine residue was treated in its disulfide bridged form, so it was represented by only one molecule comprised of two residues each having a different backbone conformation. We assigned the dicystine backbone and sidechain conformations based on data in the PDB survey carried out by Thornton [20]. She found that right-handed (χ_3 = +90°) and left-handed (χ_3 = -90°) disulfides occured in relatively equal numbers. The right-handed disulfides displayed a greater variety of conformations than did the left-handed ones, however, with 70% of the left-handed disulfides occuring with χ_2 = χ_2' = -80° and χ_1 = χ_1' = -60°. We therefore chose to use the predominant left-handed conformation for assigning the sidechain dihedrals. Thornton [20] found that cystine residues occured primarily with random coil backbones (59%), but also occured quite frequently in α-helical (25%) and β-sheet (18%) conformations. Based on this data, we gave one of the cystine residues an extended backbone and the other an α-helical one. The final minimization was carried out without restraints on the ϕ, ψ dihedrals of the α-helical backbone. The minimized values of ϕ and ψ were -60° and -27°. The conformations chosen for each amino acid are described in Table I.

descr

lowir

where

and

In the

(ESP)

limits

the tig

approp

λ_1 are

requir

which I

B. RESP fitting methodology

The procedure of RESP fitting to obtain atomic charges has been previously described [11,12]. The term "RESP" refers to restrained ESP charge fitting using the following equation:

$$f(q_1,....,q_{natoms}) = \chi_{esp}^2 + \chi_{hyp.restr}^2 + \lambda_1 g_1 + \cdots + \lambda_w g_w, \tag{1}$$

where

$$\chi_{esp}^2 = \sum_{i=1}^{ESPpoints} (V_i - \sum_{j=1}^{natoms} \frac{q_j}{r_{ij}})^2 \tag{2}$$

and

$$\chi_{hyp.restr}^2 = a \sum_{j=1}^{natoms} ((q_j^2 + b^2)^{1/2} - b). \tag{3}$$

In the above formulas V_i is the quantum mechanically calculated electrostatic potential (ESP) at point i, q_j are the resultant charges, a is a scale factor defining the asymptotic limits of the strength of the hyperbolic restraint according to equation (3), and b defines the tightness of the hyperbola around the minimum, A value of 0.1 was found to be appropriate for b [12]. The g_i are additional constraints imposed on resultant charges and λ_i are Lagrange multipliers. The minimum of the $f(q_1,....,q_{natoms})$ function is sought by requiring that:

$$\frac{\partial f}{\partial q_k} = 0, \quad \frac{\partial f}{\partial \lambda_l} = 0, \quad \text{for each k,l,} \tag{4}$$

which leads to the matrix equation of the type:

$$A \; q = B \qquad (5)$$

which must be solved for q. This is done iteratively when using nonzero hyperbolic restraints, since the left-hand side of the equation (5) (matrix A) depends on charges q.

The RESP fitting scheme, which we have demonstrated to be useful, involves a two-stage procedure with hyperbolic restraints which we denoted as (wk.fr./st.eq.) in earlier works [11,12]. In the first stage, a weak hyperbolic restraint (a=0.0005) to a target value of 0.0 is applied to all heavy atoms. Hydrogen atoms are not restrained in either stage, as they are never buried within a molecule and are always well-defined by the esp points. In the second stage, charges on all atoms are kept frozen to their values obtained in the first stage, except for those in methyl and methylene groups. CH_3 and CH_2 groups are refit with the hydrogens within a given group constrained to have equivalent charges. The hyperbolic restraint applied during the second stage is twice as strong as the one in stage one (a=0.001). The two-stage restrained standard ESP charges exhibit less conformational dependence compared to the standard ESP charges, result in good conformational energies, and gives good results for hydrogen bonding energies and free energies of solvation [11,12].

The necessity of a two-stage fit arises from the need to constrain atoms which are not symmetrically equivalent within the static conformation of the molecule used for the calculation, but which become equivalent under dynamical conditions when rotation can occur. One example of this would be the three methyl hydrogens in methanol. If these inequivalent atoms are forced to have the same charge during a one stage fit, the charge on the oxygen is reduced to a value which does not yield good free energies of solvation or interaction energies. The two stage fit then allows for the "best" charges to be fit on the heteroatoms during the first stage, with the maximum number of degrees of freedom

available to the molecule. Then methyl and sometimes methylene hydrogens are constrained to be equivalent in the second stage of the fit.

C. The role of Lagrange constraints.

The role of Lagrange constraints (conditions) in equation (1) is manifold. In the standard RESP procedure, described in the previous section, they were used for two purposes. In the first and simplest case, they were used to keep the sum of charges to be equal to the total molecular charge. In the second case, they were used to force identical charges on equivalent atoms during the fit. The most common example applies to methyl hydrogens in the second stage, as mentioned above; however, chemically equivalent atoms which are not refit in stage two can be constrained to have equivalent charges in stage one. The two oxygens in the sidechain of aspartic acid are an example of the latter situation. When two different methyl groups in a molecule were defined to be symmetric, the two carbons were constrained to have the same charge during the first stage of the fit, but each hydrogen was allowed to optimize freely.

In the present paper, Lagrange multipliers are shown to have some additional uses. They will be used for equivalencing atomic charges on the same atoms of different conformers of the same molecule. This was extensively tested in our earlier study for propylamine [11], and will be applied here to equivalence the atoms in the different conformers of the amino acids. Multiple conformation fitting has been shown to be useful in dealing with non-physical conformational variation of charges. Unlike Reynolds et al. [10], however, we do not use Boltzmann weighting for different conformers, since we do not know the relative energies in solution or the dielectric environment of a protein.

on

ato

to c

in a

resp

amii

two

amm

in or

of q

splic

the C

timol

devel

III. D

A. B

A

that t

methyl

the reg

The Lagrangian multiplier method will also be applied to equivalence some charges on atoms during multiple molecule fitting. This will be used to force similar groups of atoms in different residues to have the same atomic charges. This strategy was employed to derive charges for sugar atoms in different nucleosides and -CO-NH- backbone atoms in amino acids. Using this approach in creating our database is especially important with respect to its further application for any free energy perturbation calculations [21].

Finally, Lagrange constraints can be used to force the proper net unit charge on the amino acid residue minus its blocking groups (-NH-CHR-CO-)and to splice together two fragments from different molecules. The latter application is employed to splice an ammonia group from methyl ammonium onto the central residue of a blocked amino acid in order to create the N-terminal charged residues without carrying out an additional set of quantum mechanical calculations on those residues. A similar strategy is used to splice the carboxylate group from methyl acetate onto the blocked amino acid to create the C-terminal charged amino acids.

The features described above, i.e. restrained electrostatic potential (RESP), multimolecular and multiconformation fitting, have all been employed in our charge development for proteins, DNA, and RNA.

III. Derivation of the Amino Acid Charges

A. Basis for evaluating different charge sets

A known weakness of the Weiner et al. [6,7] and some other protein force fields is that the energy calculated for the $C7_{eq}$ and C7 conformations of N-acetyl and N-methylamide blocked alanyl and glycyl dipeptides is significantly too stable compared to the results from high level quantum mechanical calculations. A reason for this

discrepancy is that at the time that the Weiner *et al.* force field was developed (mid 1980's), computer limitations were such that no high level quantum mechanical data was available for these molecules for use in calibration of the force field. A number of such calculations have recently been carried out [22-28] at varying levels of theory and employing either the methyl-blocked residues described above or ones in which those methyl groups are replaced by hydrogen atoms. Thus, a key motivation in the evaluation of the charge models described below is to choose those that, firstly, are representative of important conformations, and, secondly, come closest to reproducing the quantum mechanical conformational energies.

B. Conformational energies from single and multiple conformation charge sets

RESP charges were first calculated using the potentials generated for some of the quantum mechanically optimized (HF/6-31G**) conformers of the alanyl dipeptide [23]. Previous studies [10,11] have shown that multiple conformation fits produce charge sets which perform better at reproducing the electrostatic potentials of more of the low energy conformations than does a single conformation fit. We thus wished to examine the conformational energies which resulted from charge sets derived from different conformations. We calculated RESP charges derived from single conformation fits of the $C7_{eq}$, $C7_{ax}$, $C5$ (extended β-sheet), and α_R (α-helical) conformations; as well as from multiple conformation fits of C5 and α_R; $C7_{eq}$, C5, and α_R, and all four conformations together. Those charges are presented in Table II.

The conformational energies calculated using the various charge sets are presented in Table III. All of the charge sets result in conformational energies with similar trends. The $C7_{ax}$ conformation is from 1.0-1.4 kcal/mole too low in energy and the C5 confor-

mation is from 0.5-2.7 kcal/mole too high in energy relative to the quantum mechanical values. Like the C5 conformation, the α_R conformation is fairly sensitive to the charge model used, ranging from being 1.1-3.4 kcal/mole too high in energy. It is interesting that the highest energy conformation, the α_R, yields charges which produce the best set of energies. This is consistent with results seen earlier for dopamine [29] and propylamine [11].

In Table IV are presented the molecular mechanically optimized values of phi and psi obtained for the $C7_{eq}$, $C7_{ax}$, and C5 conformations with the different charge sets. The molecular mechanically optimized values for a given angle in a particular conformation vary by at most 5 degrees. The variation from the quantum mechanically optimized values ranged from 3 to 15 degrees. The different charge sets are thus seen to result in minimized geometries which agree well with each other but vary more in comparison with the quantum mechanically optimized values.

The dipole moments of the molecular mechanically optimized structures using the various charge sets are presented in Table V. Previous studies [10, 11] have shown that charges derived from a single conformation or multiple conformations reproduce the dipole of that (those) conformation(s) fairly accurately, but perform less well at reproducing the dipole moments of other conformations. Disagreement between the dipole moments of these optimized structures and the quantum mechanical dipole moments arise from differences in the charge distribution and also from differences in geometry. When RESP charges derived from a single conformation were used for the geometry optimization of that same conformation, the molecular mechanical dipole moments were as much as 28% in error ($C7_{eq}$) by comparison with the quantum mechanical values. The RMS deviation between the quantum mechanical and molecular mechanical dipole

moments is given for each charge set. The best agreement is seen with the RESP charges derived from the C5 conformation.

In order to allow the effects of the charge model to be separated from the effects of the geometry used for the calculation of the dipole moments, in Table VI the dipole moments are given for each of the quantum mechanically optimized geometries as a function of the charge set. For the conformations included in the charge derivation, the dipole moment of that conformation is given after both stage one and stage two of the RESP fit. The RMS deviations for these sets of dipole moments are considerably smaller than the ones which were observed when the molecular mechanically optimized geometries were used. The dipole moment of the α_R conformation is seen to be up to 1 D too large and this is caused by errors introduced in the second stage of the fit during the methyl refitting.

The next step in the development and analysis of a final charge model involved the application of multiple lagrange constraints on the net charge of the central ala residue, the acetyl (ACE) blocking group, and the N-methyl (NME) blocking group. Before these constraints were applied, the net charge on each of the three residues was determined for the various charge models. For the four single conformation RESP fits, the central ala residue had a negative charge of magnitude less than 0.1. Each acetyl group had a positive net charge of less than 0.1. The net charge on the N-methyl group was the smallest on average and was either positive or negative. The small magnitudes of the net charges on the residues suggested that constraining each to have a net neutral charge was a reasonable approximation.

Subsequent calculations were carried out using a $C5/\alpha_R$ multiple conformation model. The single conformation α_R fit provided better alanyl dipeptide energies, but we

wanted to use a multiple conformation fit since such charges perform better in modelling the electrostatic potential of many conformations of the amino acid. The three conformation model was rejected as its energies were nearly identical to those resulting from the two conformation model. Furthermore, PDB data was available which provided information on side chain conformations preferred for the two different types of secondary structure -- β-sheet (C5) and α-helix (α_R). Using these two conformations it was then a straightforward task to assign side chain conformations for all of the amino acids (see Methods section).

It is apparent that even the simplest model for calculating the dipeptide charges, i.e. one with no additional constraints, did not result in molecular mechanical energies in good agreement with the quantum mechanical energies. We therefore proceeded in our development of a charge model for the amino acids by testing the effect of each aspect of the model on the conformational energies calculated.

C. Effects of multiple Lagrange constraints and equivalencing of backbone amide atoms

In Table VIII we present the conformational energies calculated for glycyl and alanyl dipeptides with and without additional Lagrange constraints to produce three neutral residues and with and without forcing equivalent charges on the amide groups on either side of the α carbon (Figure 1). The constraint of three neutral residues is necessary when deriving charges for the amino acid database since each amino acid residue must have the appropriate net integral charge. When carrying out the quantum mechanical electrostatic potential calculation, however, one must used a blocked form of the residue so as to have a chemically reasonable structure. The desireability of employing the constraint to equivalence amide groups is related to the desire to have a consensus set of

backbone charges that would be used for all of the amino acids. The need for this simplification arises from results obtained by Sun *et al.* [21], as discussed below. We thus decided to make the simplifying assumption that different residues having the same net charge should have common backbone amide atom charges. The question of equivalencing the amide group charges from the blocking groups in the charge fit then arises. These C=O or N-H groups in the blocking groups are mimicing the adjacent residues in the protein or peptide, and, in principle, those adjacent residues should have C=O and N-H charges which are identical to the ones found in the central residue.

The results presented in Table VIII employed the $C5/\alpha_R$ multiple conformation fit charges derived from the quantum mechanically optimized structures for both glycyl and alanyl dipeptides. The first set of molecular mechanical energies corresponds to charges derived with each dipeptide treated as a single residue and with no constrained equivalence of the two backbone amides. These alanyl dipeptide energies differ slightly from the ones in the $C5/\alpha_R$ column of Table III, because in Table VIII the charges were derived without constraining the two conformations to have common charges on their N-terminal methyls and common charges on their C-terminal methyls. It was necessary to remove these two constraints, otherwise there were not enough degrees of freedom in the fit when the molecule was treated as three neutral residues for the charge fit algorithm to converge.

The two different types of constraints on the fit have similar effects. Both cause the alanyl dipeptide $C7_{ax}$ conformation to be lowered in energy by about 0.4 kcal/mole. The C5 and α_R conformations of alanyl dipeptide both increase in energy with the application of either constraint, the C5 conformation by 0.5-0.9 kcal/mole and the α_R conformation by 1.7-2.1 kcal/mole. The energy of glycyl dipeptide's α_R conformation also goes up in

energy in both cases, whereas the C5 conformation goes up by 0.5 kcal/mole in one case and down in the other. The effect of the simultaneous application of both types constraints is to lower the energy of the C5 conformation of glycyl dipeptide by about 0.5 kcal/mole and to raise the energy of the α_R conformation by 1.4 kcal/mole. The effect on the alanyl dipeptide is to lower the energy of the $C7_{ax}$ conformation by 0.4 kcal/mole, to lower the energy of the C5 conformation by 0.3 kcal/mole, and to increase the energy of the α_R conformation by 1.7 kcal/mole. The conformation which is most in error is the α_R conformation of glycyl dipeptide, which is 3.3 kcal/mole higher in energy than the quantum mechanical reference energy.

D. Effect of molecular mechanical versus quantum mechanical geometry optimization

The above charge calculations were carried out on structures which had been optimized with nearly the same basis set that was used for calculating the electrostatic potential for the charge calculation (6-31G** vs 6-31G*). One might expect such optimized geometries to provide the most reasonable charges. However, such high level *ab initio* geometry optimizations are fairly costly, so for the remaining amino acids we settled for structures optimized using molecular mechanics -- in this case the Weiner *et al.* force field. [6]. In Table IX we compare energies calculated with charges derived from AMBER optimized structures with those derived from quantum mechanically optimized structures. All charge sets were derived from multiple conformation fit employing the C5 and α_R conformations and did not employ the three neutral residue constraint or the equivalent amide constraint. These energies show the same general trends as the ones calculated from the QM optimized geometry charges, but the energies of the α_R conformations are even farther off from the quantum mechanical target energies. It would

therefore seem that the quantum mechanically optimized structures yielded the best charges. We lack sufficient computer resources to carry out the optimizations on the remaining amino acids, however, and therefore were limited to dealing with the molecular mechanically optimized structures.

E. Multiple molecule and multiple conformation fitting

With the above tests, we established that the constraints of three neutral residues and equivalent amide groups and the use of molecular mechanically optimized structures produced charge sets which yielded conformational energies in reasonable agreement to the "ideal" charge sets, derived from the quantum mechanically optimized geometries and with no additional Lagrange constraints imposed. We next set out to carry out a multiple molecule fit for the purpose of deriving a set of consensus charges for the backbone amide atoms. As noted above, the use of consensus charges for the amide atoms was motivated by results obtained by Sun et al. [21] in a free energy perturbation study involving the perturbation of alanyl to valine. They found that the majority of the change in free energy was derived from interactions between water molecules and the atoms in the backbone of each residue. Corresponding backbone atoms had fairly different charges for each residue.

While the nature of a given side chain would be expected to have some effect on the electrostatic character of the backbone, this effect would probably be fairly subtle. The large variation in charge seen by Sun et al. [21] was likely more of an artifact of the esp fitting procedure. Although this variation is reduced with the RESP procedure, we chose to use consensus amide charges in order to completely avoid the problem. The use of a simplified charge model having consensus charges on the amide atoms then restricts the

charge variation to the side chains and the α carbon and hydrogen.

The first multiple molecule/conformation fit included the amino acids glycine, alanine, serine, valine, asparagine, aspartic acid, and protonated histidine. We chose this group to include the two simplest amino acids as well as a β-branched and hydrophobic chain, a short and a longer polar chain, and negative and positively charged side chains. This fit resulted in α carbon charges ranging from -0.084 to 0.038 for the neutral amino acids. The α carbon charges for aspartic acid and protonated histidine were -0.252 and 0.210, respectively. The larger charges on the α carbons of the charged residues suggested that their backbone amide groups were sufficiently different from the neutral residues as to merit separate fitting.

We therefore settled on three separate fits to determine consensus charges for the backbone amide atoms. The main fit consisted of the five neutral amino acids from the set of seven above -- glycine, alanine, serine, valine, and asparagine. The second fit consisted of the two negatively charged amino acids -- aspartic acid and glutamic acid. The third fit consisted of the three positively charged amino acids -- lysine, arginine, and protonated histidine. For the fits of the two groups of charged amino acids, the charges on the C=O and N-H groups in the blocking groups were constrained to have the consensus charges derived for the neutral amino acid backbone amide atoms. This modelled the presence of neutral amino acids on either side of the central charged residue. This was an approximation, since charged residues may be found adjacent to each other, but certain simplifying assumptions are necessary when deriving charges in this fashion for residues of a heteropolymer. The consensus amide charges from the neutral amino acid fit were applied to the remaining neutral amino acids through a constrained fit. The charges on the methyl hydrogens in the blocking groups were left free (not constrained to have

the same charge within a group) since the blocking groups were discarded after the fit. This allowed the best set of charges to be calculated for the amino acid residue.

Charges for the acetyl and N-methyl terminal blocking groups were taken from the C5 conformation of alanyl dipeptide. Hydrogens in a given methyl group were constrained to have a common charge during a "third-stage" fit where the charges on the remaining atoms were constrained to have the values determined in the two-stage five residue fit. The use of acetyl and N-methyl charges derived from different molecules or conformations was shown to have a minimal effect on the conformational energies calculated.

The final set of charges is given in Table X along with the charges from the Weiner et al. force field for comparitive purposes. The backbone amide charges are seen to be fairly similar to those used by Weiner et al. In the Weiner et al. force field, these charges were determined from a 6-31G calculation on N-methyl acetamide, and then scaled by a net factor of 0.75; 0.82 to make them more like the "optimal" 6-31G** derived charges and 0.91 to scale back a 6-31G** dipole moment to an experimental value. The consensus amide charges calculated for the negatively and positively charged amino acids differ from the consensus charges for the neutral amino acids primarily at the nitrogen and carbon atoms. The sum of the charges on the CONH backbone atoms is -0.115 for the neutral residues, 0.072 for the positive residues, and -0.267 for the negative residues.

The α-C and α-H charges were allowed to optimize independently for each amino acid in the new force field, in contrast to Weiner et al. where there was a common value. The charges on side chain heteroatoms are similar to the ones found in Weiner et al., but they are larger, reflecting the greater polarity of the 6-31G* basis set. The charges on the

sulfur atoms exhibit perhaps the greatest variation between the two force fields, but this is partly due to the absence of lone pairs in the new force field.

Although the new force field employs restrained esp (RESP) fitting, the charges on the atoms in alkyl groups are often larger in the new force field by comparison to Weiner *et al.* This effect is likely due to the use of the 6-31G* basis set in the current work versus the less polar STO-3G basis set used for side chain esp calculations in the Weiner *et al.* force field. The fact that two conformations were used to fit each amino acid lends further support to this argument, since large charges on buried atoms are less common when multiple conformations are employed, due to better statistical sampling.

In the cases of valine, leucine, and isoleucine, the charges on some of the alkyl atoms seemed to be particularly large. This was unexpected because RESP fitting applied to butane was shown to result in much smaller charges on the atoms compared to standard esp fitting. In each of these three amino acids, a large charge was observed on a carbon adjacent to a methine group. For example, the charge on the γ-C in valine is -0.319, the one on the δ-C in leucine is -0.412, and the one on the γ-C in isoleucine is -0.320. Unlike methyl and methylene groups, methine groups are not refit during the second stage of the RESP fit, and this was thought to be a possible cause of this seemingly aberrant result. To test this hypothesis, the second stage of each fit was repeated but with the methines allowed to reoptimize. The results of these fits are presented in Table XI. The large charges observed in the standard RESP fits are seen to be greatly reduced while the RRMS of the fit is improved in one case, worsened in another, and unchanged in the third. The dipole moments are also not significantly changed by refitting the methines.

A second aspect of the fit is the use of a stronger restraining function (0.0010 vs. 0.0005) in the second stage of the fit. The effect of this stronger restraint was tested by

carrying out the second stage of the fits with the weaker restraint and also refitting the methines. In the cases of valine and isoleucine, this model was seen to result in charges on the "aberrant" atoms about halfway between those seen with standard RESP fitting and RESP fitting with methines refit. However, in the case of leucine, although the charge on the carbon adjacent to the methine group was reduced over its standard RESP value, the charge on the methine carbon increased from its standard RESP value of 0.353 to 0.428! This exercise suggested that there was no obvious improvement to our current standard RESP approach for fitting charges. While the large alkyl charges may seem counter-intuitive, the true test of a charge model is its performance, and even the relatively large butane standard esp charges were shown to result in accurate conformational energies [11].

G. Effect of diffuse functions

The quantum mechanical calculations carried out for the purpose of deriving RESP fitted charges employed the 6-31G* basis set. This basis set is well studied and has been shown to perform well at modelling molecular properties. This basis set was derived to reproduce the properties of neutral atoms, however, and the addition of diffuse functions has been shown to be necessary to model certain properties of negatively charged systems.

In order to assess the effect of the inclusion of diffuse functions on the charge derivation, 6-31+G* calculations were carried out on the two aspartic acid and two glutamic acid conformers. The resulting wavefunctions gave rise to charges on the carboxylate atoms that were significantly higher than those derived from the 6-31G* wavefunctions. The aspartic acid 6-31+G* charges were 0.936 for the carboxylate carbon and

-0.868 for the oxygen as compared to 0.799 and -0.801 from the 6-31G* wavefunction.

The glutamic acid 6-31+G* charges were 0.878 for the carboxylate carbon and -0.855 for the oxygen as compared to 0.805 and -0.819 from the 6-31G* wavefunction.

The consensus amide charges were calculated to be -0.533, 0.299, 0.562, and -0.607 for the N, H, C, and O atoms, respectively, with the 6-31+G* wavefunction, which are fairly similar to the ones derived with the 6-31G* wavefunction. The net charge on those four backbone atoms only changes from -0.267 to -0.279 upon the addition of the diffuse functions. Although greater changes are seen in the charges of the carboxylate atoms, the net charge on each group changes by at most 0.003. There is thus no major "migration" of charge seen when diffuse functions are added, suggesting that the 6-31G* basis set does not force part of the net charge away from the carboxylate group.

H. N- and C-terminal amino acids

We have also calculated charges for the charged N- and C-terminal versions of the amino acids. These charges were derived by splicing the ammonia group from methylammonium or the carboxylate group from acetic acid onto the blocked versions of the amino acids. Figure 2 illustrates how this procedure was carried out. A Lagrange constraint was applied which forced the charges on the atoms within the two boxed regions together to sum to 0.0. In this way the proper charge was attained on the resulting residue. In addition, the charge on the methyl carbon in methylammonium or acetic acid was constrained to have the same value as the charge on the α carbon. Also, in the N-terminal residue fits the N-terminal N and H atoms were constrained to have the same charges for both conformations of a given amino acid even though these atoms were "discarded" after the fit. In the C-terminal fits, the C-terminal C and O atoms were similarly con-

strained. The charges for N-terminal, C-terminal, and central versions of glycine, alanine, serine, valine, and asparagine are shown in Table XII.

IV. Discussion

We have presented an application of multimolecule, multiconformation RESP charge fitting to amino acids. This approach leads to more reasonable charges for buried atoms as compared to the standard ESP approach. Also, the use of a 6-31G* basis set rather than the STO-3G basis set used by Weiner *et al.* has been shown to lead to hydrogen bond energies closer to those found with the highest level *ab initio* calculations [22]. The use of the splicing approach described above allows for an algorithmic merging of the charges of separate molecules. More extensive molecular mechanical simulations employing these charges will be presented elsewhere [30].

We have also presented the derivation of charges for the amino acids using RESP fitting and multiple molecules, and conformations. The amino acid charges also differ from those in the previous force field in that 6-31G* level calculations were carried out on blocked versions of entire amino acids, rather than fitting the backbone and the sidechains separately. Furthermore, Lagrange constraints were employed to obtain residues of the appropriate integral charge and also to splice ammonium and carboxylate groups onto the charged N- and C-terminal residues.

A variety of charge models were evaluated to determine their ability to reproduce the quantum mechanical conformational energies of the glycyl and alanyl dipeptides. It is curious that the RESP charges calculated for alanyl and glycyl dipeptides using both the C5 and α_R quantum mechanically optimized conformations for each, no additional Lagrange constraints for multiple residues of integral charge, and no equivalencing of

amide charge within and between the molecules did not better reproduce the quantum mechanically calculated energies (Table VIII).

This results runs counter to ones obtained for butane, simple alcohols, simple amines, ethane diol, and a series of substituted 1,3-dioxanes [31], where RESP (and standard ESP) charges combined with simple dihedral potentials performed quite well at reproducing relative conformational energies. The dipeptide molecules are more complicated, however, with their two amide groups, and with 6-31G* charges, the intramolecular hydrogen bonding in the C7$_{eq}$ and C7$_{ax}$ conformations may be exaggerated, leading to an overestimate of the stability of these two conformations relative to C5 and α_R.

Further support for this interpretation comes from simply scaling the charges by 0.88 as part of a non-additive model for peptide conformational analysis, which leads to significant improvement in the dipeptide energies [32]. It is not clear what the physical basis is for the fact that C7$_{ax}$ is ~2 kcal/mole less stable than C7$_{eq}$ at the quantum mechanical level, but only ~1 kcal/mole in the molecular mechanical models. We should emphasize that the dipeptide conformational energies can and have been improved through the addition of torsion parameters. The development of those torsion parameters is described in the next paper in this series [30].

We have presented charges for three forms of the amino acids: central residues and charged N- and C-terminal residues. One can analyze the charges on similar groups of atoms for a given residue type. For the N-terminal residues, the total charge on the ammonium group can be seen to vary from about 0.70 to 0.80. For the C-terminal residues, the total charge on the carboxylate group varies from about -0.79 to -0.85. The charges on the α-C's and α-H's exhibit the effects of induction caused by the adjacent charged group. Some inductive effect is also seen at the β position.

One can also compare charges for a particular amino acid in its three different forms. The charges on the atoms in the serine and valine sidechains are quite similar for all three versions of each residue. The charges on the asparagine sidechain, however, are less consistent. The charge on the γ-C has a value of about 0.71 for the central and C-terminal residue, but a value of 0.58 for the N-terminal residue. The charges on the remaining atoms in that sidechain are fairly consistent among the three residue types. We have only gone into detail about the five amino acids used for the neutral amino acid consensus fit, however charges have been calculated for the N- and C-terminal versions of all of the amino acids.

V. Conclusions

We have presented general methods for the derivation of charges for amino acids and nucleic acids for use in molecular mechanics calculations. Our strategy employs multiple conformation fitting to reduce the conformational dependence of the charges and multiple molecule fitting to derive consensus charges for certain common atoms where appropriate. Furthermore, the charges were fit to the electrostatic potential of each molecule with restraints applied to the charges in order to attenuate the charges of statistically ill-determined atoms. The charges have been placed into a database for use in carrying out molecular mechanical simulations on nucleic acids and proteins with a new force field [32].

Acknowledgements

References

(1) W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657, (1988).

(2) S. Debolt and P. A. Kollman, *J. Am. Chem. Soc.*, **112**, 7515, (1990).

(3) M. Aida, G. Corongiu, and E. Clementi, *Int. J. Quant. Chem.*, 1992, **42**, 1353.

(4) S. R. Cox and D. E. Williams, *J. Comp. Chem.*, **2**, 304, (1981). T. R. Stouch and D. E. Williams, *J. Comp. Chem.*, **13**, 622, (1992).

(5) U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, **5**, 129, (1984).

(6) S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7**, 230, (1986).

(7) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765, (1984).

(8) W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.*, **79**, 926, (1983).

(9) H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman, Reidal, Dordrecht, Holland, 1981, p. 331.

(10) C. A. Reynolds, J. W. Essex, and W. G. Richards, *J. Am. Chem. Soc.*, **114**, 9075, (1992).

(11) W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.*, **115**, 9620, (1993).

(12) C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.*, **97**, 10269, (1993).

(13) W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56**, 2257, (1972).

(14) M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople, *Gaussian 90*, Revision J, Gaussian, Inc., Pittsburgh PA, 1990.

(15) D. A. Pearlman, D. A. Case, J. W. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner and P. A. Kollman, *AMBER 4.0* (UCSF), University of California, San Francisco (1992).

(16) (a) F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Bio.*, **112**, 535, (1977). (b) E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng, "Protein Data Bank" in *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*, eds. F. H. Allen, G. Bergerhoff, and R. Sievers, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, 107-132.

(17) M. J. McGregor, S. A. Islam, and M. J. E. Sternberg, *J. Mol. Biol*, **198**, 295, (1987).

(18) T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company, N.Y., (1984).

(19) M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.*, **218**, 397, (1991).

(20) J. M. Thornton, *J. Mol. Biol.*, **151**, 261, (1981).

(21) Y. Sun, D. Spellmeyer, D. Pearlman, and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 6798, (1992).

(22) I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, **116**, 2493, (1994).

(23) I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 9255, (1992).

(24) I. R. Gould, W. D. Cornell, and I. H. Hillier, *J. Am. Chem. Soc.*, in press, (1994).

(25) T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks, and J. A. Pople, *J. Am. Chem. Soc.*, **113**, 5989, (1991).

(26) H. Bohm and S. J. Brode, *J. Am. Chem. Soc.*, **113**, 7129, (1991).

(27) L. Shafer, S. Q. Newton, M. Cao, A. Peeters, C. Vanalsenoy, K. Wolinski, and F. A. Momany, *J. Am. Chem. Soc.*, **115**, 272, (1993).

(28) R. F. Frey, J. Coffin, S. Q. Newton, M. Ramek, V. K. W. Cheng, F. A. Momany, and L. Schafer, *J. Am. Chem. Soc.*, **114**, 5369, (1992).

(29) J. J. Urban and G. R. J. Famini, *J. Comp. Chem.*, **14**, 353, (1993).

(30) W. D. Cornell, P. Cieplak, P. A. Kollman, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, and J. W. Caldwell, manuscript in preparation.

(31) A. E. Howard, P. Cieplak, and P. A. Kollman, in press J. Comp. Chem. (1994).

(32) W. D. Cornell, J. W. Caldwell, and P. A. Kollman, manuscript in preparation.

Figure Captions

Figure 1. The scheme used for fitting the central amino acids. Lagrange constraints are used to define three residues of net integral charge. Blocking groups are neutral.

Figure 2. The scheme used for fitting the N- and C-terminal amino acids. Lagrange constraints define the appropriate blocking group as neutral. The appropriate charged group is spliced onto the other end of the amino acid by defining a Lagrange constraint which forces the sum of the charges on all atoms within the two boxed groups to sum to zero. The sums of the charges within each boxed group are then equal and opposite, so that the net charge on the charged end group has the same charge as the group that it replaces, ensuring the appropriate net integral charge on the resulting N- or C-terminal charged amino acid.

Figure 1

Figure 2

Table I. AMBER minimized conformations used for each amino acid.[a]

amino acid	backbone	phi	psi	chi1	chi2	chi3	chi4
Val	c5	208	154	61 (g-)			
	aR	-60*	-40*	181 (t)			
Ser	c5	197	174	47 (g-)	299 (g+)		
	aR	-60*	-40*	187 (t)	43 (g-)		
Asn	c5	217	160	297(g+)	263		
	aR	-60*	-40*	180 (t)	269		
Cys	c5	206	141	172 (t)	179 (t)		
	aR	-60*	-40*	284 (g+)	303 (g+)		
Cyx	c5	218	157	290 (g+)	303 (g+)	276	
	aR	-60*	-27	293 (g+)	302 (g+)	276	
Asp	c5	218	160	299 (g+)	85		
	aR	-60*	-40*	160 (t)	92		
Thr	c5	220	165	46 (g-)	300 (g+)		
	aR	-60*	-40*	333 (g+)	185 (t)		
Ile	c5	207	154	184 (t)	62 (g-)		
	aR	-60*	-40*	300 (g+)	176 (t)		
Leu	c5	209	150	181 (t)	63 (g-)		
	aR	-60*	-40*	302 (g+)	179 (t)		
Hid	c5	217	162	314 (g+)	94 (g-)	270*	
	aR	-60*	-40*	180* (t)	266 (g-)	90*	
Hie	c5	217	161	279 (g+)	88 (g-)	271*	
	aR	-60*	-40*	186 (t)	276 (g-)	95	
Hip	c5	208	145	312 (g+)	88 (g-)	271*	
	aR	-60*	-40*	184 (t)	271 (g-)	82	
Phe	c5	217	160	308 (g+)	119		
	aR	-60*	-40*	185 (t)	264 (g-)	82	

Table I. AMBER minimized conformations used for each amino acid, cont'd.

amino acid	backbone	phi	psi	chi1	chi2	chi3	chi4
Tyr	c5	217	160	299 (g+)	96	177	
	aR	-60*	-40*	180 (t)	256	357	
Trp	c5	216	161	299 (g+)	100		
	aR	-60*	-40*	191 (t)	282		
Met	c5	205*	153*	187 (t)	176 (t)	180 (t)	
	aR	-60*	-40*	298 (g+)	306 (g+)	302 (g+)	
Pro(t)	beta	295*	150*	28	326		
	alpha	299*	325*	334	35		
Pro(c)	beta	305	143	334	35		
	alpha	274*	1*	33	324		
Glu	c5	206*	153*	297 (g+)	189 (t)	65	
	aR	-60*	-40*	192* (g+)	88	90* (g+)	
Gln	c5	205	153	192 (t)	177 (t)	120	
	aR	-60*	-40*	308 (g+)	310 (g+)	123	
Arg	c5	214	145	298 (g+)	186 (t)	181 (t)	177 (t)
	aR	-60*	-40*	179 (t)	160 (t)	291 (g+)	283 (g+)
Lys	c5	205	153	188 (g+)	180 (t)	198 (t)	73 (g-)
	aR	-60*	-40*	293 (g+)	189 (t)	177 (t)	183 (t)

[a] Asterisk indicates that dihedral was constrained.

Table II. Charges for alanyl dipeptide as a function of charge model.

	STD ESP	conformation(s) used in RESP charge fit						
	$C7_{eq}/C7_{ax}$ $/C5/\alpha R$	$C7_{eq}$	$C7_{ax}$	C5	αR	$C5/\alpha R$	$C7_{eq}/C5$ $/\alpha R$	$C7_{eq}/C7_{ax}$ $/C5/\alpha R$
C (N-term CH$_3$)	-0.603	-0.491	-0.340	-0.321	-0.258	-0.314	-0.359	-0.345
H (N-term CH$_3$)	0.162	0.141	0.106	0.094	0.078	0.092	0.104	0.102
C	0.864	0.701	0.611	0.702	0.592	0.679	0.686	0.664
O	-0.626	-0.607	-0.564	-0.600	-0.541	-0.577	-0.585	-0.575
N	-0.730	-0.499	-0.522	-0.526	-0.431	-0.526	-0.520	-0.512
H	0.361	0.316	0.323	0.294	0.288	0.305	0.309	0.310
C (α)	0.246	0.074	0.079	0.048	-0.030	0.013	0.038	0.050
H (α)	0.032	0.071	0.076	0.073	0.114	0.090	0.080	0.076
C (β-CH$_3$)	-0.330	-0.226	-0.188	-0.129	-0.086	-0.083	-0.122	-0.148
H (β-CH$_3$)	0.094	0.073	0.071	0.052	0.037	0.038	0.047	0.056
C	0.632	0.598	0.497	0.555	0.551	0.593	0.610	0.576
O	-0.588	-0.565	-0.563	-0.531	-0.550	-0.551	-0.562	-0.563
N	-0.506	-0.507	-0.350	-0.432	-0.376	-0.436	-0.466	-0.436
H	0.326	0.329	0.300	0.318	0.261	0.299	0.302	0.298
C (C-term CH$_3$)	-0.196	-0.055	-0.214	-0.351	-0.245	-0.295	-0.209	-0.197
H (C-term CH$_3$)	0.116	0.073	0.107	0.154	0.122	0.139	0.116	0.110

Table III. Alanyl dipeptide conformational energies calculated using charges derived from single and multiple conformations fits. [a]

| conformation | E(QM) | conformation(s) used in RESP charge fit | | | | | | | STD / ESP |
		$C7_{eq}$	$C7_{ax}$	C5	α_R	C5/α_R	$C7_{eq}$/C5 /α_R	$C7_{eq}$/$C7_{ax}$ /C5/α_R	$C7_{eq}$/$C7_{ax}$ /C5/α_R
$C7_{eq}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$C7_{ax}$	2.1	0.9	0.9	0.8	0.9	0.7	1.0	0.9	1.1
C5	1.5	4.2	3.7	4.2*	2.0	3.5	3.4	3.2	5.3*
α_R	3.9	7.0	7.3	6.6	5.0	5.6	5.8	6.1	5.9

[a] Energies marked with an asterisk were calculated using dihedral constraints which were necessary to keep the structure in the desired local minimum.

Table IV. Phi and psi values (in degrees) of molecular mechanically minimized conformations as a function of charge model. a

conformation		QM opt	conformation(s) used in RESP charge fit							STD ESP
			C7eq	C7ax	C5	αR	C5/αR	C7eqC5 /αR	C7eqC7ax /C5/αR	C7eqC7ax /C5/αR
C7eq	φ	274	285	284	284	281	283	284	283	286
	ψ	79	67	67	67	70	67	67	68	65
C7ax	φ	75	65	66	66	68	66	66	66	63
	ψ	305	300	301	301	298	301	300	300	302
C5	φ	203	218	213	*	214	216	216	215	*
	ψ	160	148	146	*	150	145	147	147	*

a Asterisks are shown for conformations where the structure did not stay within the desired local minimum upon optimization.

Table V. Classical dipole moments (in Debyes) of molecular mechanically minimized conformations as a function of charge model. [a]

conformation	QM opt	conformation(s) used in RESP charge fit							STD ESP
		C7$_{eq}$	C7$_{ax}$	C5	α$_R$	C5/α$_R$	C7$_{eq}$/C5 /α$_R$	C7$_{eq}$/C7$_{ax}$ /C5/α$_R$	C7$_{eq}$/C7$_{ax}$ /C5/α$_R$
C7$_{eq}$	2.90	3.71	3.78	3.36	3.16	3.29	3.41	3.37	3.43
C7$_{ax}$	3.89	4.03	4.11	3.65	3.57	3.62	3.74	3.71	3.69
C5	2.58	2.36	2.35	2.25*	2.21	2.19	2.20	2.38	2.58*
α$_R$	6.58	7.97	8.23	7.56	7.35	7.42	7.56	7.71	7.57
RMS(μ_{MM}-μ_{QM})		0.82	0.95	0.34	0.47	0.52	0.59	0.63	0.57

[a] Values marked with an asterisk were calculated from a constrained geometry since the structure did not optimize to the proper minimum.

Table VI. Classical dipole moments (in Debyes) of quantum mechanically minimized conformations as a function of charge model. [a]

conform	QM opt	conformation(s) used in RESP charge fit							STD ESP
		C7eq	C7ax	C5	αR	C5/αR	C7eq/C5 /αR	C7eq/C7ax /C5/αR	C7eq/C7ax /C5/αR
C7eq	2.90	2.86/2.94	2.94	2.64	2.73	2.63	2.74	2.87/2.82	2.87/2.86
C7ax	3.89	3.87	3.87/3.88	3.47	3.63	3.50	3.62	3.84/3.70	3.84/3.74
C5	2.58	2.67	2.73	2.56/2.67	2.43	2.60	2.56	2.56/2.53	2.57/2.93
αR	6.58	7.52	7.76	7.05	6.52/6.86	6.90	7.06	6.58/7.24	6.59/7.12
RMS(μMM -μQM)		0.47	0.60	0.34	0.22	0.29	0.29	0.35	0.33

[a] When two numbers are given, they represent the dipole moments after the first and second stage of the RESP fit, respectively.

Table VII. Net charge on each residue in alanyl dipeptide from various charge models.

	conformation(s) used in RESP charge fit							STD ESP
	$C7_{eq}$	$C7_{ax}$	$C5$	α_R	$C5/\alpha_R$	$C7_{eq}/C5$ $/\alpha_R$	$C7_{eq}/C7_{ax}$ $/C5/\alpha_R$	$C7_{eq}/C7_{ax}$ $/C5/\alpha_R$
N-term. acetyl	0.027	0.026	0.064	0.027	0.062	0.054	0.048	0.129
ala	-0.012	-0.083	-0.060	-0.032	-0.046	-0.028	-0.044	-0.086 ($C7_{eq}$)
C-term.N-methyl	-0.014	0.057	-0.004	0.005	-0.017	-0.026	-0.005	-0.043
								0.127
								-0.081 ($C7_{ax}$)
								-0.045
								0.121
								-0.094 ($C5$)
								-0.028
								0.126
								-0.096 (α_R)
								-0.030

Table VIII. Glycyl and alanyl dipeptide conformational energies calculated using RESP charges derived with and without multiple lagrange constraints and constrained equivalence of the two amide groups.[a,b]

			E(MM) Charge Model[c,d]							
			1 res no am eq.		3 res no am eq.		1 res am eq.		3 res am eq.	
	E(QM)									
Conf	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala
$C7/7_{eq}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$C7_{ax}$		2.1		0.8		0.8		0.9		0.8
C5	2.0	1.5	4.6	3.4	3.8	3.6	4.8	3.2	3.8	2.4
α_R	4.0	3.9	6.6	5.6	7.3	6.0	6.6	5.6	7.3	5.6

[a] Energies in kcal/mole. [b] Charges derived from multiple conformation fits of c5 and α_R conformations of alanyl dipeptide. Corresponding atoms constrained to have equivalent charges between the two conformations, with the exception of the terminal methyl groups. Charges for those methyl groups taken from the C5 conformation. Three lagrange constraints applied to achieve neutrality of the acetyl blocking group, the N-methyl blocking group, and the central amino acid residue. [c] "1 res" and "3 res" signify that the charges were fit to produce either 1 or 3 neutral residues from the molecule. [d] "no am eq" and "am eq" signify that the two amide groups were not constrained to have equivalent charges, or that they were constrained.

Table IX. Glycyl and alanyl dipeptide conformational energies calculated using RESP charges derived from quantum mechanically (HF/6-31G**) and molecular mechanically (AMBER/Weiner *et al.* force field [6,7,16]) optimized geometries.[a,b]

| | | | | E(MM) | | |
| | E(QM) | | QM opt geom | | MM opt geom | |
Conf	GLY	ALA	GLY	ALA	GLY	ALA
C7/C7$_{eq}$	0.0	0.0	0.0	0.0	0.0	0.0
C7$_{ax}$		2.1		0.8		0.9
C5	2.0	1.5	4.6	3.4	4.3	3.5
α$_R$	4.0	3.9	6.6	5.6	7.7	6.6

[a] Energies in kcal/mole. [b] Charges derived from multiple conformation fits of either glycyl or alanyl dipeptides using both C5 and α$_R$ conformations. Corresponding atoms constrained to have equivalent charges between the two conformations, with the exception of the terminal methyl groups. Charges for those methyl groups taken from the C5 conformation. A single Lagrange constraint was imposed resulting in overall neutrality of the molecule and the two amides groups were allowed to optimize independently and have different charges.

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields

	Weiner *et al.*	Cornell *et al.*		Weiner *et al.*	Cornell *et al.*
BACKBONE (neutral residue)			CG2	-0.191	-0.244
N	-0.463	-0.416	HG21	0.065	0.064
H	0.252	0.272	HG22	0.065	0.064
C	0.616	0.597	HG23	0.065	0.064
O	-0.504	-0.568	OG1	-0.550	-0.676
			HG1	0.310	0.410
GLY					
CA	0.035	-0.025	**ASN**		
HA2	0.032	0.070	CA	0.035	0.014
HA3	0.032	0.070	HA	0.048	0.105
			CB	-0.086	-0.204
ALA			HB2	0.038	0.080
CA	0.035	0.034	HB3	0.038	0.080
HA	0.048	0.082	CG	0.675	0.713
CB	-0.098	-0.183	OD1	-0.470	-0.593
HB1	0.038	0.060	ND2	-0.867	-0.919
HB2	0.038	0.060	HD21	0.344	0.420
HB3	0.038	0.060	HD22	0.344	0.420
SER			**GLN**		
CA	0.035	-0.025	CA	0.035	-0.003
HA	0.048	0.084	HA	0.048	0.085
CB	0.018	0.212	CB	-0.098	-0.004
HB2	0.119	0.035	HB2	0.038	0.017
HB3	0.119	0.035	HB3	0.038	0.017
OG	-0.550	-0.655	CG	-0.102	-0.065
HG	0.310	0.428	HG2	0.057	0.035
			HG3	0.057	0.035
THR			CD	0.675	0.695
CA	0.035	-0.039	OE1	-0.470	-0.609
HA	0.048	0.101	NE2	-0.867	-0.941
CB	0.170	0.365	HE21	0.344	0.425
HB	0.082	0.004	HE22	0.344	0.425

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*		Weiner *et al.*	Cornell *et al.*
CYS			**MET**		
CA	0.035	0.021	CA	0.035	-0.024
HA	0.048	0.112	HA	0.048	0.088
CB	-0.060	-0.123	CB	-0.151	0.034
HB2	0.038	0.111	HB2	0.027	0.024
HB3	0.038	0.111	HB3	0.027	0.024
SG	0.827	-0.312	CG	-0.054	0.002
HG	0.135	0.193	HG2	0.065	0.044
LP1	-0.481	-----	HG3	0.065	0.044
LP2	-0.481	-----	SD	0.737	-0.274
			LP1	-0.381	-----
CYX			LP2	-0.381	-----
CA	0.035	0.043	CE	-0.134	-0.054
HA	0.048	0.077	HE1	0.065	0.068
CB	-0.098	-0.079	HE2	0.065	0.068
HB2	0.050	0.091	HE3	0.065	0.068
HB3	0.050	0.091			
SG	0.824	-0.108			
LP1	-0.405	-----			
LP2	-0.405	-----			

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*		Weiner *et al.*	Cornell *et al.*
VAL			**ILE**		
CA	0.035	-0.088	CA	0.035	-0.060
HA	0.048	0.097	HA	0.048	0.087
CB	-0.012	0.299	CB	-0.012	0.130
HB	0.024	-0.030	HB	0.022	0.019
CG1	-0.091	-0.319	CG2	-0.085	-0.320
HG11	0.031	0.079	HG21	0.029	0.088
HG12	0.031	0.079	HG22	0.029	0.088
HG13	0.031	0.079	HG23	0.029	0.088
CG2	-0.091	-0.319	CG1	-0.049	-0.043
HG21	0.031	0.079	HG12	0.027	0.024
HG22	0.031	0.079	HG13	0.027	0.024
HG23	0.031	0.079	CD1	-0.085	-0.066
			HD11	0.028	0.019
LEU			HD12	0.028	0.019
CA	0.035	-0.052	HD13	0.028	0.019
HA	0.048	0.092			
CB	-0.061	-0.110			
HB2	0.033	0.046			
HB3	0.033	0.046			
CG	-0.010	0.353			
HG	0.031	-0.036			
CD1	-0.107	-0.412			
HD11	0.034	0.100			
HD12	0.034	0.100			
HD13	0.034	0.100			
CD2	-0.107	-0.412			
HD21	0.034	0.100			
HD22	0.034	0.100			
HD23	0.034	0.100			

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*		Weiner *et al.*	Cornell *et al.*
PHE					
CA	0.035	-0.002	HE2	0.102	0.166
HA	0.048	0.098	CD2	-0.002	-0.191
CB	-0.100	-0.034	HD2	0.064	0.170
HB2	0.108	0.030			
HB3	0.108	0.030	**TRP**		
CG	-0.100	0.012	CA	0.035	-0.028
CD1	-0.150	-0.126	HA	0.048	0.112
HD1	0.150	0.133	CB	-0.098	-0.005
CE1	-0.150	-0.170	HB2	0.038	0.034
HE1	0.150	0.143	HB3	0.038	0.034
CZ	-0.150	-0.107	CG	-0.135	-0.142
HZ	0.150	0.130	CD1	0.044	-0.164
CE2	-0.150	-0.170	HD1	0.093	0.206
HE2	0.150	0.143	NE1	-0.352	-0.342
CD2	-0.150	-0.126	HE1	0.271	0.341
HD2	0.150	0.133	CE2	0.154	0.138
			CZ2	-0.168	-0.260
TYR			HZ2	0.084	0.157
CA	0.035	-0.001	CH2	-0.077	-0.113
HA	0.048	0.088	HH2	0.074	0.142
CB	-0.098	-0.015	CZ3	-0.066	-0.197
HB2	0.038	0.030	HZ3	0.057	0.145
HB3	0.038	0.030	CE3	-0.173	-0.239
CG	-0.030	-0.001	HE3	0.086	0.170
CD1	-0.002	-0.191	CD2	0.146	0.124
HD1	0.064	0.170			
CE1	-0.264	-0.234			
HE1	0.102	0.166			
CZ	0.462	0.323			
OH	-0.528	-0.558			
HH	0.334	0.399			
CE2	-0.264	-0.234			

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*			Weiner *et al.*	Cornell *et al.*
HID				**PRO**		
CA	0.035	0.019		N	-0.229	-0.255
HA	0.048	0.088		CA	0.035	-0.027
CB	-0.098	-0.046		HA	0.048	0.064
HB2	0.038	0.040		CB	-0.115	-0.007
HB3	0.038	0.040		HB2	0.061	0.025 .
CG	-0.032	-0.027		HB3	0.061	0.025
ND1	0.146	-0.381		CG	-0.121	0.019
HD1	0.228	0.365		HG2	0.063	0.021
CE1	0.241	0.206		HG3	0.063	0.021
HE1	0.036	0.139		CD	-0.012	0.019
NE2	-0.502	-0.573		HD2	0.060	0.039
CD2	0.195	0.129		HD3	0.060	0.039
HD2	0.018	0.115		C	0.526	0.590
				O	-0.500	-0.575
HIE						
CA	0.035	-0.058				
HA	0.048	0.136				
CB	-0.098	-0.007				
HB2	0.038	0.037				
HB3	0.038	0.037				
CG	0.251	0.187				
ND1	-0.502	-0.543				
CE1	0.241	0.164				
HE1	0.036	0.144				
NE2	-0.146	-0.280				
HE2	0.228	0.334				
CD2	-0.184	-0.221				
HD2	0.114	0.186				

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*		Weiner *et al.*	Cornell *et al.*
BACKBONE (positive residue)					
N	-0.463	-0.348	HG2	0.074	0.029
H	0.252	0.275	HG3	0.074	0.029
C	0.616	0.734	CD	-0.228	0.049
O	-0.504	-0.589	HD2	0.133	0.069
			HD3	0.133	0.069
LYS			NE	-0.324	-0.530
CA	0.035	-0.240	HE	0.269	0.346
HA	0.048	0.143	CZ	0.760	0.808
CB	-0.098	-0.009	NH1	-0.624	-0.863
HB2	0.038	0.036	HH11	0.361	0.448
HB3	0.038	0.036	HH12	0.361	0.448
CG	-0.160	0.019	NH2	-0.624	-0.863
HG2	0.116	0.010	HH21	0.361	0.448
HG3	0.116	0.010	HH22	0.361	0.448
CD	-0.180	-0.048			
HD2	0.122	0.062	**HIP**		
HD3	0.122	0.062	CA	0.035	-0.135
CE	-0.038	-0.014	HA	0.048	0.121
HE2	0.098	0.114	CB	-0.098	-0.041
HE3	0.098	0.114	HB2	0.086	0.081
NZ	-0.138	-0.385	HB3	0.086	0.081
HZ1	0.294	0.340	CG	0.058	-0.001
HZ2	0.294	0.340	ND1	-0.058	-0.151
HZ3	0.294	0.340	HD1	0.306	0.387
			CE1	0.114	-0.017
ARG			HE1	0.158	0.268
CA	0.035	-0.264	NE2	-0.058	-0.172
HA	0.048	0.156	HE2	0.306	0.391
CB	-0.080	-0.001	CD2	-0.037	-0.114
HB2	0.056	0.033	HD2	0.153	0.232
HB3	0.056	0.033			
CG	-0.103	0.039			

Table X. Comparison of Charges from Weiner *et al.* and Cornell *et al.* Force Fields, cont'd

	Weiner *et al.*	Cornell *et al.*
BACKBONE (negative residue)		
N	-0.463	-0.516
H	0.252	0.294
C	0.616	0.537
O	-0.504	-0.582
ASP		
CA	0.035	0.038
HA	0.048	0.088
CB	-0.398	-0.030
HB2	0.071	-0.012
HB3	0.071	-0.012
CG	0.714	0.799
OD1	-0.721	-0.801
OD2	-0.721	-0.801
GLU		
CA	0.035	0.040
HA	0.048	0.111
CN	-0.184	0.056
HB2	0.092	-0.017
HB3	0.092	-0.017
CG	-0.398	0.014
HG2	0.071	-0.043
OHG3	0.071	-0.043
CD	0.714	0.805
OE1	-0.721	-0.819
OE2	-0.721	-0.819

Chapter 5

A Second Generation Force Field for the Simulation of

Proteins and Nucleic Acids

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

by

Wendy D. Cornell, [1]

Piotr Cieplak, [2]

Christopher I.Bayly, [3]

Ian R. Gould, [4]

Kenneth M. Merz, Jr., [5]

David M. Ferguson, [6]

David C. Spellmeyer, [7]

Thomas Fox,

James W. Caldwell,

and

Peter A. Kollman [1*]

Contribution from The Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA

[1] Graduate Group in Biophysics, University of California at San Francisco.

[2] Permanent Address: Department of Chemistry, University of Warsaw, Pasteura, 1, 02-093, Warsaw, Poland.

[3] Current Address: Merck Frosst Canada, Inc., C.P. 1005 Pointe Claire-Dorval, Quebec, H9R 4P8, Canada.

[4] Current Address: Department of Chemistry, University of Manchester, Lancs, M13 9PL, U.K.

[5] Current Address: Department of Chemistry, The Pennsylvania State University, State College, PA 16802.

[6] Current Address: Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455.

[7] Current Address: Chiron Corporation, Emeryville, CA 94608.

[*] To whom correspondence and reprint requests should be addressed.

Abstract

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

Introduction

The application of computer based models using analytical potential energy functions within the framework of classical mechanics has proven to be an increasingly powerful tool to study molecules of biochemical and organic chemical interest. These applications of molecular mechanics have employed energy minimization, molecular dynamics, and Monte Carlo methods to move on the analytical potential energy surfaces. Such methods have been used to study a wide variety of phenomena including intrinsic strain of organic molecules, structure and dynamics of simple and complex liquids, thermodynamics of ligand binding to proteins, and conformational transitions in nucleic acids. In principle, they are capable of giving insight into the entire spectrum of non-covalent interactions between molecules, and, when combined with quantum mechanical electronic structure calculations to model covalent bonding changes, essentially all molecular reactions and interactions. Given their importance, much effort has gone into consideration of both the functional form and the parameters that must be established in order to apply such analytical potential energy functions (or "force fields").

In the area of organic molecules, the book by Allinger and Burkert [1] provides a thorough review pre-1982 and the subsequent further development of the MM2 [2] and MM3 [3] force fields by Allinger and co-workers has dominated the landscape in this area. The number of force fields developed for application to biologically interesting molecules is considerably greater, probably because of the greater complexity of the interactions which involve ionic and polar groups in aqueous solution and the difficulty of finding an unequivocal test set to evaluate such force fields. Many of these force fields developed prior to 1987 are described briefly by McCammon and Harvey [4].

Given the complexities and subjective decisions inherent in such biological force fields, we have attempted to put the development of the force field parameters on a more *explicitly*

stated algorithmic basis than done previously, so that the force field could be extended by ourselves and others to molecules and functional groups not considered in the initial development. This is important, because, if the assumptions, approximations, and inevitable imperfections in a force field are at least known, one can strive for some cancellation of errors.

Approximately a decade ago, Weiner *et al.* [5,6] developed a force field for proteins and nucleic acids which has been widely used. Important independent tests of this force field were performed by Pavitt and Hall for peptides [7] and Nilsson and Karplus[8] for nucleic acids and it was found to be quite effective. Nonetheless, it was developed in the era before one could routinely study complex molecules in explicit solvent. Weiner *et al.* attempted to deal with this issue by showing that the same force field parameters could be effectively used both without explicit solvent (using a distance dependent dielectric constant $(\varepsilon=R_{ij})$) and with explicit solvent $(\varepsilon=1)$ on model systems. Further support for this approach was provided by molecular dynamics simulations of proteins [9,10,11] and DNA [12,13] which compared the implicit and explicit solvent representations.

As computer power has grown, it has become possible to carry out more realistic simulations which employ explicit solvent representations. It is therefore appropriate that any new force field for biomolecules focus on systems modelled in the presence of an explicit solvent representation. This approach has been pioneered by Jorgensen and co-workers in their OPLS (Optimized Potentials for Liquid Simulations) model [14]. In particular, the development of parameters which reproduce the enthalpy and density of neat organic liquids as an essential element ensures the appropriate condensed phase behavior. The OPLS non-bonded parameters have been combined with the Weiner *et al.* bond, angle and dihedral parameters to create the OPLS/Amber force field for peptides and proteins [15], which has also been effectively used in many systems [16].

We have been influenced by the OPLS philosophy of balanced solvent-solvent and solute-solvent interactions in our thoughts about a second generation force field to follow that of Weiner *et al.* [5,6]. The Weiner *et al.* force field used quantum mechanical calculations to derive electrostatic potential (esp) fit atomic centered charges, whereas the OPLS charges were derived empirically, using mainly the liquid properties as a guide. For computational expediency, Weiner *et al.* relied principally on the STO-3G basis set for their charge derivation. This basis set leads to dipole moments that are approximately equal to or smaller than the gas phase moment, but tends to underestimate quadrupole moments. Thus, it is not well *balanced* with the commonly used water models (SPC/E [17], TIP3P [18], TIP4P [18]) which have dipole moments that are about 20% higher than the gas phase value for water. These water models, which have empirically derived charges, include condensed phase electronic polarization implicitly. Kuyper *et al.* [19] suggested that the logical choice of a basis set for deriving esp-fit partial charges for use in condensed phases is the 6-31G* basis set, which uniformly overestimates molecular polarity. Standard ESP charges derived with that basis set were shown to lead to excellent relative free energies of solvation for benzene, anisole, and trimethoxyanisole [19].

A 6-31G* based esp-fit charge model, like the OPLS model, is capable of giving an excellent reproduction of condensed phase *inter* molecular properties such as liquid enthalpies and densities and free energies of solvation [20]. A major difference between our two force fields and most others is the magnitude of the charges on hydrocarbons. For example, 6-31G* standard ESP charges derived from the trans conformation of butane have values of -0.344 for the methyl carbon and 0.078 for the methyl hydrogen. In both cases, however, the carbon and hydrogen charges offset each other, resulting in small net charges on the methyl groups of -0.110 and -0.059 for the trans and gauche charges, respectively. Furthermore, free energy perturbation calculations involving the perturbation

of methane with standard ESP charges (q_C= -0.464 and q_H= 0.116) to methane with charges of 0.0 in solution yields essentially no change in free energy [21]. The standard ESP charges also result in conformational energies for butane which are in reasonable agreement with experiment, when used with a 1-4 electrostatic scale factor of 1/1.2 [20].

Nevertheless, the 6-31G* standard ESP charges are less than ideal for two reasons. First, when charges generated using different conformations of a molecule are compared, there is often considerable variation seen. This was demonstrated by Williams who studied the conformational variation of esp-fit charges in alanyl dipeptide for 12 different conformations [22]. Butane is another example, where charges from the gauche conformation have values of -0.197 and 0.046 for the methyl carbon and hydrogen, respectively. Another example is propylamine, which was studied at length by Cornell *et al.* [20]. Five low energy conformations can be identified for propylamine and the 6-31G* standard ESP charges calculated for each conformation show significant variation. The average and standard deviation for the charge on a given atom over the five conformations are: α-carbon q_{av} = 0.339 and σ = 0.059, β-carbon q_{av} = 0.033 and σ = 0.060, and γ-carbon q_{av} = -0.205 and σ = 0.146. This inconsistency is potentially problematic in terms of deriving other force field parameters which may be sensitive to the variation. Furthermore, it reduces the reproducibility of a particular calculation, which is not a problem in other force fields where the charges are assigned empirically.

The second reason that the 6-31G* standard ESP charges are less than ideal is that the charges on "buried" atoms (such as the sp^3 carbons described above for butane and propylamine) are statistically underdetermined and often assume unexpectedly large values for nonpolar atoms. Bayly *et al.* [23] found that the electrostatic potential of methanol could be fit almost equally well using either the standard ESP charges determined by the linear least

squares fit or an alternative set of charges derived with the methyl carbon constrained to have a much smaller value.

Considering the problems associated with the standard ESP charge model, it might seem tempting to adopt the OPLS approach of empirically derived charges. However, any empirically derived charge model cannot easily describe transition states and excited states, as can an electrostatic potential fit model. Furthermore, the conformational dependence of N-methyl acetamide (NMA) is better represented with an esp-fit model [24]. Finally, the requirement of Monte Carlo calculations on requisite liquids including appropriate fragments makes it more problematic to make an empirical charge model that will cover most or all of chemical/biochemical functionality.

Given the above mentioned deficiencies in the standard ESP model, along with the desire to retain the general strategy of fitting charges to the electrostatic potential, Bayly *et al.* [23] were motivated to develop the RESP (restrained esp-fit) charge model. The RESP model still involves a least squares fit of the charges to the electrostatic potential, but with the addition of hyperbolic restraints on charges on non-hydrogen atoms. These restraints serve to reduce the charges on atoms which *can* be reduced without impacting the fit, such as buried carbons. The final RESP model requires a two-stage fit, with the second stage needed to fit methyl groups which require equivalent charges on hydrogen atoms which are not equivalent by molecular symmetry. The new charge model has been shown to perform well at reproducing interaction energies and free energies of solvation. When used with a 1-4 electrostatic scale factor of 1/1.2 (as opposed to the scale factor of 1/2 employed by Weiner *et al.*), both the RESP (and standard ESP) charges also result in good conformational energies for many of the small molecules studied to date without the necessity for an elaborate dihedral potential [20].

In addition to the new charges which have been tailored for condensed phase simulations, new van der Waals (VDW) parameters have also been adopted and developed which are optimized for reproducing liquid properties. The VDW parameters in the Weiner *et al.* [5,6] force field are primarily a modification of a set originally proposed by Hagler-Euler-Lifson [25] which were fit to lattice energies and crystal structures of amides. The new VDW parameters for aliphatic and aromatic hydrogens take into account the effects of any vicinal electronegative atoms [26,27].

High level quantum mechanical data is now available on the conformational energies of the glycyl and alanyl dipeptides [28] and this data is critical for developing ϕ and ψ dihedral parameters for the peptide backbone. Because such high level data was unavailable at the time the Weiner *et al.* force field was developed, the ϕ and ψ parameters were left as 0.0 kcal/mole since the resulting molecular mechanical energies seemed to be in reasonable agreement with the best theoretical data available at that time. That force field led to conformational energies for glycyl dipeptide where the C5 extended conformation was about 1 kcal too high in energy and for alanyl dipeptide where the C5 conformation was nearly 2 kcal too high in energy but the $C7_{ax}$ conformation was about 1 kcal too low in energy. The error in the alanyl dipeptide $C7_{ax}$ energy is not critical since it is rarely found in proteins [29] (only in γ-turns), but the errors in the energies of the C5 conformations are more important since that is the conformation found in β-sheets. Any errors in the energies of the C5 conformations are multiplied by the length of the secondary structure. The new force field includes V_1, V_2, V_3, and V_4 dihedral parameters for ϕ and ψ which result in good agreement between the molecular mechanical and quantum mechanical energies of the dipeptides.

Finally, the benzene molecule as modelled by the Weiner *et al.* all-atom force field has been shown to possess excessive flexibility for out-of-plane distortions [30]. This was caused by the use of the V_2 potential derived for the united atom model. This underestimate of the

benzene V_2 parameter is noteworthy, because it not only affects the flexibility of benzene and benzene-like moieties, but it also affects the interpolation scheme used for determining the V_2 barriers for X-C-N-X and X-C-C-X dihedrals in conjugated rings. These V_2 parameters are determined by interpolating according to the bond length either between a pure single bond and a partial double bond (benzene) or between a partial double bond and a pure double bond. The excessive out-of-plane motion of benzene has been easily fixed by adjusting the V_2 parameter from 5.5 to 14.5 kcal/mole to match the experimental normal mode frequencies.

General Description of the Model

The model presented here can be described as "minimalist" in its functional form, with the bond and angles represented by a simple diagonal harmonic expression, the VDW interaction represented by a 6-12 potential, electrostatic interactions modelled by a Coulombic interaction of point charges, and dihedral energies represented (in most cases) with a simple set of parameters, often only specified by the two central atoms. Electrostatic and van der Waals interactions are only calculated between atoms in different molecules or for atoms in the same molecule separated by at least three bonds. Those non-bonded interactions separated by exactly three bonds ("1-4 interactions") are reduced by the application of a scale factor.

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(\eta\phi - \gamma)] +$$

$$\sum_{i<j} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}}]$$

Our assumption is that such a simple representation of bond and angle energies is adequate for modelling most unstrained systems. The goal of this force field is to accurately model conformational energies and intermolecular interactions involving proteins, nucleic acids, and

other molecules with related functional groups which are of interest in organic and biological chemistry.

A. Atom Types

The atom types employed are similar to those defined previously and are given in Table I. The one significant departure is the definition of new atom types for hydrogens bonded to carbons which are themselves bonded to one or more electronegative atoms. This is similar in spirit to the electronegativity based bond length correction used in MM2 and MM3.

B. Bond and Angle Parameters

The r_{eq}, θ_{eq}, K_r, and K_θ values were used as starting values and adjusted as necessary to reproduce experimental normal mode frequencies. These values were initially derived by fitting to structural and vibrational frequency data on small molecular fragments that make up proteins and nucleic acids. For example, in complex fragments such as the nucleic acid bases, the r_{eq} and θ_{eq} values have been taken from X-ray structural data, the K_r values determined by linear interpolation between pure single and double bond values using the observed bond distances and the K_θ value taken from vibrational analysis of simple sp^2 atom containing fragments such as benzene and NMA. That this approach was reasonably successful is supported by the reasonable agreement found in nucleic acid base vibrational analysis and suggested by the critical analysis of Halgren of the diagonal force constants used in different force fields [31].

One "difficulty" arose in the development of this new force field compared to that of Weiner *et al.* which was related to the switch to the 6-31G* basis set for charge derivation. With 6-31G* standard ESP charges and a 1-4 electrostatic scale factor of 1/1.2 rather than 1/2.0 (see below), we found that the exocyclic -NH$_2$ groups of the bases moved considerably away from their r_{eq} and θ_{eq} values upon energy minimization. This problem

was considerably reduced with RESP charges and a 1-4 electrostatic scale factor of 1/1.2, so we chose not to selectively increase the K_θ values around the -NH_2 group to force it to more "canonical" geometries.

In general, however, one might have resorted to a more complex optimization of r_{eq}, θ_{eq}, K_r, and K_θ to insure that the geometries of simple fragments were as close as possible to experiment *after energy minimization*, rather than taking r_{eq} and θ_{eq} from experiment and assuming little distortion would occur (which is generally the case, with the slight exception of the case of the -NH_2 groups noted above). We chose not to undertake a more time-consuming iterative self-consistent derivation of geometrical parameters, because of our assumption that any such errors which we were making were of much smaller consequence for accurately representing conformations and intermolecular interactions than the inaccuracies remaining in the dihedral and non-bonded (charge and VDW) parameters.

C. Dihedral Parameters

Weiner *et al.* [5,6] developed a limited set of general and specific dihedral parameters which were appropriate for the functionalities found in proteins and DNA and calibrated to adjust the energies of small model compounds. In this strategy, a dihedral parameter is optimized on the simplest molecule possible and then applied to larger and more complex molecules. This approach is in contrast to one employed by many other force field developers where the parameters are optimized to best reproduce the conformational energies of a large number of molecules.. An advantage of our approach is the lack of dependence of the resulting parameters on the particular molecules chosen for the test set.

For the most part, a minimalist approach has been retained with regards to dihedral parameters. For example, we have only a three-fold Fourier component (V_3) for dihedrals around-C-C bonds, with the exception of cases such as E-C-C-E' where E and E' are

electronegative atoms like O or F. In these cases, there is a "gauche" effect which stabilizes the gauche conformation over the trans and this can be modelled with a two-fold Fourier component (V_2). The rotation around phosphorus-ester bonds (CT-OS-P-OS) also requires a two-fold component. In these cases, we have been able to go beyond the Weiner *et al*. force field by making use of reasonably high level *ab initio* models (MP2/6-31G*) to fit the values of such V_2 Fourier components.

Two exceptions we made to the principle of adding extra Fourier terms to the dihedral energies only in the presence of a compelling physical basis. These exceptions are the dipeptide ψ and ϕ and the nucleoside χ dihedrals. Here we used Fourier components to try to reproduce as well as possible the relative energies of the alanyl and glycyl dipeptides and a model nucleoside fragment calculated at a high level of theory without the requirement of "a physical picture." An alternative approach would be to empirically adjust the atomic partial charges to achieve the same aim. Given the power of the RESP methodology for deriving atomic partial charges which lead to good representations of intermolecular interactions and the importance of maintaining an accurate balance between intra and intermolecular interactions, we chose to empirically adjust the terms in the Fourier series for ψ and ϕ as well as χ.

In our previous force field, the V_2 parameters for X-C-N-X and X-C-C-X involving sp^2 hybridized atoms in conjugated rings were determined by a two stage linear interpolation approach (according to bond length) between the known barriers of pure single, pure double and partial double bonded systems (benzene for X-C-C-X and NMA for X-C-N-X). We have used the same approach here, but have adjusted the V_2 term of benzene to more accurately describe its out-of-plane frequencies (Weiner *et al*. [5,6] had used the V_2 derived for a united atom model of benzene, which was significantly different). Table II presents the parameters used.

D. VDW Parameters

Given the success of the OPLS approach in modelling liquids, we have developed all-atom sp^3 carbon and aliphatic hydrogen VDW parameters by carrying out Monte Carlo simulations on CH_4, C_2H_6, C_3H_8 and C_4H_{10} liquids and empirically adjusting R* and ε for the C and H to reproduce the densities and enthalpies of vaporization of these liquids [37]. Such parameters have also been employed in calculations of relative free energies of solvation of CH_4, C_2H_6 and C_3H_8 [21,38]. We also derived VDW parameters for sp^2 C and aromatic H employing Monte Carlo simulations on benzene liquid and adjusting the R* and ε of these atoms to reproduce the density and enthalpy of liquid benzene [37]. At the time these parameters were developed, such all-atom parameters were unavailable for the OPLS force field. These Monte Carlo simulations were the first calculations carried out as part of the development of this new force field, and as such employed 6-31G* standard ESP charges. The electrostatic contribution for the n-alkanes was very small regardless of the charge model -- at most a few tenths of a kcal/mole. We note that the standard ESP charges for benzene (q_C= -0.145 and q_H= 0.145) accurately reproduce the quadrupole moment of that molecule.

We have taken most of the remaining VDW parameters from the OPLS model [15] -- sp^2 and sp^3 N; sp^2 O, ether ester (OS), hydroxyl (OH) and TIP3P water (OW) sp^3 oxygens; and sulfur (SH and S) -- since it has been optimized for reproducing liquid properties. The Weiner *et al.* [5,6] phosphorus (P) parameters were not re-optimized since that atom is most frequently found buried inside of four other heavy atoms.

The VDW model is minimalist as well, with sets of exceptions. A standard VDW parameter is used for a given atom and hybridization, e.g. all sp^2 carbons have the same VDW parameters. The only heavy atom exceptions are sp^3 O, where oxygens in water

s

A

h

p

H

b

h

H

Co

sen

R*

sen

$\underline{H}$ h

-0.

We h

nucle

depen

studie

electr

H2. an

reduce

(OW), alcohol (OH), and ether (OS) have slightly different parameters, as found in OPLS. We suspect that this is due to the use of a zero VDW radius on hydrogens bond to oxygen, so that an effectively larger R* is required for a water oxygen than alcohol than ether.

A significant departure has been made from the previous model in the treatment of hydrogens. The current model does not employ 10-12 hydrogen bonding H...X parameters, although these are still supported within the AMBER software. The original Hagler et al. [25] and OPLS approach [14,15] suggested a zero R* and ε for hydrogen binding hydrogens. Thus the TIP3P water model, has R* and ε equal to 0.0 for its hydrogen (HW). We opted not to develop a new water model, but to use the TIP3P one.

Hydrogen is unique in the periodic table in not having an inner shell of electrons. Consequently, it makes physical sense for its VDW radius, unlike other atoms, to be very sensitive to its bonding environment. This has been extensively analyzed for the hydrogen R* in X-C-H systems by Gough et al. and Veenstra et al., [26,27] who demonstrated the sensitivity of R* to the electron withdrawing properties of X. For example, a "normal" C-H has VDW R*=1.487 Å; whereas in CF3-H it is ~0.3 Å shorter and in CH3NH3+ it is ~0.4 Å shorter still.

We have employed the following approach here. A C-H has R*=1.487Å and, based on nucleic and base pairing energy minimization, an N-H has R*=0.6Å. This qualitative dependence on electronegativity makes physical sense. Based on the Veenstra et al. [27] studies we have chosen to reduce the R* on sp³ C-H atoms by 0.1 Å for each electronegative (O, N, F, S) substituent. The hydrogen atom types are then defined as H1, H2, and H3 for 1, 2 and 3 electronegative groups, respectively The hydrogen R* is reduced by 0.4 Å for each neighboring positively charged group (atom type HP). For sp²

C-H, R* has been reduced by 0.05 Å for each electronegative neighbor (atom types H4 and H5).

Given our retention of the simplicity of a 6-12 rather than a 6-exponential VDW representation, we have continued to reduce 1-4 VDW interactions since the 6-12 approximation and the lack of polarization in the model both will lead to exaggerated short range repulsion. It is difficult to determine the scale factor unambiguously so we have retained the value of 1/2.0 used by Weiner *et al.* [5,6]

D. Electrostatic Energies

In Cornell*et al.* [20] and Cieplak *et al.* [39], we have extensively analyzed the development of our electrostatic model, which relies on the use of 6-31G* derived electrostatic potentials, multiple molecules, multiple conformations, and the RESP fitting approach. The multiple molecules/conformations and RESP fitting all serve to reduce the problem of statistically under-determined charges on buried atoms. We have further validated these models in their ability to calculate liquid enthalpies and densities [40] and free energies of solvation [21] of the prototypal polar molecules methanol and NMA in good agreement with experiment. We have not used lone pairs on sulfur in the new force field, despite their importance in hydrogen bond directionality [5,6] because of the PDB analysis of Gregoret *et al.* which showed that neutral sulfur functions only extremely rarely as a proton acceptor in proteins [41].

The new RESP charge model employs a scale factor of 1/1.2 for 1-4 electrostatics, which was calibrated on 1,2-ethanediol and also performed well on tests on simple alcohols, amines, and butane. The RESP and standard ESP charge models were shown by Howard *et al.* to perform better than MM2 and MM3 in the conformational analysis of substituted

1,3 dioxanes [42], requiring only the addition of a single dihedral parameter optimized on 2,4-dioxapentane.

Methods

ESP and RESP charges were calculated from electrostatic potentials derived using the Gaussian 90 and Gaussian 92 programs [43]. These programs were also employed for *ab initio* calculations of conformational energies. All minimization and normal mode calculations reported for this work were carried out using the AMBER package [44]. Scale factors of 1/1.2 and 1/2 were applied to 1-4 electrostatic and VDW interactions, respectively.

Free energy perturbation calculations for perturbing methane thiol to methanol and dimethyl thioether to dimethyl ether were carried out using the AMBER program and the slow growth method [45]. Simulations were run for 200 ps with a time step of 2 fs. SHAKE [46] was applied to constrain all bonds and perturbed bonds were shrunk. Only the solution perturbation was carried out (with TIP3P water [18] and periodic boundary conditions) and the intramolecular components were not included. Calculations were carried out in both the forward and reverse directions. The PMF correction was included to account for the free energy change associated with perturbed bonds [47].

Free energy perturbation calculations for the perturbation of 9-methyl-adenine to methane were carried out using the SPASMS [48] module of the AMBER program and the windows method with the acceptance ratio [49] with the electrostatic and VDW perturbations decoupled. All intramolecular components were included. The gas phase electrostatic run was carried out with 11 windows with 5K of equilibration and 10K of collection. The gas phase VDW run was carried out with 51 windows with 1K of equilibration and 5K of collection. The solution perturbation was carried out with TIP3P

water and periodic boundary conditions. The electrostatic part of the solution calculation was carried out analogously to the gas phase electrostatic calculation. The VDW part of the solution calculation was carried out with 51 windows, 1K of equilibration and 4K of collection. A 9.0 A cut-off with no switch functions was employed for non-bonded interactions and the time step was 1 fs. The coupling constants were 0.2 ps (temperature) and 0.4 ps (pressure).

Molecular dynamics simulations of ubiquitin were carried out using the AMBER program. The simulations were carried out at 300 K with a time step of 1.5 fs and a non-bonded cut-off of 8.0 A. SHAKE was applied to bonds containing hydrogens.

Results

We begin the development of the force field with ethane, the fundamental unit for hydrocarbons. The general V_3 (X-CT-CT-X) dihedral was changed from 1.3 kcal/mole to 1.4 kcal/mole in order to reproduce the experimental barrier to rotation (Table III). Ethane charges have been shown to be particularly sensitive to the conditions of the esp fit [55]. Nonetheless, changing the charges on hydrogen from 0.0 to 0.1 changes the barrier only from 2.89 to 2.92 kcal/mole. In contrast to MM2/MM3 [2,3], only this general V_3 dihedral potential is used for hydrocarbons. As one can see in Table III, the conformational energies and structures are well represented for the simple model hydrocarbons with such an approach. At this point, we should note the difference of our approach from that of MM3 [3], where the rotational barrier in ethane is ~0.5 kcal/mole smaller than experiment. The parameters in MM3 were derived by fitting to a wide variety of data for hydrocarbons, whereas our approach is to start with ethane as the simplest model and add additional dihedral parameters in a conservative way. As one can see, the barriers and geometry of n-butane are well described with such a model, as is the energy to eclipse the first and second methyl group of propane with the methylene.

We next turn to the alcohols and ethers (Table IV). Here we use only two general V_3 dihedral, as in Weiner *et al.* [5,6], for X-CT-OH-X and X-CT-OS-X. This leads to essentially exact reproductions of the dihedral barriers in methanol and dimethyl ether. The cis-trans energy difference is about 0.5 kcal greater than that calculated by the Weiner *et al.* force field, however the Weiner *et al.* value matched the experimental data originally used. When these dihedral parameters are applied to methyl ethyl ether (MEE) and tetrahydrofuran (THF), one finds that a small V_2 (CT-CT-OS-CT) dihedral of 0.1 kcal/mole (Weiner *et al.* had such a parameter with magnitude 0.2 kcal/mole) leads to an excellent reproduction of the g/t energy difference in MEE and a slight preference for C_2 THF over Cs, as inferred from experiments. The calculations overestimate the barrier to planarity of THF, but not by as much as MM3.

We next turn to dimethyl phosphate, the model for the backbone of nucleic acids. We have carried out *ab initio* calculations (MP2/6-31G*//HF/6-31G*) on dimethyl phosphate in its g,g; g,t; and t,t conformations and adjusted the V_2 (OS-P-OS-CT) parameter to reproduce the (g,g)/(g,t) energy difference of 1.41 kcal/mole. These results are reported in Table V. The reoptimized V2 parameter has a value of 1.20 as opposed to the value of 0.75 determined by Weiner *et al.* with the V3 parameter of 0.25 left unchanged. Reasonable agreement with *ab initio* calculations and consensus structural values from x-ray data has been achieved. The normal mode frequencies calculated with such a model are also compared with those developed based on experimental frequencies of diethyl phosphate [66]. Given the difference in molecules, the agreement between calculation and experiment for the low frequency modes reported in Table V is acceptable.

The low frequency modes for the simple hydrocarbons, alcohols, ethers, and thio-compounds are presented in Table VI. The average error between the calculated and

experimental frequencies is 31 cm⁻¹ for the 36 low frequency examples where experiemental data is available, compared to an error of 21 cm⁻¹ with MM3. Again, it should be noted that our parameters have been optimized using this limited set of simple molecules whereas the test set of molecules used to derive the MM3 parameters is much larger.

Next to consider in the development of a force field for nucleic acids are the bases. Elsewhere, we have reported the hydrogen bond energies and structures of A:T and G:C pairs and these appear to be in good agreement with the highest level of *ab initio* data currently available [69]. However, a critical element in the development of planar functionalities such as the bases are the dihedral potentials for out-of-plane motion, as discussed by Weiner *et al.* As in the development of our previous force field, normal mode analysis of benzene and NMA are important. The results for the normal mode analyses applied to these molecules are presented in Table VII. We have readjusted the X-CA-CA-X V_2 value and the improper out-of-plane dihedral X-X-CA-HA to ensure correct representation of the lowest frequency modes of benzene, with the four lowest modes (<700 cm⁻¹) in good agreement with experiment.

We next turn to NMA, the model for the peptide backbone. With a few adjustments to the Weiner *et al.* [5,6] bonded parameters, the agreement with experiment for the six lowest frequency modes is again excellent. In NMA, a key adjustment was the V_1 (H-N-C-O) dihedral potential, which, given the change in electrostatic and non-bonded parameters from Weiner *et al.*, had to be modified from 0.65 to 2.00 kcal/mole to ensure that the in vacuo cis/trans NMA energy difference was ~2.3 kcal/mole.

The re-optimized X-CA-CA-X parameter was used to interpolate new V_2 dihedral potentials for X-C-N-X and X-C-C-X dihedrals in conjugated rings (Table VII). The

normal mode frequencies for the four nucleic acid bases-- guanine, adenine, cytosine, and thymine -- were then calculated. The calculated and experimental frequencies for modes ~<600 cm^{-1} are reported in Table VIII. The agreement is qualitatively reasonable; in particular, the cost of out-of-plane distortion is approximately correct in these lowest frequency modes.

We then proceeded to the study of a larger fundamental unit of nucleic acids, deoxy adenosine nucleoside (dA). Table IX presents the results of calculations of the energy of dA as a function of sugar pucker and the dihedral angles γ (C5'-O5'-C4-C3') and χ (O1'-C1'-N9-C4), using both a pure gas phase ($\varepsilon=1$) and an implicit solvent ($\varepsilon=4$) model. Although this force field is primarily intended for use with explicit solvent, calculations by Sun et al. on conformational free energies of 18-crown-6 suggest that a model with $\varepsilon=4$ provides an approximate and qualitatively reasonable representation of aqueous free energies [78].

Encouragingly, the C2' endo/C3' endo energy difference is 0.6-1.0 kcal/mole, in good agreement with experiment. The barrier between these conformations through the O1' endo conformation is ~1.9-2.9 kcal/mole, somewhat larger (and perhaps more realistic) than that found by Weiner et al. The barrier through O1' exo is not ε dependent and is ~5.9 kcal/mole, which is in reasonable agreement with what is known. Experimentally, it is known that a γ in the g$^+$ range is preferred for nucleosides in solution, followed by trans, with little g$^-$ observed. The relative conformational energies with $\varepsilon=4$ are quite consistent with this trend, whereas the gas phase values ($\varepsilon=1$) are not.

Finally, adenosine and deoxyadenosine are known to prefer the anti conformation over the syn conformation, but the syn conformation is low enough in free energy to be observable. The gas phase ($\varepsilon=1$) energy difference between anti and syn is very large, but the $\varepsilon=4$ value

is much more reasonable. However, we wished to assess the reasonability of our calculated energies as a function of χ with an *ab initio* model. We thus constructed a simple test case where adenine is attached to CH(OH)-CH3, with the dihedrals constrained to mimic the C2' endo conformation of a sugar ring (Figure 1) and carried out MP2/6-31G*//HF/6-31G* *ab initio* calculations as a function of χ with this model. As one can see from Table X, with no additional dihedral parameters, the energy difference between the syn and anti minima is significantly overestimated with our initial model. We thus chose to add explicit dihedrals (V_1 and V_2) around the glycosidic bond to bring the two minima into qualitative agreement. This has very little effect on the γ and sugar pucker energies, so only the values of the final parameter set are reported in Table IX.

We next turned to studies of peptide conformations. Table XI presents the local minima and Figures 2a and 2b the (ϕ,ψ) maps for glycyl and alanyl dipeptides. Here, as in the case of glycosidic χ, we were forced to add explicit dihedral parameters in order to reproduce the *ab initio* quantum mechanical energies for these models. As one can see, the agreement with high level *ab initio* data is very good for all but alanyl dipeptide C_{7ax} and glycyl dipeptide α_R. The ala C_{7ax} conformation is rarely found in proteins and gly occurs relatively infrequently in α-helices, due to the loss of conformational entropy, so these conformations were reasonable ones in which to tolerate any error.

One of the important features in our force field is the attempt to reproduce the solvation free energies of a representative set of molecules. In Table XII, we present such a representative set. As one can see, the absolute solvation free energy of methane is somewhat (0.5 kcal/mole) too large with our model, but the relative solvation free energies of methane, ethane and propane are within 0.3 kcal/mole of experiment. For our protypal polar molecules, methanol and NMA, the agreement with experimental solvation free energies is within ~0.5 kcal/mole. We wished also to assess the solvation free energies for

sulfur compounds and the relative solvation free energies of those are in reasonable agreement with experiment (again within 0.5 kcal/mole). The calculated free energy of 9-methyl adenine is a prediction, because there are no precise experiments [84], but the relative free energies of this force field and that of Weiner et al. [5,6] suggest that the experimental determination of this quantity would be of great interest. Turning to the ionic molecules, our results make clear that a typical two body additive force field will tend to overestimate ion solvation (when corrected for long range cut-off) unless its parameters are significantly modified, but fully non-additive calculations with exactly the same parameters reproduce experiment very well.

The results described above were obtained on model systems that were relatively very simple (neat liquids) and/or small (dipeptides and nucleosides). In order to test the performance of the new force field on a more complex system, we carried out an MD simulation of ubiquitin in water with periodic boundary conditions. The RMS difference was calculated for structures along the trajectory relative to the crystal structure [85] for (1) the backbone atoms and (2) all of the heavy atoms. These results were then compared to those obtained with the Weiner et al. [5,6] force field (Figure 3). The RMS values are reported only for the first 72 residues, since the four residues of the carboxy terminus were mobile. The behavior of the new force field presented here is better in two ways. First, the protein structure seems to have stabilized after 50 ps of simulation with the new force field, while the RMS deviation continues to increase throughout the trajectory with the Weiner et al. [5,6] force field. Second, the RMS deviation for all of the heavy atoms after 180 ps of simulation is about 2.0 A with the force field presented here and about 2.5 A with the Weiner et al. [5,6] force field. Alonso and Daggett have also reported the results of a long MD simulation of ubiquitin, and they found an RMS deviation of 1.4 A from the crystal structure [86]. This greater stability is likely due to the use of a shorter (shifted)

electrostatic cut-off in their force field, which is adequate for reproducing local structure but is unlikely to reproduce free energies very accurately.

Even closer agreement with a protein crystal structure has been obtained by York *et al.*, [11] who carried out a 1000 ps MD simulation of BPTI with the long range electrostatic forces of the crystal environment treated using the particle mesh Ewald method and the Weiner *et al.*. force field. With this model they obtained an RMS deviation from the crystal structure of 0.33 A for backbone atoms. These results serve to illustrate the difference between errors arising from the force field itself and those arising from its implementation in a given calculation. Currently, most MD simulations employ an 8 or 9 A cut-off for non-bonded interactions in order to reduce this rate determining part of the calculation. In systems where long-range electrostatics play an important role, this approximation is clearly inadequate. Although the Ewald method is only fully appropriate for periodic crystal systems, other methods also exist which allow for the more accurate treatment of long range electrostatics [87]. Thus, it appears that the way electrostatic interactions are handled is significantly more important than the detailed force field parameters in ensuring that a molecular dynamics trajectory stays near an experimental (x-ray or NMR) structure. We suggest, however, that comparing two force fields with the same cut-off protocol can be illustrative and we conclude, on that basis, that the new force field performs at least as well as, or slightly better than, that of Weiner *et al.* [5,6] for full solution simulations.

Discussion

We have presented the development and the description of a new force field for proteins, nucleic acids and organic molecules. Previously, we have attempted to give a coherent description of the underlying basis for the Weiner *et al.* force field [5,6], in order that it could be extended by others as well as ourselves for studies of molecular interactions and conformations. We should emphasize again that our goal is to describe molecular

conformational energies and structures as accurately as possible in condensed phases with a simple, transferable, and general model. This goal has framed our approach, which has been to focus mainly on the electrostatic, VDW and dihedral energies and use both *ab initio* calculations, empirical liquid and solvation data, and experiment to calibrate the model. However, our approach differs significantly from that of many in building from the ground up with the simplest model and defining relatively few general principles, which are elucidated in the section "General Description of the Model" above.

We will attempt to summarize the salient features of some of the more commonly used force fields here, in order to compare and contrast our approach with theirs. They can be roughly grouped into four different categories, depending on the nature and complexity of the force field equation: (1) those with rigid or partially rigid geometries, (2) those without electrostatics, (3) simple diagonal force fields, and (4) more complex force fields.

The ECEPP force field of Scheraga [88] employs rigid internal geometries which allow a more efficient exploration of conformational space. This approach has the disadvantage that it can cause certain conformations and conformational barriers to be too high in energy. A second force field which uses only partially rigid geometries is JUMNA [89], developed by Lavery and co-workers. This force field has been developed for nucleic acids and allows flexibility in the sugar ring but uses mainly internal geometries and keeps the bases rigid.

The SYBYL force field [90] has been developed for the calculation of internal geometries and conformational energies. Because it contains no electrostatic term, it is inappropriate for studying detailed condensed phase properties. The YETI force field [91], developed by Vedani and Huhta, is a modification of the Weiner *et al.* force field with highly damped electrostatics and an angular dependent hydrogen bond (and metal ligation) potential added.

This approach could be valuable in some modelling situations, where large and difficult to handle electrostatic energies are present, but it is also unlikely to be general and extendable to condensed phase phenomena.

The category of simple diagonal force fields includes the Weiner *et al.* (AMBER) [5,6], GROMOS [92], CHARMm [93], and OPLS/AMBER [15] force fields. All of these force fields employ a simple harmonic diagonal representation for the bond and angle terms. Descriptions of the non-bonded and dihedral eneries are given in Table XIII. The Weiner *et al.* force field derived charges from fits to the electrostatic potential of a molecule whereas the other two force fields used empirical fits to interaction energies (CHARMm) or liquid and solid state data (GROMOS). The Weiner *et al.*, CHARMm, and GROMOS force fields all employ VDW parameters derived from crystal data, whereas the VDW parameters in the OPLS/AMBER and Cornell *et al.* force fields are derived from liquid simulations. The OPLS/AMBER and GROMOS force fields specify values for "A" and "B", the repulsive and attractive coefficients, respectively, whereas Weiner *et al.*, Cornell *et al.*, and CHARMm specify values for R* and ε. (Some force fields use "C" instead of "B.") For heteronuclear interactions, the OPLS/AMBER and GROMOS force fields determine values for A and B using geometric mean combining rules. By comparison, Weiner *et al.*, Cornell *et al.*, and CHARMm employ arithmetic combining rules for R* and geometric combining rules for ε. (See Table XII for the relationship between A, B, ε*, and R*.) GROMOS makes the further distinction of using different values for A and B for a particular atom type, depending on the second atom involved in the interaction. This has been shown to result sometimes in anomalous behavior [96,97].

Two new sets of CHARMm hydrocarbon VDW parameters have recently been published [94,95] and tested by Kaminski *et al.* [96] for their ability to reproduce condensed phase properties. The CHARMm92 [94] parameters resulted in a density for liquid butane which

was 63% in error. The CHARMm94 [95] parameters performed much better, reproducing the density and heat of vaporization of the neat liquid alkanes (excluding methane) with average errors of 3.2% and 4.5%, comparable to the results obtained with the new AMBER parameters reported here, where the average errors were 3.8% and 4.4%. Nonetheless, the CHARMm94 model is more complex, using a different R^* and ε for CH_2 and CH_3 carbons. Kaminski *et al.* also reported new all-atom VDW parameters for the OPLS force field, and these were shown to result in average errors of 3.5% and 1.2% for the density and heat of vaporization of the neat hydrocarbon liquids tested (again excluding methane). The OPLS all-atom parameters also performed better at reproducing the relative free energies of solvation of methane, ethane, propane, and butane. It should be noted that while the OPLS parameters result in the lowest overall error for the systems described/included above, this is achieved at the expense of fitting the neat liquid properties of methane.

While all five force fields employ a simple Fourier expansion to represent the dihedral energy, some variation is also seen in the assignment of that energy, with Weiner *et al.*, Cornell *et al.*, OPLS/AMBER, and later versions of CHARMm distributing the energy equally among equivalent bond paths (such as the nine HC-CT-CT-HC dihedrals in ethane), and GROMOS allowing user specification of that parameter. In earlier versions of CHARMm the dihedral energy was assigned to only one specific bond path (quartet of atoms).

Finally, the category of "more complex" force fields includes not only the MM2 and MM3 force fields for small molecules but also two macromolecular force fields. These force fields go beyond the simple diagonal potential function in their inclusion of higher ordered terms as well as cross-terms for representing bonds and angles. The MM3 force field is the state-of-the-art for modelling organic molecules in the gas phase and has been carefully

calibrated to reproduce many properties of these molecules. The focus of MM3 is quite different from that of the force field presented here in that it is not oriented towards the representation of polar and ionic molecules in condensed phases, although, for example, some crystal minimizations were used to calibrate some of the non-bonded parameters. Its complex functional form is necessary for reproducing vibrational frequencies and subtleties of molecular geometries. The use of a 6-exponential non-bonded potential is more accurate than the 6-12 used here, particularly for close contacts such as those found in highly strained organic molecules. The MM2/MM3 model uses a point dipole approach for electrostatic interactions which has often worked well for modelling intramolecular properties but has not been rigorously established as a general model for modelling intermolecular interactions. MM2/MM3 has a large number of dihedral parameters specific to four-atom bond quartets which have been fit to a large set of data.

A second complex force field is the "Class II" one under development by Hagler and co-workers [99]. This force field has a functional form of similar complexity to that of MM2/MM3, but differs in the extensive use of quantum mechanical energies and gradients for its calibration. The developers of this force field are pioneering new ways of deriving parameters and analyzing molecular interactions. This force field currently suffers, however, from the lack of a general charge model of the same caliber as the other parameters.

The third complex force field is the Merck Molecular Force Field (MMFF) under development by Halgren [100]. The stated purpose of this force field is to be able to handle all of the functional groups of interest in pharmaceutical design. The non-bonded function is a "buffered" 7-14 potential, which Halgren found to give the best fit to rare gas interactions and an empirical bond dipole model is used to assign partial charges. The key calibration test set is a series of conformational energies calculated at a very high level of ab

initi

have

large

char

Cor

We

prot

stren

emp

electi

the u

RES

algo

and (

work

appe

the d

comp

quali

Furth

studie

potent

justific

3IG* R

body lev

initio theory (MP4SDQ/TZP//MP2/6-31G*). Thus far, no condensed phase simulations have been carried out, but are planned. This approach has the advantage of generality to a large number of molecules, but at the expense of the use of a simple, empirical, generic charge model and a large number of dihedral parameters.

Conclusion

We have described the development of a second generation force field for the simulation of proteins, DNA, and related organic molecules primarily in the condensed phase. The strengths of the approach presented here are: (1) the general and algorithmic strategy employed to develop the force field; (2) the emphasis on the accurate reproduction of electrostatic interactions — a demonstrated strength of the Weiner *et al*. force field [5,6]; (3) the use of a new approach for deriving electrostatic potential fit charges (multi-conformer RESP) which are better behaved than the previous standard ESP model; (4) general and algorithmic approaches to describe the non-bonded interactions, particularly for hydrogens; and (5) a minimalist approach to adding dihedral potentials to the energy function. This work represents as complete of a description of the development of a force field as has appeared in the literature. Through our approach we have minimized the coupling between the different terms in the force field equation. Although only the total energy can be compared directly with experiment, the force field has the potential of providing additional qualitative insight when the results agree with experiment.

Further applications will be required to assess how successful the new model is. In the studies described above, the major weakness was the necessity of adding dihedral potentials for the ψ and ϕ of peptides and χ of nucleic acids without obvious physical justification. This effect is at least partially due to the somewhat too large polarity of the 6-31G* RESP model, which is needed to accurately simulate solvation at the effective two body level. The magnitude of the re-optimized ψ and ϕ dihedral parameters is considerably

reduced in a non-additive force field with reduced gas phase-like polarity [101], and the magnitudes are slightly reduced for χ [102]. A better behaved set of charges which yielded more accurate conformational energies and still reproduced solvation free energies could possibly be derived through empirical adjustment. But then the generality and simplicity of the model would be sacrificed. These examples do emphasize the degree to which the non-bonded and dihedral terms dominate any complex intramolecular function, particularly when the charges are optimized for an effective two-body model to reproduce the energies of polar and ionic molecules in solution.

This new force field has retained some of the features of the Weiner *et al.* force field [5,6], with its emphasis on the accurate representation of electrostatics and simple representation of bond and angle energies, while offering electrostatic and VDW parameters which are optimized for state-of-the-art condensed phase simulations. Further work is being carried out in this laboratory to investigate the improved performance of models which incorporate either off-center charges (lone pairs) [103] or electronic polarization [40,81,101,102]. It is our belief, however, that with this new force field we have reached the limit for accurately representing biomolecular systems with an effective two-body additive potential employing quantum mechanically derived atom centered charges.

Acknowledgements

References

1. Burkert, U. and Allinger, N.J. *Molecular Mechanics;* American Chemical Society: Washington, 1992.

2. Allinger, N.L. *J. Amer. Chem. Soc.* 1977, **99**, 8127, and subsequent versions, e.g. MM2-87, MM2-89, MM2-91.

3. Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. *J. Amer. Chem. Soc.* **1989**, *111*, 8551-8566; ibid, 8566-8576; ibid, 8576-8582.

4. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids;* Cambridge University Press: Cambridge, 1987.

5. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Amer. Chem. Soc.* **1984**, *106*, 765-784.

6. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Amer. Chem. Soc.* **1986**, *7*, 230-252.

7. Pavitt, N.; Hall, D. *J. Comput. Chem.* **1984**, *5*, 441.

8. Nilsson,L.; Karplus, M. *J. Comp. Chem.* **1986**, *7*, 591-616.

9. Tilton, R.F.; Singh, U.C.; Weiner, S.J.; Connolly, M.L.; Kuntz, I.D., Jr., Kollman, P.A.; Max, N.; Case, D. *J. Molec. Biol.* **1986**, *192*, 443-456.

10. Guenot, J.M.; Kollman, P.A. *Protein Science* **1992**, *1*, 1185-1205.

11. York, D.M.; Wlodawer, A.; Redersen, L.; Darden, T.A. *Proc. Nat. Acad. Sci. USA* **1994**, *91*, 8715-8718.

12. Singh, U.C.; Weiner, S.J.; Kollman, P.A. *Proc. Nat.. Acad. Sci. USA* **1985**, *82*, 755-759.

13. Seibel, G.L.; Singh, U.C.; Kollman, P.A. *Proc. Nat. Acad. Sci. USA* **1985**, *83*, 2923-2933.

14. Jorgenson, W.L.; Pranata, J. *J. Amer. Chem. Soc.* **1990**, *112*, 2008-2010.

15. Jorgenson, W.L.; Tirado-Rives, J. *J. Amer. Chem. Soc.* **1988**, *110*, 1657-1666.

16. (a) Tirado-Rives, J.; Jorgensen, W.L. *J. Amer. Chem. Soc.* **1990**, *112*, 2773-2781. (b) Orozco, M.; Tirado-Rives, J.; Jorgensen, W.L. *Biochemistry* **1993**, *32*, 12864-12874.

17. Berendsen, H.J.C.; Grigera, J.R.; Straatsma, T.P. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

18. Jorgensen, W.L.; Chandreskhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. *J. Chem. Phys.* **1982**, *79*, 926-935.

19. Kuyper, L.; Ashton, D.; Merz, K.M., Jr.; Kollman, P.A. *J. Phys. Chem.* **1991**, *95*, 6661-6666.

20. Cornell, W.; Cieplak, P.; Bayly, C.; Kollman, P.A. *J. Amer. Chem. Soc.* **1993**, *115*, 9620.

21. Sun, Y.X.; Spellmeyer, D.; Pearlman, D.A.; Kollman, P.A. *J. Amer. Chem. Soc.* **1992**, *114*, 6798-6801.

22. Williams, D.E. *Biopolymers*, **1990**, *29*, 1367-1386.

23. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P.A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

24. Cieplak, P.; Kollman, P.A. *J. Comput. Chem.* **1991**, *12*, 1232-1236.

25. Hagler, A.; Euler, E.; Lifson, S. *J. Amer. Chem. Soc.* **1974**, *96*, 5319.

26. Gough, C.; DeBolt, S.; Kollman, P.A. *J. Comput. Chem.* **1992**, *13*, 963-970.

27. Veenstra, D.; Ferguson, D.; Kollman, P.A. *J. Comput. Chem.* **1992**, *13*, 971-978.

28. (a) Gould, I.R.; Kollman, P.A. *J. Phys. Chem.* **1992**, *96*, 9255-9258. (b) Gould, I.R.; Cornell, W.D.; Hillier, I.H. *J. Amer. Chem. Soc.* **1994**, *116*, 9250-9256.

29. Creighton, T.E. *Proteins, 2nd. ed.*, W.H. Freeman: New York, 1984.

30. Lipkowitz, K.B.; Peterson, M.A. *J. Comput. Chem.* **1993**, *14*, 121-125.

31. Halgren, T.A. *J. Amer. Chem. Soc.* **1990**, *112*, 4710.

32. Harmony, M.; Laurie, V.; Kuezkowski.; Schwendeman, R.; Ramsay, D.; Lovas, F.; Lafferty, W.; Maki, A. *J. Phys. Chem. Ref. Data* **1979**, *8*, 619.

33. Benedetti, E. in *Peptides-Proceedings of the 5th American Peptide Symposium*, Goddman, M.; Meienhofer, J., Eds.; New York, 1977, p 257.

34. Douglas, J.; Rabinovich; B.S.; Looney, F. *J. Chem. Phys.* **1955**, *23*, 315.

35. Momany, F.; McGuire, R.; Burgess, A.; Scheraga, H. *J. Phys. Chem.* **1975**, *79*, 2361.

36. Lehn, J. *Theor. Chim. Acta* **1970**, *16*, 351.

37. Spellmeyer, D. unpublished.

38. Sun, Y.; Kollman, P.A. "Hydrophobic Solvation of Methane and Nonbond Parameters of the TIP3P Water Model," *J. Comput. Chem.*, submitted.

39. Cieplak, P.; Cornell, W.D.; Bayly, C.; Kollman, P.A. "Application of the Mutlimolecule and Multiconformation RESP Methodology to Biopolymers: Derivation for DNA, RNA and Proteins," *J. Comp.Chem.*, submitted.

40. Caldwell, J.; Kollman, P., The Structures and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methyl Acetamide," *J. Phys. Chem.*, submitted.

41. Gregoret, L.M.; Rader, S.D.; Fletterick, R.J.; Cohen, F.E. *Proteins-Struct. Funct. Genet.* **1991**, *9*, 99-107.

42. Howard, A.; Cieplak, P.; Kollman, P.A. "A Molecular Mechanical Model that Reproduces the Relative Energies for Chair and Twist-Boat Conformations of 1,3-Dioxanes", *J. Comput. Chem.*, in press.

43. (a) Frisch, M.J.; Head-Gordon, M.; Trucks, G.W.; Foresman, J.B.; Schlegel, H.B.; Raghavachari, K.; Robb, M.A.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seger, R.; Melius, C.F.; Baker, J.; Martin, L.R.; Kahn, L.R.; Stewart, J.J.P.; Topiol, S.; Pople, J.A. *Gaussian 90*, Gaussian, Inc., Pittsburgh, PA 1990. (b) Frisch, M.J.; Trucks, G.W.; Head-Gordon, M.; Gill, P.M.W.; Wong, M.W.; Foresman, J.B.; Johnson, B.G.; Schlegel, H.B.; Robb, M.A.; Replogle, E.S.; Gomperts, R.; Andres, J.L.; Raghavachari, K.; Binkley, J.S.; Gonzalez, C.; Martin, R.L.; Fox, D.J.;

Defrees, D.J.; Baker, J.; Stewart, J.J.P.; Pople, J.A. *Gaussian 92*, Revision A, Gaussian, Inc., Pittsburgh, PA 1992.

44. Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Seibel, G.L.; Singh, U.C.; Weiner, P.A.; Kollman, P.A. *AMBER 4.0 (UCSF)*; Department of Pharmaceutical Chemistry, University of California: San Francisco, CA, 1991.

45. van Gunsteren, W.F.; Berendsen, H.J.C. *J. Comput. Aided Mol. Des.* **1987**, *1*, 171.

46. (a) van Gunsteren, W.F.; Berendsen, H.J.C. *Mol. Phys.* **1977**, *34*, 1311. (b) Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. *J. Comput. Phys.* **1977**, *23*, 327.

47. Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.* **1991**, *94*, 4532.

48. Spellmeyer, D.C.; Swope, W.C., Evensen, E.-R.; Ferguson, D.M. *SPASMS*, University of California: San Francisco, CA, 1994.

49. Ferguson, D.M. *J. Chem. Phys.* **1993**, *99*, 10086.

50. Hirota, E.; Emlo, Y.; Saito, S.; Duncan, J. *J. Mol. Spect.* **1981**, *89*, 285.

51. Heenan, R.K.; Bartell, L.S. *J. Chem. Phys.* **1983**, *78*, 1270.

52. Compton, D.A.C.; Montero, S.; Murphy, W.F. *J. Phys. Chem.* **1980**, *84*, 3587.

53. Allinger, N.L.; Greu, R.S.; Schaefer III, H.F. *J. Amer. Chem. Soc.* **1990**, *112*, 113.

54. See review by P. Payne and L.C. Allen in *Modern Theoretical Chemistry, Applications of Electronic Structure Theory*, H.F. Schaefer, ed., Plenum: New York, 1987, Chapter 2.

55. Miller, M. personal communication.

56. Allinger, N.L.; Rahman, M.; Lii, J-H, *J. Amer. Chem. Soc.* **1990**, *112*, 8293.

57. Blukis, U.; Kasei, P.H.; Myers, R.J., *J. Chem. Phys.* **1963**, *38*, 2753.

58. Almenningen, A.; Seip, H.M.; Willadsen, T. *Acta Chem.Scand.* **1969**, *23*, 2748.

59. Engerholm, G.G.; Luntz, A.C.; Gwinn, W.D.; Harris, D.O. *J. Chem. Phys.* **1969**, *50*, 2446.

60. Cremer, D.; Pople, J. *J. Amer. Chem. Soc.* **1975**, *97*, 354.

61. Geise, H.; Adams, W.; Bartell, L. *Tetrahedron* **1969**, *25*, 3045.

62. Kitagawa, T.; Miyazawa, T. *Bull Chem. Soc. Japan* **1968**, *41*, 1986; MP2/6-31G*//HF/6-31G* *ab intio* calculations lead to a ΔE=1.4 kcal/mole.

63. Oyanagi, K.; Kutchitsu, K. *Bull. Chem. Soc. Japan* **1978**, *51*, 2237.

64. Hayashi, M.; Kuwada, K. *J. Mol. Struct.* **1982**, *78*, 53.

65. Lees, R.M.; Baker, J.G. *J. Chem. Phys.* **1968**, *48*, 5299.

66. Brown, E.; Peticolas, W. *Biopolymers* **1975**, *14*, 1259.

67. Allinger, N.L.; Quinn, M.; Rahman, M.; Chen, K. *J. Phys. Org. Chem.* **1991**, *4*, 647-658.

68. Allinger, N.L.; Quinn, M.; Rahman, M.; Chen, K. *J. Phys. Org. Chem.* **1991**, *4*, 659-666.

69. Gould, I.R.; Kollman, P.A. *J. Amer. Chem. Soc.* **1994**, *116*, 2493-2499.

70. Rey-Lafon, M.; Ford, M.T.; Garrigen-Lagrange, C. *Spect. Acta, Part A* **1973**, *29A*, 471.

71. Shimananchi, T. "Tables of Molecular Vibrational Frequencies," National Stand. Ref. Data, Washington, D.C., 1967; parts 1-3.

72. Dhasuadi, Z.; Ghomi, M.; Austin, Y.C.; Girling, R.B.; Hester, R.E.; Mojres, P.; Chinosky, L.; Tarpin, P.Y.; Coulombeau, C.; Yobic, H.; Tomhinson, Y. *J. Phys. Chem.* **1993**, *97*, 1074.

73. Susi, H.; Ard, Y.S.; Purcell, Y.M. *Spectrochim. Acta.* **1973**, *29A*, 725.

74. (a) Delaber, J-M.; Majoube, M. *Spectrochim. Acta* **1978**, *37A*, 129; (b) Delabar, J-M. *J. Raman Spect.* **1978**, *7*, 261.

75. Beetz, C.P., Jr.; Ascarelli, G. *Spect. Acta* **1980**, *36A*, 299.

76. Cremer, D.; Pople, J. *J. Amer. Chem. Soc.* **1975**, *97*, 1354.

77. Saenger, W. Principles of Nucleic Acid Structure; Springer-Verlag: Tokyo, 1984.

78. Sun, Y.; Kollman, P.A. *J. Chem. Phys.* **1992**, *97*, 2487.

79. Ben Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.

80. Wolfenden, R. *Biochemistry* **1978**, *17*, 201.

81. Meng, E.; Caldwell, J.; Cieplak, P.; Kollman, P. "Accurate Solvation Free Energies of Acetate and Methylammonium Calculated with a Polarized Water Model," *J. Amer. Chem. Soc.*, accepted.

82. Analyzed by Cramer, C.J.; Truhlar, D.G., *J. Amer. Chem. Soc.* **1991**, *113*, 8305-8311.

83. Hine, J.; Mookerjee, P.K. *J. Org. Chem.* **1975**, *40*, 292-298.

84. Ferguson, D.M.; Pearlman, D.A.; Swope, W.C.; Kollman, P.A. *J. Comput. Chem.* **1992**, *13*, 362-370.

85. Vijay-Kumar, S.; Bugg, C.E.; Cook, W.J. *J. Mol. Bio.* **1987**, *194*, 531.

86. Alonso, D.O.V.; Daggett, V. "Molecular Dynamics Studies of Partially Unfolded Conformations of Ubiquitin in Methanol and Their Refolding in Water," submitted.

87. Saito, M. *J. Chem. Phys.* **1994**, *101*, 4055-4061.

88. (a) Roterman, I.K.; Lambert, M.H.; Gibson, K.D.; Scheraga, H.A., *ibid*, **1989**,7, 421-453. (b) see also Kollman, P.A.; Dill, K.A. , *ibid*, **1991**, *8*, 1103; Gibson, K.D.; Scheraga, *J. Biomol. Struct. Dyn.* **1991**, *8*, 1109-1111;

89. Lavery, R.; Hartmann, B. *Biophys. Chem.* **1994**, *50*, 33-45.

90. Clark, M.; Cramer, R.D.; van Oppendosch, N. *J. Comput. Chem.* **1989**, *10*, 982-1012.

91. Vedani, A.; Huhta, D.A. *J. Amer. Chem. Soc.* **1990**, *112*, 4759-4767.

92. van Gunsteren, W.F.; Berendsen, H.J.C. Groningen Molecular Simulations (GROMOS) Library Manual, Biomos, Groningen 1987.

93. Brooks, B.R.; Brucoleri, R.E.; Olafson, B.D.; Slater, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

94. Smith, J.C.; Karplus, M. *J. Amer. Chem. Soc.* **1992**, *114*, 801-812.

95. Woolf, T.B.; Roux, B. *J. Amer. Chem. Soc.* **1994**, *116*, 5916-5926.

96. Kaminski, G.; Duffy, E.M.; Matsui, T.; Jorgensen, W.L. "Free Energies of Hydration of Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model," *J. Phys. Chem.*, submitted.

97. Mark, A.E.; van Helden, S.; Smith, P.E.; Janssen, L.H.M.; van Gunsteren, W. *J. Amer. Chem. Soc.* **1994**, *116*, 6293-6302.

98. Aqvist, J.; Medina, C.; Samuellson, J.E., *Prot. Engin.* **1994**, 385-391.

99. (a) Maple, J.R.; Dinur, U.; Hagler, A.T. *Proc. Nat. Acad. Sci. USA* **1988**, *85*, 5350-5354. (b) Dinur, U.; Hagler, A.T. *J. Chem. Phys.* **1989**, *91*, 2949-2958. (c) Maple, J.R.; Hwang, M.J.; Stockfisch, T.P.; Dinur, U.; Waldman, M.; Ewig, C.S.; Hagler, A.T. *J. Comput. Chem.* **1994**, *15*, 162-182.

100. Halgren, T. *J. Amer. Chem. Soc.* **1992**, *114*, 7827-7843. Halgren, T. Results to be submitted for publication.

101. Cornell, W.D.; Caldwell, J.W.; Kollman, P.A., manuscript in preparation.

102. Cieplak, P.; Kollman, P.A., unpublished results.

103. Dixon, R.; Kollman, P.A., work in progress.

Figure Captions

Figure 1. Model of deoxyadenosine employed in the quantum mechanical and molecular mechanical conformational studies reported in Table X. In the quantum mechanical calculations, the HO4'-O4'-C1'-N9 and H2'3-C2'-C1'-N9 dihedrals were held fixed at values characteristic of a C2'-endo sugar, in order to mimic the conformation of the sugar ring. In the molecular mechanical calculations, the dihedrals were restrained to those values with dihedral restraints of 500 kcal/mole.

Figure 2a. The molecular mechanical (ϕ,ψ) map for methyl-blocked glycyl dipeptide generated using the new force field presented here. Contours are drawn every 2 kcal.

Figure 2b. The molecular mechanical (ϕ,ψ) map for methyl-blocked alanyl dipeptide generated using the new force field presented here. Contours are drawn every 2 kcal.

Figure 3. RMS deviation (A) between the crystal structure of ubiquitin and structures along an MD trajectory as modelled by the Weiner *et al.* and Cornell *et al.* force fields. The lower lines correspond to the RMS deviation of the heavy backbone atoms only and the upper lines to the RMS deviation for all heavy atoms in each residue.
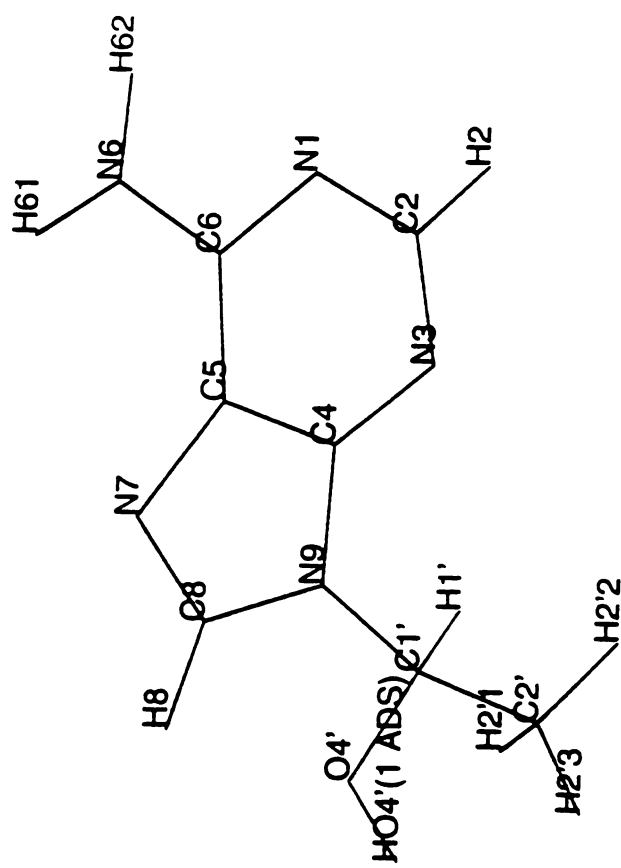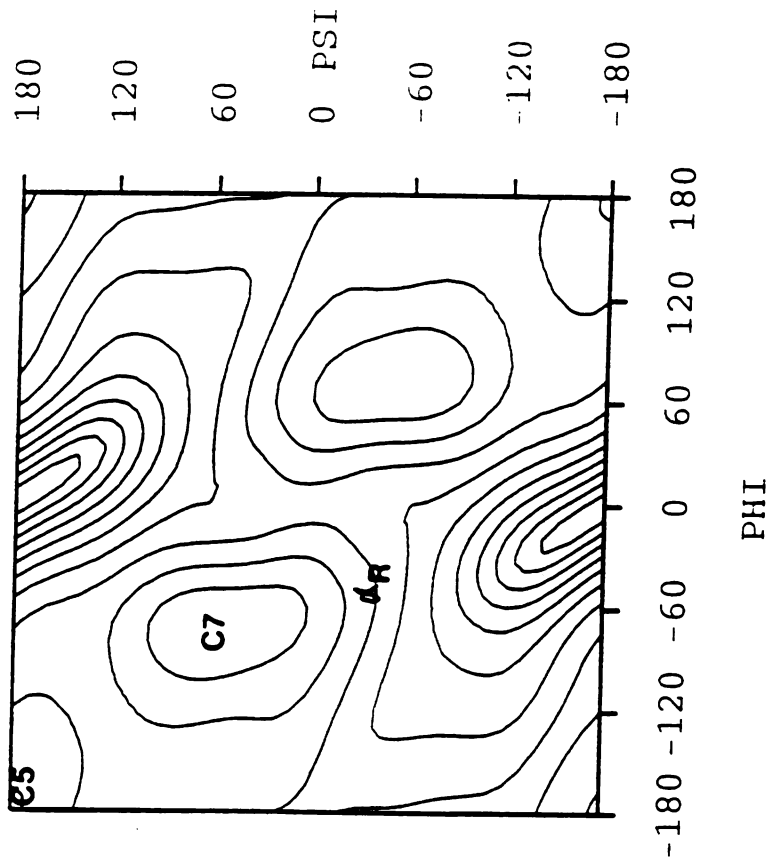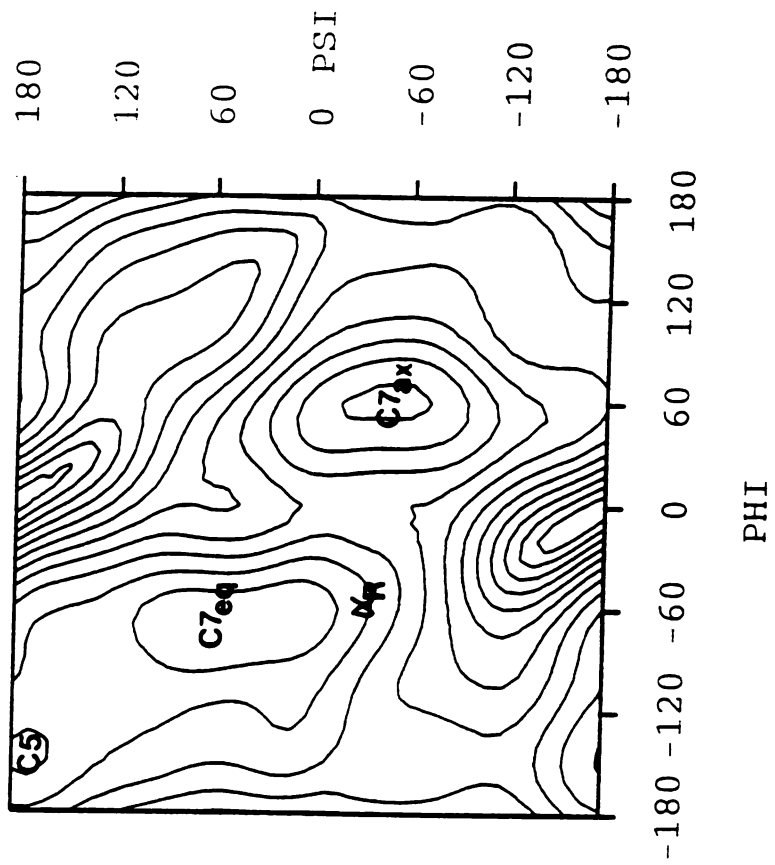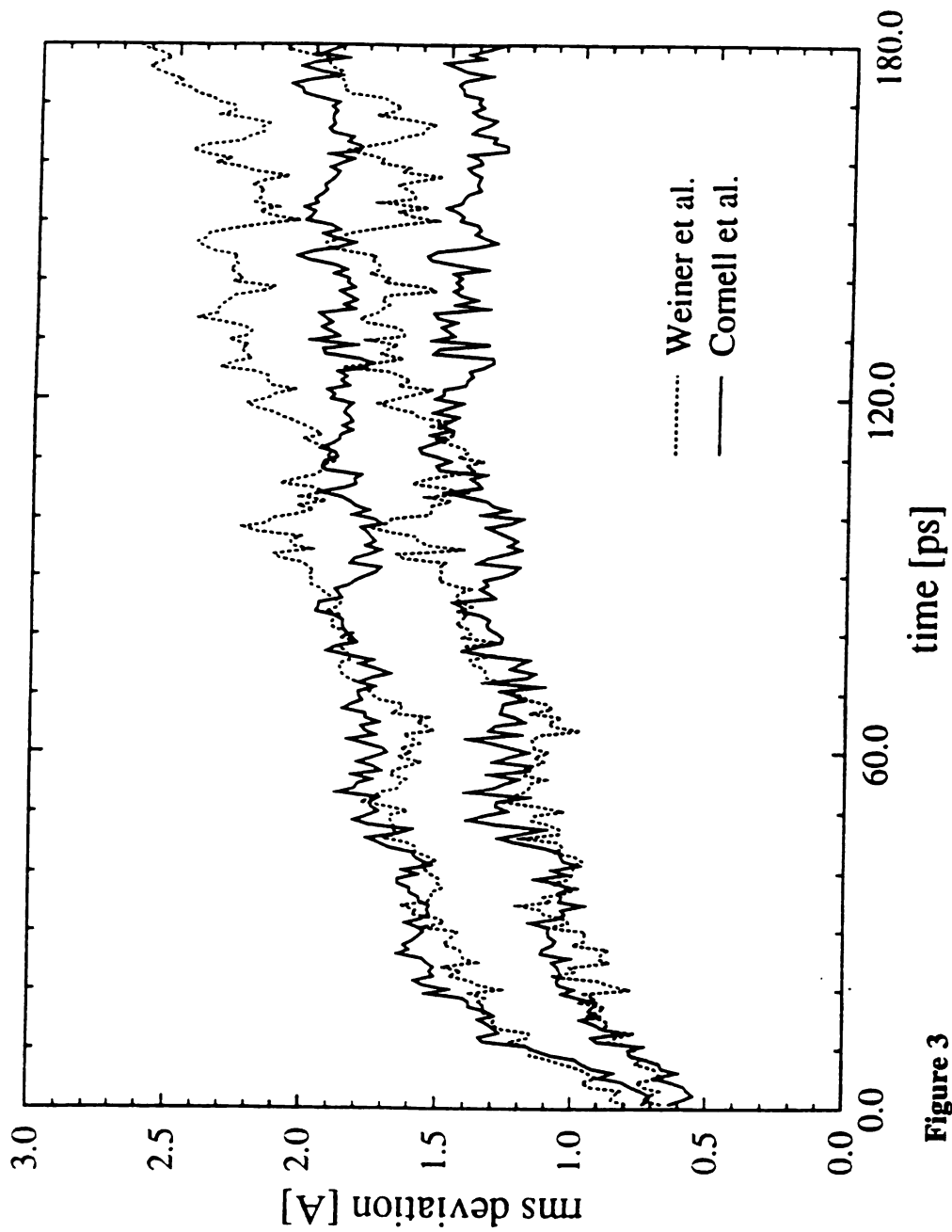
Figure 1

Figure 2a

Figure 2b

Figure 3

Table I. List of Atom Types [a]

Atom	Type	Description
carbon	CT	any sp^3 carbon
carbon	C	any carbonyl sp^2 carbon
carbon	CA	any aromatic sp^2 carbon and (Cε of Arg)
carbon	CM	any sp^2 carbon, double bonded
carbon	CC	sp^2 aromatic in five membered ring with one substituent + next to Nitrogen (Cγ in His)
carbon	CV	sp^2 aromatic in five membered ring next to carbon and lone pair nitrogen (e.g. Cδ in His (δ).
carbon	CW	sp^2 aromatic in five membered ring next to carbon and NH (e.g. Cδ in His (ε) and in Trp)
carbon	CR	sp^2 aromatic in five membered ring next to two nitrogens (Cγ andCε in His)
carbon	CB	sp2 aromatic at junction of 5 and 6 membered rings (Cδ in Trp) and both junction atoms in Ade and Gua.
carbon	C*	sp^2 aromatic in five membered ring next to two carbons (e.g. Cγ in Trp)
carbon	CN	sp^2 junction between five and six membered rings and bonded to CH and NH (Cε in Trp)
carbon	CK	sp^2 carbon in five membered aromatic between N and N-R (C8 in purines)
carbon	CQ	sp^2 carbon in six membered ring between lone pair nitrogens (e.g. C2 in purines)

Table I. List of Atom Types (cont'd.) [a]

nitrogen	N	sp^2 nitrogen in amides
nitrogen	NA	sp^2 nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)
nitrogen	NB	sp^2 nitrogen in five membered ring with lone pair (e.g. N7 in purines)
nitrogen	NC	sp^2 nitrogen in 6 membered ring with lone pair (e.g. N3 in purines)
nitrogen	N*	sp^2 nitrogen in 5 membered ring with carbon substituent (in purine nucleosides)
nitrogen	N2	sp2 nitrogen of aromatic amines and guan ions
nitrogen	N3	sp^3 nitrogen
oxygen	OW	sp^3 oxygen in TIP3P water
oxygen	OH	sp^3 oxygen in alcohols, tyrosine and protonated carboxyl acids
oxygen	OS	sp^3 oxygen in ethers
oxygen	O	sp^2 oxygen in amides
oxygen	O2	sp^2 oxygen in anionic acids
sulfur	S	sulfur in methionine and cysteine
sulfur	SH	sulfur in cysteine
phosphorus	P	phosphorus in phosphates

Table I. List of Atom Types (cont'd.) [a]

hydrogen	H	H attached to N
hydrogen	HW	H in TIP3P water
hydrogen	H O	H in alcohols and acids
hydrogen	HS	H attached to sulfur
hydrogen	HA	H attached to aromatic carbon
hydrogen	H C	H attached to aliphatic carbon with no electron-withdrawing substituents
hydrogen	H1	H attached to aliphatic carbon with one electron-withdrawing substituent
hydrogen	H2	H attached to aliphatic carbon with two electron-withdrawing substituents
hydrogen	H3	H attached to aliphatic carbon with three electron-withdrawing substituents
hydrogen	HP	H attached to carbon directly bonded to formally positive atoms (e.g. C next to NH_3^+ of lysine)
hydrogen	H4	H attached to aromatic carbon with one electronegative neighbor (e.g. hydrogen on C5 of Trp, C6 of Thy)
hydrogen	H5	H attached to aromatic carbon with two electronegative neighbors (e.g. H8 of Ade and Gua and H2 of Ade)

[a] See refs. 5 and 6.

Table II. Standardized Parameters for Scaling Algorithms

bond	r_{eq} [a]	K_r [b]
pure C-C	1.507 [c]	317 [d]
pure C=C	1.336 [e]	570 [f]
pure C-N	1.449 [g]	337 [h]
pure C=N	1.273 [i]	570 [j]

torsion	r_{eq} [a]	V_2 [k]
pure X-C-C-X	1.507 [c]	0.0 [l]
partial X-C=C-X	1.397 [m]	14.5 [n]
pure X-C=C-X	1.336 [e]	30.0 [o]
pure X-C-N-X	1.449 [g]	0.0 [p]
partial X-C=N-X	1.335 [q]	10.0 [r]
pure X-C=N-X	1.273 [i]	30.0 [s]

[a] in Å. [b] in kcal/mol Å². [c] Microwave data from acetone (ref. 32). [d] Value taken from MM2, ref 2. [e] Microwave data from propene (ref. 32). [f] Default from NMA normal mode analysis for carbonyl force constant. [g] Benedetti structural data (ref. 33). [h] Value derived from normal mode analysis on NMA. [i] Microwave data from methylenimine (ref. 32). [j] Default value, see footnote f. [k] In kcal/mol. [l] Assumed free rotation about pure C-C single bond. [m] Structural data from benzene (ref. 32). [n] From normal modes analysis of benzene. [o] Approximate rotational barrier of ethylene is ~60 kcal/mol (see ref 34). [p] Assumed free rotation about a pure single C-N bond. [q] Benedetti structural data (ref. 33). [r] Ref. 35. [s] Calculated rotational barrier in methylene imine is 57.5 kcal/mol (see ref 36).

Table III. Results for hydrocarbons (energies in kcal/mole, angles in degrees)

parameter	this work	MM3	experiment or high level theory
Ethane			
ΔE (eclipsed staggered)	2.89	2.41	2.88[a,b]
Butane			
ΔE (gauche-trans)	0.67	0.81	0.75±0.25[c]
ΔE (cis-trans)	5.16	4.83	4.56[d], 4.89[e]
structural parameters			
ϕ (gauche)	68.0	64.5	71±5[c]
δ(C-C-C) (cis)	117.2	--	--
δ(C-C-C) (trans)	111.3	112.4	113±4[c]
δ(C-C-C) (gauche)	113.5	113.7	--
Propane			
ΔE (V1)[d]	3.30	--	3.3[f]
ΔE (V2)[d]	3.74	--	3.9[f]

[a] Ref. 3. [b] Ref. 50. [c] Ref. 51. [d] Energy for methylene group to eclipse first methyl group, relative to all staggered conformation (V1) and energy for methylene to eclipse second methyl group, relative to first eclipsed conformation (V2) (ref. 52). [e] Ref. 53. [f] Ref. 54.

Table IV. Results for alcohols and ethers (energies in kcal/mole, angles in degrees)

parameter	this work	MM3[a]	experiment
		Dimethyl Ether	
ΔE (eclipsed-staggered)	2.74	2.45	2.72 [b]
δ(C-O-C)(staggered)	112.3	111.9	111.8 [b]
δ(C-O-C)(eclipsed)	113.3	-	-
		Tetrahydrofuran	
$\Delta E(C_s$-$C_2)$	0.12	0.094	0±0.3 [c]
$\Delta E(C_{2v}$-$C_2)$	3.98	4.41	3.5 [d]
structural parameters			
C_2 conformation			
q [e]	0.40	-	0.39 [c]
θ(C-O-C)	108.8	108.7	110.5 [c]
θ(C-O-O)	106.8	106.7	106.5 [c]
θ(C-C-C)	100.4	101.1	101.8 [c]
C_s conformation			
q [e]	0.38	-	0.364 [c], 0.38 [f]
θ(C-O-C)	105.4	104.0	106.2 [c]
θ(C-O-O)	105.1	105.0	105.0 [c]
θ(C-C-C)	103.6	103.6	104.1 [c]
		Methyl Ethyl Ether	
ΔE(gauche-trans)	1.46	1.49	1.5±0.2 [g]
ΔE(cis-trans)	6.46	6.02	7.01 [h]

Table IV. Results for alcohols and ethers, continued (energies in kcal/mole, angles in degrees)

parameter	this work	MM3[a]	experiment
structural parameters			
φ (gauche)	76.0	74.5	84±6 [i]
δ(C-O-C) (trans)	112.3	112.1	111.7 [j]
δ(C-C-O) (trans)	108.3	108.7	108.9 [k]
		Methanol	
ΔE (eclipsed-staggered)	1.03	0.78	1.06 [k]

[a] Ref. 56. [b] Ref. 57. [c] Ref. 58. [d] Ref. 59. [e] Ref. 60. [f] Ref. 61. [g] Ref. 62. [h] *Ab initio* MP2/6-31G*//HF/6-31G* calculations. [i] Ref. 63. [j] Ref. 64. [k] Ref. 65.

Table V. Dimethyl Phosphate Energies, Structures, and Low Frequency Vibrational Modes

relative energies [a] (kcal/mole)		
conformation [b]	E(MM)	E(QM)
g,g	0.00	0.00
g,t	1.42	1.41
t,t	2.83	3.45

geometrical parameters (angles in degrees) [b,c]			
	MM	QM	x-ray [d]
φ1,φ2 (g,g)	67.7, 67.7	75.2, 75.2	73, 73
φ1,φ2 (g,t)	74.2, 179.2	73.7, 189.4	74, 169
Θ(COP) (g,g)	122.1	118.5	121.7
Θ(OPO) (g,g)	103.8	99.3	104.8
Θ(OPO') (g,g)	108.2	107.5	110.6
Θ(O'PO') (g,g)	119.3	124.9	119.7
Θ(COP) (g,t)	120.5	118.0	
Θ(OPO) (g,t)	102.5	96.7	
Θ(O'PO) (g,t)	108.2	108.5	
Θ(O'PO') (g,t)	120.1	122.8	
Θ(COP) (t,t)	120.2	116.5	
Θ(OPO') (t,t)	103.0	94.3	
Θ(O'PO) (t,t)	108.2	109.6	
Θ(O'PO') (t,t)	119.9	120.9	

Table V. Dimethyl Phosphate Energies, Structures, and Low Frequency Vibrational Modes, continued.

vibrational frequencies <500 cm^{-1} (cm^{-1})	
MM	exp [e]
78	---
109	---
196	195
262	210
295	321
302	345
359	357
383	393
421	503

[a] Absolute energies for g,g conformations are -40.77 kcal/mole (MM) and -720.606019 au (QM). The quantum mechanical calculations used the model MP2//6-31G*//HF/6-31G*. [b] Dihedral angles around C-O-P-O. [c] Bond angles, O is ester oxygen and O' is anionic oxygen. [d] See Table IV in ref. 5. [e] Ref. 66.

Table VI. Low Frequency ($<1000\text{cm}^{-1}$) Vibrational Modes for Small Hydrocarbons, Ethers, Alcohols, and Sulfur Compounds

symm.	ϑ (this work)	ϑ (MM3) [a]	ϑ (exp) [b,c]	mode [d]
		Ethane		
A2u	312	279	283	CH_3-CH_3 torsion
E2u	811	908	822	CH_3 asym rocking
E2u	811	908	822	CH_3 asym rocking
A1g	898	962	995	C-C stretch
		Propane		
A2	231	208	217	CH_3-CH_2 torsion
B2	275	255	265	CH_3-CH_2 torsion
A1	356	375	379	C-C-C bend
B2	733	803	748	CH_2 rock + CH_3 def
A1	809	850	868	CH3 rock + sym C-C str/str
B1	866	938	921	CH3 rock + asym C-C str/str
A2	877	961	899	CH_2 twist + CH_3 def
		Butane		
Au	127	122	121	CH_2-CH_2 torsion
Au	236	216	--	CH_3-CH_2 torsion
Bg	271	245	266	CH_3-CH_2 torsion
Bu	272	287	--	asym C-C-C bend+ C-C-C bend
Ag	364	394	427	sym C-C-C bend+ C-C-C bend
		Methanol		
A"	297	263	270	CH_3-O torsion
A'	867	1052	1034	C-O stretch

Table VI. Low Frequency (<1000 cm^{-1}) Vibrational Modes for Small Hydrocarbons, Ethers, Alcohols, and Sulfur Compounds, continued.

symm.	ϑ (this work)	ϑ (MM3)[a]	ϑ (exp)[b,c]	mode [d]
			Dimethyl Ether	
A2	212	188	198	CH$_3$-O sym torsion
B2	279	273	242	CH$_3$-O asym torsion
A1	416	400	424	C-O-C bend
A1	798	924	918	C-O sym stretch
			Methyl ethyl ether	
A"	123	114	--	C$_2$H$_5$-O torsion
A"	225	216	--	CH$_3$-C torsion + CH$_3$-O torsion
A"	271	257	238	CH$_3$-O torsion + CH$_3$-C torsion
A'	283	296	308	C-O-C bend + C-C-O bend
A'	404	420	472	C-C-O bend + C-O-C bend
	755	870	820	CH$_3$ rock + CH$_2$ rock + CH$_2$ twist
	806	897	855	C-O str + CH$_3$ wag + C-C str
			Methane thiol	
	707	695	704	C-S
	801	823	803	C-S-H
			Dimethyl sulfide	
	279	285	282 (285)	C-S-C
	691	683	691 (683)	S-C sym
	720	702	741 (704)	S-C asym

Table VI. Low Frequency (<1000 cm^{-1}) Vibrational Modes for Small Hydrocarbons, Ethers, Alcohols, and Sulfur Compounds, continued.

symm.	ϑ (this work)	ϑ (MM3)[a]	ϑ (exp)[b,c]	mode [d]
		Dimethyl disulfide		
	105	116	102 (106)	C-S-S-C torsion
	236	241	239 (242)	S-S-C bend
	275	279	272	S-S-C bend
	509	514	509 (514)	S-S stretch
	710	701	689 (694)	S-C stretch
	713	703	(694)	S-C stretch

[a] Refs. 2, 56, 67, and 68. [b] See refs. 2, 5, 56, 67, and 68 for experimental frequencies. [c] Experimental frequencies given in parentheses refer to those used as reference for MM3 values. [d] See refs. 2, 5, 56, 67 and 68 for the mode assignments.

Table VII. Normal Modes of trans-NMA and Benzene (cm^{-1})

		trans-NMA	
nmode #	symm.	this work	experiment [a]
1)	A"	44	---
2)	A"	97	---
3)	A"	184	192
4)	A'	286	289
5)	A'	440	439
6)	A"	587	600
7)	A'	591	628
8)	A"	696	725
9)	A'	801	883
10)	A'	963	991
11)	A"	1037	1044
12)	A"	1046	---
13)	A'	1075	1114
14)	A'	1082	1161
15)	A'	1209	1300
16)	A'	1395	1374
17)	A'	1398	1414
18)	A"	1402	1441
19)	A"	1407	1451
20)	A'	1428	1458
21)	A'	1516	1471
22)	A'	1614	1569
23)	A'	1693	1660

Table VII. Normal Modes of trans-NMA and Benzene (cm^{-1}), continued.

nmode #	symm.	this work	experiment [a]
24)	A'	2868	2935
25)	A'	2869	2935
26)	A"	2980	2981
27)	A'	2982	2981
28)	A'	2982	2994
29)	A'	2983	2994
30)	A'	3304	3307

Table VII. Normal Modes of trans-NMA and Benzene (cm⁻¹), continued.

		Benzene		
nmode #	symm.	this work	experiment [b]	mode
1)	e2u	410	410	ring def.
2)	e2g	609	606	ring def.
3)	a2u	661	673	CH bend
4)	b2g	704	703	ring def.
5)	e1g	900	849	CH bend
6)	e2u	979	975	CH bend
7)	a1g	941	992	ring stretch (breathing)
8)	b2g	947	995	CH bend
9)	b1u	1167	1010	ring def.
10)	e1u	1124	1038	CH bend
11)	b2u	1194	1150	CH bend
12)	e2g	1129	1178	CH bend
13)	b2u	1331	1310	ring stretch (kekule)
14)	a2g	1729	1326	CH bend
15)	e1u	1493	1486	ring stretch + deformation
16)	e2g	1706	1596	ring stretch
17)	e2g	3064	3047	CH stretch
18)	a1g	3062	3062	CH stretch
19)	e1u	3064	3063	CH stretch
20)	b1u	3068	3068	CH stretch

[a] Ref. 70. [b] Ref. 71.

Table VIII. Low Frequency Normal Modes of the Bases (cm^{-1})

nmode #	this work	experiment [a]	mode [a]
Adenine			

nmode #	this work	experiment [a]	mode [a]
In-plane vibrations			
1)	467	337	
2)	529	540	
3)	556	620	
4)	667	665	
Out-of-plane vibrations			
1)	193	184	ring torsion
2)	248	194	ring torsion
3)	290	238	ring torsion
4)	292	310	ring torsion
5)	456	331	C6-N6 torsion
6)	553	550	C6-N6 wag
7)	569	624	
8)	623	655	
9)	672	686	

Table VIII. Low Frequency Normal Modes of the Bases (cm^{-1}), cont'd.

	Cytosine		
nmode #	this work	experiment [b]	mode [b]
In-plane vibrations			
1)	320	400	C-NH$_2$ bend
2)	518	533	C=O bend
3)	539	549	ring def.
4)	659	600	ring def.
Out-of-plane vibrations			
1)	202	197	C-NH$_2$ wag
2)	217	232	C=O wag
3)	414	421	ring def.
4)	493	485	NH$_2$ wag
5)	536	548	NH$_2$ rock
6)	602	566	ring def.

Table VIII. Low Frequency Normal Modes of the Bases (cm^{-1}), cont'd.

	Guanine		
nmode #	this work	experiment [c]	mode [c]
In-plane vibrations			
1)	301	343	C-NH$_2$ bend
2)	344	400	C=O bend
3)	505	501	ring def. (py)
4)	534	557	ring def. (py)
5)	554	645	ring def (py)
6)	644	690	ring def (Im)
Out-of-plane vibrations			
1)	140	142	C-NH$_2$ wag
2)	185	170	C=O wag
3)	233	214	ring (butterfly) def.
4)	300	243	ring (propeller) def.
5)	445	416	ring (py) def.
6)	448	490	ring (Im) def.
7)	534	601	ring (Im) def.
8)	593	654	NH$_2$ rock
9)	674	690	ring (py) def.

Table VIII. Low Frequency Normal Modes of the Bases (cm^{-1}), cont'd.

		Thymine	
nmode #	this work	experiment [d]	mode [d]
In-plane vibrations			
1)	348	321	C-CH$_3$ bend
2)	372	392	C=O bend (out-of-phase)
3)	463	475	ring def.
4)	549	560	ring def.
5)	592	617	C=O bend (in-phase)
Out-of-plane vibrations			
1)	43	---	CH$_3$ rot.
2)	132	---	C-CH$_3$, C=O wag
3)	188	206	C=O wag
4)	332	285	C-CH$_3$ wag
5)	454	433	ring def.
6)	599	635	ring def.

[a] Ref. 72. [b] Ref. 73. [c] Ref. 74. [d] Ref. 75.

Table IX. Conformational Energies for Deoxyadenosine (angles in degrees, energies in kcal/mole). [a]

Pucker	q[a]	W[a]	γ[b]	χ[b]	3'OH[c]	5'OH[d]	E[e]	ΔE[f]
				Sugar Pucker Profile				
				$\varepsilon=1^g$				
C2'endo	0.40	146.1	51.3	-158.4	176.5	171.4	-52.40	0
C3'endo	0.37	5.7	56.9	-162.3	-178.3	-179.3	-51.87	0.63
O4'endo	0.38	65.3	54.1	-156.7	-175.5	-175.4	-49.53	2.87
O4'exo	0.29	276.6	42.6	-178.7	175.8	178.5	-46.54	5.86
				$\varepsilon=4^g$				
C2'endo	0.39	144.9	55.3	-153.1	176.2	179.2	-1.65	0
C3'endo	0.38	14.1	56.3	-156.1	178.1	179.9	-0.61	1.04
O4'endo	0.39	56.6	55.6	-153.1	179.1	179.5	0.21	1.86
O4'exo	0.30	285.0	42.7	176.9	177.3	179.8	4.03	5.68

Table IX. Conformational Energies for Deoxyadenosine (kcal/mole) continued.

			Gamma Dependence					
Pucker	q^a	W^a	γ^b	χ^b	$3'OH^c$	$5'OH^d$	E^e	ΔE^f
				$\varepsilon=1^g$				
C2'endo	0.40	146.1	51.3	-158.4	176.5	171.4	-52.40	0
C2'endo	0.42	141.9	-168.6	-168.6	179.6	-179.9	-50.39	2.01
C2'endo	0.42	151.7	-62.3	-169.9	-179.7	-179.7	-50.92	1.48
				$\varepsilon=4^g$				
C2'endo	0.39	144.9	55.5	-153.1	176.2	179.2	-1.65	0
C2'endo	0.40	148.0	-172.9	-162.4	180.0	-179.9	-1.31	0.34
C2'endo	0.40	150.0	-66.6	-169.9	-179.9	-179.9	-0.13	1.52

Table IX. Conformational Energies for Deoxyadenosine (kcal/mole) continued.

Pucker	q[a]	W[a]	γ[b]	χ[b]	3'OH[c]	5'OH[d]	E[e]	ΔE[f]
				χ Dependence				
				ε=1[g]				
C2'endo	0.40	146.1	51.3	-158.4	176.5	171.4	-47.46	0
C2'endo	0.40	166.7	-63.3	60.8	-179.1	179.8	-41.52	5.94
				ε=4[g]				
C2'endo	0.39	145.4	55.3	-141.6	176.5	179.2	3.24	0
C2'endo	0.40	144.0	52.6	37.6	180.0	179.6	1.84	-1.40

[a] q and W defined in ref. 76. [b] γ and χ defined in ref. 77. [c] 3'OH refers to C4'-C3'-O3'-HO3' dihedral. [d] 5'OH refers to HO5'-O5'-C5'-C4' dihedral. [e] Absolute molecular mechanical energy. [f] Relative conformational energy. [g] ε is the dielectric screening factor used in equation (1).

Table X. χ Angle Profile for Base with Sugar Fragment (kcal/mole)

	Model of Deoxyadenosine		
	ab initio	AMBER (ε=1)	
χ [a,b]	MP2/6-31G*// HF/6-31G*	no specific dihedral	with specific dihedral
60	0.94	4.63	1.53
min	0.63 (74.7°)[d]	4.62 (61.1°)[d]	1.45 (54.5°)[d]
120	3.37	5.48	5.07
180	0.06	0.38	0.40
min	0.00 (198.2°)[d]	0.00 (196.7°)[d]	0.00 (197.2°)[d]
210	0.22	0.20	0.15
240	1.45	1.58	1.20
300	5.33	6.57	3.61
360	---	9.68	4.72

	Model of Deoxythymidine		
	ab initio	AMBER (ε=1)	
χ [a,b]	MP2/6-31G*// HF/6-31G*	no specific dihedral	with specific dihedral
60	2.27	6.00	2.94
min	2.02 (72.3°)[d]	5.99 (61.0°)[d]	2.83 (55.4°)[d]
120	7.02	8.48	8.05
180	0.74	1.37	1.43
min	0.00 (210.0°)[d]	0.00 (205.2°)[d]	0.00 (205.5°)[d]
210	0.00	0.05	0.03
240	1.29	1.77	1.40
300	8.15	8.94	5.82
360	---	13.11	8.18

[a] Ref. 77. [b] Degrees [c] Specific V_1 and V_2 dihedral terms were added for OS-CT-N*-CK (purines) and OS-CT-N*-CM (pyrimidines) dihedral angles. [d] Minimized value of χ.

Table XL Conformational Energies of Glycyl and Alanyl Dipeptides (kcal/mole).

	glycyl dipeptide			alanyl dipeptide	
	E(MM)	E(QM) [a]		E(MM)	E(QM) [a]
C7	0.0	0.0	C7$_{eq}$	0.0	0.0
C5	2.0	2.0	C7$_{ax}$	1.6	2.1
α_R	5.9	4.0	C5	1.5	1.5
			α_R	3.9	3.9

[a] Quantum mechanical energies calculated at MP2/TZP//HF/6-31G* level on methyl-blocked versions of the dipeptides. See ref. 28 for further details.

Table XII. Solvation Free Energies for Model Compounds (kcal/mole)

Molecule	$\Delta\Delta G$ (calc)	$\Delta\Delta G$ (exp)
$CH_4 \rightarrow$ nothing	2.5 ± 2 [a]	2.0 [b]
$C_2H_6 \rightarrow CH_4$	0.0 ± 0.1 [c]	-0.2 [b]
$C_3H_8 \rightarrow C_2H_6$	0.2 ± 0.1 [c]	0.2 [b]
$CH_3OH \rightarrow CH_3CH_3$	6.9 ± 0.1 [d]	6.9 [b]
$NMA \rightarrow CH_4$	11.6 ± 0.2 [d]	12.1 [e]
$CH_3NH_3^+ \rightarrow$ nothing	87.6 ± 2.0	$77\text{-}79$ [g]
	(75.4 ± 1.7) [f]	
$CH_3CO_2^- \rightarrow$ nothing	87.1 ± 1.2	$70\text{-}71$ [h]
	(71.6 ± 1.0) [f]	
$CH_3SCH_3 \rightarrow CH_3OCH_3$	0.9 ± 0.1 [h]	0.4 [i]
$CH_3OH \rightarrow CH_3SH$	3.5 ± 0.1 [h]	3.7 [i]
$9\text{-}CH_3$ adenine $\rightarrow CH_4$	18.6 ± 2.7 [h]	----

[a] Ref. 21. [b] Ref. 79. [c] Ref. 21. [d] Ref. 20. [e] ref. 80. [f] Ref. 81, additive potential, values in parentheses are for non-additive potential. [g] ref. 82. [h] This paper. [i] Ref. 83.

Table XIII. Comparison of Cornell *et al.* (AMBER), Weiner *et al.* (AMBER), CHARMm, OPLS/AMBER, and GROMOS Force Fields

force field	electrostatics	van der Waals	VDW combining rules [a]	dihedrals
● CHARMm [93] (1983)	emp. fit to QM dimers	empirical (x-tals)	R* arithmetic mean; ε geometric mean	single bond-path [b]
● GROMOS [92]	empirical	empirical (x-tals)	A and B "non-standard" geometric mean [c]	user-specified
● OPLS/AMBER [15] (1990)	empirical (MC on liqs)	empirical (liquids)	A and B geometric means	equal division among equiv. bond-paths
● AMBER [5,6] (Weiner *et al.*, 1984)	ESP-fit (STO-3G)	empirical (x-tals)	R* arithmetic mean; ε geometric mean	equal division among equiv. bond-paths
● AMBER (Cornell *et al.*, 1994)	RESP-fit (6-31G*)	empirical (liquids)	R* arithmetic mean; ε geometric mean	equal division among equiv. bond-paths

[a] $A=\varepsilon R^{*12}$ and $B=2\varepsilon R^{*6}$. [b] In CHARMm22, the torsion representation was changed to the more commonly used equal division of the energy along equivalent bond paths. [c] GROMOS employs the geometric mean method for calculating VDW interactions, but for water - methyl interactions, for example, a smaller VDW radius is assumed for the water since it is no longer in a hydrogen bonding interaction. This has been shown to result in a "too hydrophilic" methyl group [97,98].

Chapter 6

Inclusion of Non-additive Effects Improves the Molecular Mechanical

Phi-Psi Maps Calculated for Alanyl and Glycyl Dipeptides

Inclusion of Non-additive Effects Improves the Molecular Mechanical Phi-Psi Maps Calculated for Alanyl and Glycyl Dipeptides

by

Wendy D. Cornell, [1] James W. Caldwell, and Peter A. Kollman [1*]

Contribution from The Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA

[1] Graduate Group in Biophysics, University of California at San Francisco.

[*] To whom correspondence and reprint requests should be addressed.

Abstract

We present the phi-psi maps calculated for alanyl and glycyl dipeptides using three different molecular mechanical models: 1) the parameters developed for our additive force field with phi and psi dihedral parameters set to 0.0; 2) those parameters plus a set of optimized phi and psi dihedral parameters; and 3) a non-additive variation of the additive force field which includes atom-centered polarization and phi and psi dihedral parameters set to 0.0. We compare the energies calculated using the three models for seven low energy conformers with energies obtained from high level *ab initio* calculations. Furthermore, we compare the energies calculated for 20 additional conformations which were found to be stationary points on the quantum mechanical potential energy surface. Our results show that the molecular mechanical energies calculated for alanyl dipeptide using the additive force field are improved by the addition of dihedral parameters which were optimized on the small subset of seven low energy conformations. The energy of the single additional minimum energy conformer of glycyl dipeptide is also improved. Moreover, when polarization was added to the molecular mechanical Hamiltonian, the results for alanyl dipeptide were of comparable quality to those obtained using the additive force field with optimized phi and psi dihedral parameters and were superior to those obtained using the additive force field without those optimized dihedral parameters. The low energy conformations of glycyl dipeptide were similarly improved. This improvement is seen to derive partly from the polarization contribution to the conformational energies and partly from the reduced electrostatic contribution arising from the reduced point charges. The non-additive model is also shown to result in dipole moments which are in excellent agreement with gas phase values estimated from the HF/6-31G* quantum mechanical wavefunction.

Introduction

The accurate reproduction of the conformational preferences of the peptide backbone is a necessary feature of any molecular mechanical force field intended for modelling protein structure and energetics. These preferences can be gleaned from structural analyses of known protein crystal structures, but a more direct way of determining the relative energies of different conformations is to carry out high level *ab initio* calculations. Such calculations have been carried out in this laboratory. [1,2] This reference data then allows for the critical evaluation and calibration of the molecular mechanical model. Here we evaluate the ability of additive and non-additive molecular mechanical models to reproduce the quantum mechanical potential energy surface for the glycyl and alanyl dipeptides.

Methods

The restrained electrostatic potential-fit (RESP) charges which were derived for the alanyl (ALA), glycyl (GLY), acetyl (ACE), and N-methyl (NME) groups for the new additive force field were employed. [3-6] Van der Waals and bonded parameters were also taken from the new force field. [6] Charges used in the non-additive calculations were obtained by scaling the additive charges by a factor of 0.88. [7] Isotropic atomic polarizabilities were taken from the work of Applequist. [8] The values are N= 0.530, H(N)= 0.160, C(H)= 0.878, H(C)= 0.135, C(O)= 0.616, and O(C)= 0.434 Å^3. The 1-4 van der Waals interactions were scaled by a factor of 1/2 and the 1-4 electrostatic interactions by a factor of 1/1.2. [4] The simple additive (ADD) and non-additive (NON-ADD) calculations employed dihedral parameters of 0.0 for the phi and psi dihedrals (C-N-CT-C, C-N-CT-CT, N-CT-C-N, and CT-CT-C-N) as in the Weiner *et al.* force field. [9] An optimized version of the additive model (ADD+DIHED) included phi and psi dihedral parameters which were optimized to reproduce the

relative energies of the C7, C5, and α_R conformations of glycyl dipeptide and the C7$_{eq}$, C7$_{ax}$, C5, and α_R conformations of alanyl dipeptide.

The quantum mechanical reference energies were derived from two separate studies, one at the MP2/TZVP//HF/6-31G** level which employed methyl-blocked dipeptides [1,2] and the other at the MP2/6-31+G**//HF/6-31+G* level which employed H-blocked dipeptides. [10] Methyl-blocked versions of the dipeptides were employed in the molecular mechanical studies, because this analog more closely approximates a larger peptide or protein and because our amino acid charges were derived using the methyl-blocked analogs.

A Fourier series is used to represent the contribution for the dihedral energy. The C-N-CT-C and N-CT-C-N parameters were used to adjust the conformational energies of both alanyl and glycyl dipeptides. The alanyl dipeptide conformational energies were further adjusted using the C-N-CT-CT and CT-CT-C-N parameters. The optimized values are C-N-CT-C: V2= 0.2 (γ= 180); N-CT-C-N: V4= 0.4 (γ= 180), V2= 1.35 (γ= 180), V1= 0.75 (γ= 180); CT-CT-N-C: V4= 0.5 (γ= 180), V3= 0.15 (γ= 180), V1= 0.53 (γ= 0); CT-CT-C-N: V4= 0.1 (γ= 0), V2= 0.07 (γ= 0). The large contributions from the two- and four-fold terms can be rationalized in terms of the sp^2 hybridization about the carbonyl carbon and the nitrogen (one of the central atoms in each of the dihedrals).

Results and Discussion

We begin by considering the energies of the seven low energy conformers calculated using the two additive models (Tables I and II). The energies of these seven conformers were used to optimize the dihedral parameters for the additive model and are therefore considered separately. The reference energies for these conformers as

well as glycyl dipeptide β_2 and alanyl dipeptide α_L, β, and β_2 conformations were calculated using the methyl-blocked analogs at the MP2/TZVP//HF/6-31G** level of theory, [1,2] whereas the reference energies for the remaining conformers were taken from results obtained by Head-Gordon *et al.* on the H-blocked analogs at the MP2/6-31+G**//HF/6-31+G* level of theory. [10] The latter energies were used as an approximation to the energies of the methyl-blocked analogs, in order that we might evaluate the general features of the phi-psi space.

We can estimate the errors made in using the energies of the H-blocked analogs to approximate the energies of the methyl-blocked analogs by considering the seven conformers for which data is available for both analogs (glycyl dipeptide C7 and C5 and alanyl dipeptide $C7_{eq}$, $C7_{ax}$, C5, α_L, and β_2). The relative energies calculated for the glycyl dipeptide C5 conformation differ by 0.9 kcal/mol. Less disagreement is seen with alanyl dipeptide, where the $C7_{ax}$ relative energies are nearly identical and the C5 energies differ by 0.5 kcal/mol. The α_L energies are also nearly identical and the β_2 energies differ by about half of a kcal/mol. It should be noted that these are all minimum energy conformations, so less agreement is possible for the transition structures and maxima where the geometries are more crowded. One example of an obvious disparity in the transition structures is the glycyl dipeptide C7 --> C5 transition structure. The quantum mechanical energy for the H-analog is 1.8 kcal/mol, which is higher in energy than the C7 and C5 minima which have relative energies of 0.0 and 1.1 kcal/mol, respectively. The energy of the transition structure must be higher for the methyl-blocked analog, since for that analog the C5 minimum has a relative energy of 2.0 kcal/mol. Nonetheless, overall the agreement is good enough to warrant using the quantum mechanical energies from the H-analogs as a reference against which to compare the molecular mechanical energies of the methyl-blocked analogs.

The simple additive model (no optimized dihedral parameters) does not do a very good very job of reproducing the quantum mechanical energies. The $C7_{ax}$ conformation of alanyl dipeptide is over 1 kcal/mol too low in energy and the C5 and α_R conformations are too high in energy. The glycyl dipeptide C5 conformation is also too high in energy, but by about the same amount as the alanyl dipeptide C5, making those conformations easier to adjust with common phi and psi dihedral parameters. The glycyl dipeptide α_R is 3.2 kcal/mol too high in energy as compared to alanyl dipeptide α_R which is only 1.4 kcal/mol too high. This difference makes it harder to adjust both conformational energies, even given the two additional dihedral parameters available to alanyl dipeptide (C-N-CT-CT and CT-CT-C-N) through its β-carbon. Because the alanyl dipeptide α_R was closer to it quantum mechanical energy than the glycyl dipeptide α_R, it was easier to adjust the alanyl dipeptide energy. Hence in the additive model with optimized dihedral parameters, the conformation most in error is glycyl dipeptide α_R.

In addition, the alanyl dipeptide $C7_{ax}$ conformation is still too low in energy in the optimized additive model. It was possible to raise the energy of that conformation, but doing so adversely affected the relatively low energy (1.8 kcal/mol) of the $C7_{eq}$ -- > C5 transition structure. Since the $C7_{ax}$ conformation is rarely found in proteins, [11] we decided that it was more important to emphasize the heavily populated $C7_{eq}$/C5 region of phi-psi space. [12] The additive model with optimized dihedral parameters resulted in significantly reduced errors in the additional conformations of alanyl dipeptide which were examined. This was not the case with glycyl dipeptide, however, there are only two low energy structures in the group of additional conformations (C7 --> C5 and β_2). The β_2 conformation is much better fit with the

optimized additive model while the energy of the transition structure is raised further by about half of a kcal/mole.

The simple non-additive model (no optimized dihedral parameters) greatly improves the energies calculated for the $C7_{(eq/ax)}$, C5, and α_R conformations of both alanyl and glycyl dipeptides. The energies of those three glycyl dipeptide conformations are especially well reproduced, as judged by either the sum of the absolute differences between the QM and MM energies or the RMS of the difference between those energies. The energies of the additional conformations of alanyl dipeptide are also much improved over the simple additive model. The RMS is lower than seen with both the simple and optimized additive models. If only the six additional conformers which are under 5.0 kcal/mol are considered, the sum of the absolute errors is 14.7, 8.2, and 7.6 kcal/mol and the RMS errors 2.8, 1.7, and 1.4 for the simple additive, optimized additive, and simple non-additive models, respectively. The additional low energy glycyl dipeptide structures do not realize this systematic improvement, but are still fairly well reproduced with both the optimized additive and non-additive models.

Examination of the energy components of the minimized low energy structures reveals that the improvement seen with the addition of polarization is partly due to the polarization energy itself and partly due to the reduced contribution from the "permanent" electrostatic interaction. The glycyl dipeptide C5 conformation and the glycyl and alanyl α_R dipeptide conformations are more stabilized by the polarization energy relative to the respective C7 conformations by about 0.9 kcal/mol. The alanyl dipeptide $C7_{ax}$ conformation is destabilized by the polarization energy by about 0.3 kcal/mol. For both the alanyl and glycyl dipeptides, the greater electrostatic stabilization of the C7, $C7_{eq}$, and $C7_{ax}$ conformations is reduced over the additive model by 1.0-1.5 kcal/mol.

Finally, we can examine the conformationally dependent dipole moments calculated using the various molecular mechanical models and compare the results with those obtained from quantum mechanics. The conformationally dependent dipole moments are presented for the glycyl and alanyl dipeptides in Tables V and VI. The molecular mechanical dipole moments given are for the structures minimized with each particular molecular mechanical model. The quantum mechanical dipole moments shown in the first column are derived from the 6-31G* basis set. Because this dipole moment is known to overestimate dipole moments by about 20%, the last column shows the empirically scaled back dipole moments. These scaled dipole moments are used as the reference against which to compare the ones calculated from molecular mechanics. A better set of reference dipole moments could be derived from quantum mechanical calculations using a basis set or methodology which is known to reproduce gas phase values.

Three molecular mechanical models are presented. The first is the simple standard additive model, the second is the non-additive model , and the third is the simple additive model but using the empirically scaled non-additive charges. The standard additive model uses the "full-strength" 6-31G* derived charges. Because it does not allow for polarization, the additive model must implicitly contain the amount of polarization that would be expected in solution. This model should therefore ideally reproduce the dipole moment calculated from the 6-31G* wavefunction. The additive model does this fairly well for the glycyl dipeptide C7 and C5 conformations and the alanyl dipeptide $C7_{eq}$, $C7_{ax}$, and C5 conformations. However, this model overestimates the dipole moments of the two α_R conformations by about 1 D. This could result in the α-helical conformation being overstabilized in solution with respect to the other conformations.

The non-additive model, on the other hand, does allow for explicit charge polarization. We would therefore hope that the dipole moments calculated using this model would then be "gas phase-like," and not exhibit the increased charge polarization necessary with the additive model. The non-additive model does in fact result in dipole moments which agree well with the empirically scaled quantum mechanical values. The two components of the non-additive dipole moments are also given. The permanent components arising from the fixed point charges alone are nearly identical to the dipole moments arising from the additive model using the smaller non-additive charges. The slight differences between the two sets of dipole moments arise from the small differences in minimized geometries. The induced components, arising from the atom-centered dipoles, are shown to be small for the glycyl dipeptide C7 and the alanyl dipeptide $C7_{eq}$ and $C7_{ax}$ conformations. The two C5 conformations have induced dipole moments of about 0.5 D and the two α_R conformations have induced dipole moments of about 1.3 D. These induced dipole moments are almost directly opposed to the permanent components and result in net dipole moments which are in nearly exact agreement with the scaled quantum mechanical values. The additive model which employs the smaller non-additive charges also results in dipole moments which agree fairly well with the scaled quantum mechanical values, however, the dipole moments for the α_R conformations are still too high by about 1.2 D. Clearly, the induced component of the dipole moment which arises in the non-additive model plays an important role. The dipole moments calculated using the non-additive model are thus in excellent agreement with the quantum mechanical reference values for all of the conformations examined.

Conclusion

We have presented the phi-psi maps calculated for the glycyl and alanyl dipeptides using three different molecular mechanical models. While the simple additive model does not do well at reproducing the relative quantum mechanical energies, this model can be greatly improved through the addition of optimized dihedral parameters. Furthermore, the simple model which incorporates atomic polarization does about as well as the optimized additive model and significantly better than the simple additive model. The non-additive model is also shown to result in dipole moments which are in excellent agreement with those obtained using quantum mechanical moments which have been empirically scaled back to estimate gas phase values. Additional tests will have to be carried out on more complex model systems and on proteins to assess further the performance of both the additive and non-additive models.

References

1. Gould, I.R.; Kollman, P.A. *J. Phys. Chem.* **1992**, *96*, 9255.

2. Gould, I.R.; Cornell, W.D.; Hillier, I.H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.

3. Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. *J. Phys. Chem.* **1993**, *97*, 10269.

4. Cornell, W.D.; Cieplak, P.; Bayly, C.I., Kollman, P.A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

5. Cieplak, P.; Cornell, W.D.; Bayly, C.I.; Kollman, P.A. *J. Comp. Chem.*, accepted.

6. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M., Jr.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A., submitted.

7. Caldwell, J.W.; Kollman, P.A., submitted.

8. Applequist, J.B.; Carl, J.R.; Fung K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.

9. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230.

10. Head-Gordon, T.; Head-Gordon, M.; Frisch, M.; Pople, J.A.; Brooks, C.L. *J. Am. Chem. Soc.* **1991**, *113*, 5989.

11. The $C7_{ax}$ conformation is found in γ-turns.

12. Ramachandran, G.N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 284.

Table I. Energies of key glycyl dipeptide conformations -- training set (kcal/mol).

conformation	ϕ^a	ψ^a	E(QM) methyl-analog[b]	E(QM) H-analog[c]	E (MM) (ADD)	E(MM) (ADD+DIHED)	E(MM) (NON-ADD)
C7	-85.5	72.0	0.0	0.0	0.0	0.0	0.0
C5	180.9	180.5	2.0	1.1	3.9	1.9	2.1
α_R	-60.7	-40.7	4.0		7.2	6.0	4.8
$\Sigma |(E(QM)-E(MM))|^d$					5.1	2.1	0.9
RMS(E(QM)-E(MM))[d]					2.1	1.2	0.5

[a] ϕ and ψ refer to the quantum mechanics geometry optimized values of these two parameters, except for the α_R conformation which was geometry optimized wtih constraints on ϕ and ψ. Values reported are for the methyl-analogs. In the molecular mechanics calculations, ϕ and ψ were allowed to optimize for the C7 and C5 conformations but were constrained for the α_R conformation. [b] MP2/6-31+G**//HF/6-31+G* calculations (ref. 10). [c] MP2/TZVP//HF/6-31G** calculations (refs. 1 and 2). [d] Sum of differences and RMS values calculated using QM energies of methyl-analogs.

Table II. Energies of key alanyl dipeptide conformations -- training set (kcal/mol).

conformation	ϕ[a]	ψ[a]	E(QM) methyl-analog[b]	E(QM) H-analog[c]	E (MM) (ADD)	E(MM) (ADD+DIHED)	E(MM) (NON-ADD)
C7eq	-86.1	78.8	0.0	0.0	0.0	0.0	0.0
C7ax	76.0	-55.4	2.1	2.2	1.0	1.5	1.6
C5	-157.1	159.8	1.5	1.1	3.1	1.5	2.1
αR	-60.7	-40.7	3.9		5.6	3.9	3.3
Σ I(E(QM) -E(MM))I[d]					4.4	0.6	1.7
RMS(E(QM)-E(MM))[d]					1.3	0.3	0.5

[a] ϕ and ψ refer to the quantum mechanics geometry optimized values of these two parameters, except for the αR conformation which was geometry optimized with constraints on ϕ and ψ. Values reported are for the methyl-analogs. In the molecular mechanics calculations, ϕ and ψ were allowed to optimize for the C7eq, C7ax, and C5 conformations but were constrained for the αR conformation. [b] MP2/6-31+G**//HF/6-31+G* calculations (ref. 10). [c] MP2/TZVP//HF/6-31G** calculations (refs. 1 and 2). [d] Sum of differences and RMS values calculated using QM energies of methyl-analogs.

Table III. Energies of key glycyl dipeptide conformations -- test set (kcal/mol).

conformation	ϕ^a	ψ^a	E(QM) methyl-analog[b]	E(QM) H-analog[c]	E (MM) (ADD)	E(MM) (ADD+DIHED)	E(MM) (NON-ADD)
β_2	-116.2	19.9	3.3		6.1	3.6	5.1
C7 --> C5	-85.3	126.7		1.8	3.7	4.3	2.5
C7 --> C7	0.0	0.0		9.1	9.9	6.5	10.3
C_s cusp	-180.0	0.0		9.6	9.8	6.3	6.4
C7 --> C7 (alt)	-0.9	79.9		9.6	8.5	8.3	6.8
C_s max	0.0	180.0		22.7	23.7	21.7	19.3
$\Sigma |(E(QM)-E(MM))|^d$					7.8	11.0	13.1
RMS(E(QM)-E(MM))[d]					1.5	2.1	2.4

[a] ϕ and ψ refer to the quantum mechanics geometry optimized values of these two parameters. Values reported are for the H-analogs, except for the β_2 conformation where they refer to the methyl-blocked analog. In the molecular mechanics calculations, ϕ and ψ were constrained to the given values. [b] MP2/6-31+G**//HF/6-31+G* calculations (ref. 10). [c] MP2/TZVP//HF/6-31G** calculations (refs. 1 and 2). [d] Sum of differences and RMS values calculated using QM energies of methyl-analogs, when available, and otherwise using QM energies of H-analogs.

Table IV. Energies of key alanyl dipeptide conformations -- test set (kcal/mol).

conformation	φ	ψ	E(QM) methyl-analog[b]	E(QM) H-analog[c]	E(MM) (ADD)	E(MM) (ADD+DIHED)	E(MM) (NON-ADD)
αL	67.0	30.2	4.4	4.5	8.6	7.7	6.3
β	-57.6	134.4	4.1		4.0	3.6	2.7
β2	-130.9	22.3	3.3	2.7	6.5	3.8	4.7
α'	-165.6	-40.7		5.8	7.8	6.8	5.2
C7eq --> β2	-106.0	20.0		2.7	5.2	3.0	3.7
C7eq --> C5	-100.6	133.3		1.8	3.0	3.7	2.1
C7ax --> αL	72.3	14.4		4.5	8.0	6.2	6.1
C7ax --> α'	138.9	-28.7		7.6	10.0	9.6	8.9
C7eq --> C7ax	2.6	-45.3		9.7	6.6	6.6	4.2
C5 --> C7ax	120.3	-153.0		7.9	8.0	7.1	7.6
β2 --> α'	-119.1	-48.9		8.0	9.4	9.5	7.2
C5 --> α'	-149.6	-94.5		6.7	13.6	15.1	9.6
C7eq --> αL	2.6	82.2		9.4	11.0	9.4	9.4
αL --> αD/C5	86.5	85.1		10.6	7.5	6.9	4.6
Σ l(E(QM) -E(MM))l[d]					35.3	28.7	25.0
RMS(E(QM)-E(MM))[d]					3.0	2.9	2.5

[a] φ and ψ refer to the quantum mechanics geometry optimized values of these two parameters. Values reported are for the H-analogs, except for the αL, β, and β2 conformations where they refer to the methyl-blocked analogs. In the molecular mechanics calculations, φ and ψ were constrained to the given values. [b] MP2/6-31+G**//HF/6-31+G* calculations (ref. 10). [c] MP2/TZVP//HF/6-31G** calculations (refs. 1 and 2). [d] Sum of differences and RMS values calculated using QM energies of methyl-analogs, when available, and otherwise using QM energies of H-analogs.

Table V. Dipole moments of glycyl dipeptide conformations (D). [a]

conformation	QM(6-31G*) [b]	MM/ADD	MM/NON-ADD			MM/ADD (scaled q) [c]	scaled QM (0.83*6-31G*) [d]
			perm	induced	total		
C7	3.28	3.23	2.72	0.23	2.87	2.66	2.72
C5	2.88	3.06	2.64	0.54	2.10	2.69	2.39
α_R	6.67	7.56	6.75	1.19	5.70	6.73	5.54
$\Sigma |(\mu(QM) -\mu(MM))|$ [e]					0.60	1.55	
RMS$(\mu(QM)-\mu(MM))$ [e]					0.21	0.71	

[a] Methyl-blocked analog. [b] Structures optimized at HF/6-31G** level. Additive model but using the scaled charges from the non-additive model. [d] Adjusted to account for the ~20% enhancement over gas phase dipole moments seen with the 6-31G* basis set. [e] The scaled QM dipole moments are used as a reference.

Table VI. Dipole moments of alanyl dipeptide conformations (D). [a]

conformation	QM(6-31G*) [b]	MM/ADD	MM/NON-ADD			MM/ADD	scaled QM
			perm	induced	total	(scaled q) [c]	(0.83*6-31G*) [d]
$C7_{eq}$	2.90	3.15	2.69	0.22	2.55	2.53	2.41
$C7_{ax}$	3.89	3.51	3.09	0.13	3.07	2.95	3.23
C5	2.58	2.30	2.20	0.41	1.80	1.99	2.14
α_R	6.58	7.52	6.69	1.38	5.37	6.68	5.46
$\Sigma|(\mu(QM)-\mu(MM))|$ [e]					0.73	1.77	
$RMS(\mu(QM)-\mu(MM))$ [e]					0.21	0.63	

[a] Methyl-blocked analog. [b] Structures optimized at HF/6-31G** level. [c] Additive model but using the scaled charges from the non-additive model. [d] Adjusted to account for the ~20% enhancement over gas phase dipole moments seen with the 6-31G* basis set. [e] The scaled QM dipole moments are used as a reference.

Chapter 7

Conclusions and Future Directions

I have presented the derivation of a new second generation force field for the simulation of proteins, nucleic acids, and many small molecules. A goal of this work has been to put the development of the force field on a more explicitly stated algorithmic basis, with the philosophy that "if the assumptions, approximations, and inevitable imperfections in a force field are at least known, one can strive for some cancellation of errors."

What, then, *are* the assumptions, approximations, and inevitable imperfections in this new additive force field. We have developed this force field with the assumption that it should be balanced with respect to TIP3P type (i.e. additive) water models. Because the force field is additive, the charge distribution which is derived from the gas phase quantum mechanical model is unable to respond to the polar environment of an aqueous simulation. The atom-centered charges were calculated using the 6-31G* quantum mechanical basis set, with the assumption that this basis set would provide approximately the correct amount of polarization over the actual gas phase dipole moment. The use of atom centered point charges is perhaps the approximation of greatest significance, since many biomolecular systems ae dominated by electrostatic interactions. Other charge models have been proposed, such as an atom centered multipolar expansion, which is better able to reproduce the anisotropic distribution of electron density around the nucleus. It is not clear how well such a model would perform when subjected to the molecular distortions which result during molecular dynamics or even minimization, since the higher order moments (dipole and beyond) have directional components which would not be well defined upon a change in geometry about a given nucleus.

A demonstrated imperfection of most electrostatic potential fit charge models is the conformational dependence of the resulting charges (and/or dipoles, quadrupoles,

etc.). We have attempted to address this problem through the application of multiple conformation fitting. Nonetheless, it is impossible from a practical standpoint to include all of the low energy conformations in the charge fit (or even to know them *a priori*), and for that reason most charge sets will perform well at reproducing the dipole moments of the conformations used in the charge fit but not perform as well at reproducing the dipole moments of other conformations. Yet another imperfection is introduced by the necessity to constrain methyl hydrogens, and other atoms or groups which are not equivalent in a static conformation but interconvert during molecular dynamics, to have equivalent charges. In the old standard ESP model, methyl hydrogen charges were averaged after the fit and this could result in a significant change in the resulting dipole moment. This problem is reduced with the new two-stage RESP charge model, where methyl hydrogens are refit during the second stage. The resulting charges perform better at reproducing the quantum mechanical dipole moment than the "*a posteriori* averaged" standard ESP charges did. Nevertheless, there are still cases, such as the α_R conformation of alanyl dipeptide, where the refit methyl charges significantly change the dipole moment.

These errors in the dipole moments could result in errors in relative interaction energies when different conformations of a molecule or amino acid side chain are being compared. For example, the continuums solvation model of Onsager suggests that the free energy of solvation of a molecule is expected to scale as the square of the dipole moment. When Cieplak calculated the relative free energy of solvation for methanol using different charge models, the results were consistent with this prediction.

Another inaccuracy arising from the use of our atom centered charge model has been studied by Dixon. He found that for atoms containing sp^2 nitrogens, the difference in

the interaction energy between the nitrogen as a hydorgen bond acceptor and a water molecule as a donor at 90° and 180° relative to the sp^2 plane is grossly underestimated (up to 3.5 kcal/mol for pyridine) using the atom centered RESP charge model (6-31G*). This error results from the inability of the atom centered charge model to reproduce the anisotropic distribution of charge around the nucleus, and is easily rectified through the addition of an off-center charge. A similar underestimation of the non-co-planar interaction has been seen with the DNA base pairs. Fortunately, this problem is not evidenced with sp^3 nitrogens or sp^2 oxygens as hydrogen bond acceptors, nor with any nitrogen, oxygen, or sulfur atoms as donors. In general, sp^3 oxygen and sulfur could be somewhat improved through the addition of extra charge centers.

A better representation of the true charge distribution is obviously possible through the inclusion of additional charge centers. We have chosen to avoid this approach with our new force field for two reasons. First, the additional center is difficult to manipulate during MD because, unlike a charge centered on a nucleus, the center has no true mass. Coordinate resetting during SHAKE can be time consuming. Second, the value of the extra charge is determined for the specific geometry of the molecule employed in the charge fit. When the structure is distorted during MD, the surrounding bonded atoms will be repositioned, and the new optimal position of the charge is now different. Although additional charge centers may be employed in certain specific calculations, we believe their implementation is nt sufficiently welldefined for use in a general model.

In spite of these limitations of the atom centered 6-31G* RESP charge model, we still believe this charge model to be the best choice. Overall, the charges perform well at reproducing conformational energies and most interaction energies and free energies

of solvation. Furthermore, the charge model is well defined and requires less computer time to generate the charges compared to the OPLS force field, which relies on empirical optimization based on Monte Carlo liquid simulations, or the new CHARMm force field which employs fits to the interaction energies of a variety of molecule-water configurations calculated at the MP2 level of theory.

Another assumption built into this and most force fields is the transferability of the parameters. Indeed, a basic tenet of organic chemistry is that a given functional group tends toward certain properties, whatever the environment. With force field development, however, a known problem is the coupling between the 1-4 van der Waals and dihedral energies. Many force fields, such as MM2/MM3, contain a large number of dihedral parameters which have been optimized to reproduce the energies and structures of a large number of molecules. In the Weiner *et al.* and Cornell *et al.* force fields, however, only a limited number of dihedral parameters are necessary, reducing the question of transferability to molecules not included in the initial parameterization. One notable exception are the phi and psi parameters for the peptide backbone. A relatively elaborate set of such parameters was required to reproduce the conformational energies of the alanyl and glycyl dipeptides, and even then, the glycyl dipeptide α_R and the alanyl dipeptide $C7_{ax}$ conformations were not well reproduced. The source of this non-transferability is not evident at the present time and merits further study. Further studies are planned on a large set of small molecules representing different functional groups and those results should provide more information on the need for additional dihedral parameters and the transferability of all of the dihedral parameters.

The van der Waals interaction is represented by an R^6 dispersion attraction and an R^{12} exchange repulsion interaction, both of which are approximations. The

dispersion attraction arises from the interaction of induced dipoles, and has the well defined mathematical form of $-aR^6 -bR^8 -cR^{10} -...$ The series is truncated after the R^6 term, but because the van der Waals parameters are optimized to reproduce the desired properties (density and enthalpy of vaporization for liquids and lattice distances and energies for crystals), an effective set of parameters is obtained rather than the actual set of "a" coefficients. The R^{12} exchange repulsion term has the Pauli Principle as its basis, and arises from the fact that two particles are not allowed to have the same quantum numbers and occupy the same physical space. Other researchers have shown that this interaction is better represented with an R^9 or an exponential expression. The earlier popularity of the R^{12} representation was partly due to its ease of computation, with the distance term calculated simply as the square of the already calculated sixth power of the distance. While the R^{12} representation is admittedly too repulsive at short distances, for the equilibrium properties and unstrained systems which we are primarily interested in modelling, it is quite adequate.

Another approximation in our force field is the use of a simple harmonic diagonal potential for modelling the bonded interactions. Other force field developers have focused their efforts on reproducing the entire potential energy surface of a molecule by going beyond the harmonic representation and including cross-terms in order to ensure accurate reproduction of the entropy of the system. While these more elaborate potentials are admittedly more accurate at preproducing the gas phase potential energy surface, we have developed a force field primarily for solution calculations, and when explicit water models are employed, the errors in the electrostatic and van der Waals interactions most likely far outweigh and inaccuracies in the representationof bonded interactions. This philosophy was best articulated by van Gunsteren and Berendsen:

"We note that the choice of a particular force field should depend on the system properties one is interested in. Some applications require more defined force fields than others. Moreover, there should be a balance getween the levels of accuracy or refinement of different parts of a molecular model. Otherwise the computing effort put into a very detailed and accurate part of the calculations may easily be wasted due to the distorting effects of the cruder parts of the model." (van Gunsteren, W.F. and Berendsen, H.J.C. Angew. Chem. Int. Ed. Engl. 29, 992 (1990))

Finally, I turn to the non-additive model which was developed by Caldwell based on neat liquid simulations of methanol and NMA and tested here on glycyl and alanyl dipeptides and on the DNA base pairs. This model is based on using atom centered isotropic polarizabilities determined by Applequist using optical spectroscopy. Two approximations are evident here. First, the polarizabilities are assumed to be isotropic. Second, Applequist's polarizabilities were developed within a model where, unlike in molecular mechanics calculations, bonded atoms (1-2 and 1-3) are allowed to interact. The polarizabilities as implemented in our non-additive model do not then reproduce the molecular polarizability, but they do allow for local polarization of charge. Caldwell found that when the 6-31G* RESP charges were scaled by a factor of 0.88 and used in conjunction with the polarizabilities, the enthalpies of vaporization and densities of methanol and NMA were well reproduced. This model also performed well at reproducing the conformational energies of the glycyl and alanyl dipeptides and yielded more "gas phase-like" values for the conformational dipole moments. Unfortunately, this mdel did not perform well at reproducing the interaction energies of the DNA base pairs. The energies were

severely underestimated with this model, owing to the decreased point charges and the small contribution from the induced dipoles. Similar results were obtained using unscaled charge derived from a density functional charge distribution (which reproduces the gas phase dipole moment).

The reasons for the failure of this model are currently under investigation. It is possible that a model employing additional charge centers which better reproduces the molecular multipoles is more important for modelling base pair interactions than the inclusion of polarizability. Another possibility is that charge transfer plays a significant role in this interaction, and as this quantum mechanical effect is highly system and angle dependent, it is not included in the hamiltonian. The DNA bases are different from the residues found in proteins in that they are fairly rigid with a delocalized and asymmetric charge distribution, and for that reason a different representation of their electrostatic properties does not seem unreasonable. We have tried to avoid such distinctions, however, in our pursuit of a general model. The non-additive calculations discussed in this thesis represent some of the first tests of the model, and we are optimistic that either this model can be adapted or another non-additive model can be found which will be general and useful.

I have attempted to provide some insight into the assumptions, approximations, and inaccuracies present in both our new additive force field and also our current non-additive model. The limitations described shoujld provide excellent fodder for future work in the field. In spite of these limitations, we are optimistic that our new additive force field represents the current state-of-the-art for carrying our condensed phase biomolecular simulations within a general and extendable effective two-body model.

Appendix 1

Simulation of Cyclopentane Conformational Dynamics

Using a Simple Diagonal Force Field

Simulation of Cyclopentane Conformational Dynamics
Using a Simple Diagonal Force Field

by

Wendy D. Cornell, [1,2] Maria P. Ha,[2,3,4] Yax Sun, [1,2] and Peter A. Kollman [1,2*]

Contribution from The Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA

[1] Graduate Group in Biophysics, University of California at San Francisco.

[2] Science and Education Partnership Program.

[3] Lowell High School, San Francisco, CA.

[4] Current Address: University of California at Berkeley, Berkeley, CA.

[*] To whom correspondence and reprint requests should be addressed.

Abstract

Molecular dynamics simulations have been carried out on the cyclopentane molecule using the Weiner *et al.* force field and the results compared with both experiment and a recent study which used the MM3 force field (Cui, W.; Li, F.; Allinger, N.L. *J. Amer. Chem. Soc.* **1993**, *115*, 2943). The current simulation resulted in a pseudorotational velocity of 910 deg/ps, compared to the experimentally estimated value of 400 deg/ps and the MM3 result of 1700 deg/ps. The pseudorotation amplitude was calculated to be 0.46 ±0.02 A, compared to the experimental value of 0.48 A and the MM3 value of 0.5 ±0.03 A. The two distinct average C-H bond lengths seen for the axial and equatorial conformations in the MM3 simulation were not observed. The energy barrier to passing through the planar conformation was calculated at 4.7 kcal/mol as compared to the experimental value of 5.2 kcal/mol and the MM3 value of 4.2 kcal/mol. During the simulation, the angle bending term dominated the potential energy, followed by the torsion enery, as was seen with MM3. The third largest energy term was the bond stretching, followed by the van der Waals interaction, the reverse of what was seen with MM3. The effects of carrying out the simulation under conditions of constant energy versus constant temperature are discussed.

Introduction

Most of the molecular mechanical force fields in common use have been optimized for application either to small molecules or to biological macromolecules.[1-5] In the small molecule field, Allinger's MM2 and MM3 force fields[2] have served as the gold standard. Typical applications of the MM2 and MM3 force fields have included the calculation of highly accurate heats of formation, structures, conformational energies, and rotational barriers,[2] as well as the "chemically accurate" calculation of vibrational frequencies, absolute entropies, and entropies of activation for internal

rotation. [6] The MM2 program [7] allows for minimization of structures, but until recently there was no capability for carrying out molecular dynamics (MD) calculations. With the development of MM3-MD, [8] however, the MM3 force field can now be applied to MD simulations.

A recent study by Cui *et al.* [9] examined the performance of the MM3 force field under the conditions of MD as applied to the pseudorotation of cyclopentane. This study was of interest to us, because just as the MM3 force field is being expanded to treat polypeptides and proteins [10] and applied to MD simulations, a new diagonal force field has been developed in our laboratory, [11-13] with one of the goals being to model the structural and conformational energetic properties of small molecules to a "reasonable" degree of accuracy. This capability to model the conformational energies of small molecules in a level comparable to the of MM2/MM3 has already been demonstrated for simple unstrained molecules such as butane; methyl ethyl thioether; methyl, ethyl, and propyl alcohol; methyl, ethyl, and propyl amine; and 1,2-ethane diol. [12] In one class of molecules studied, the dioxanes, the Weiner *et al.* force field used with either our standard ESP [15] or new RESP [12-13] charge model actually significantly outperformed MM2 and MM3. This superior performance was attributed to the use of electrostatic potential fit charges in the Weiner *et al.* and Cornell *et al.*. force fields as compared to the bond dipoles used in MM2/MM3.

Most molecular mechanical force fields partition the energy of a system into common components: the strain energy associated with bonds and angles which are distorted from their equilibrium values, the quantum mechanical effects represented by the dihedral term, and the electrostatic and van der Waals interactions between atoms which are separated by at least two bonds. The was this energy is actually partitioned for a given system is highly variable among force fields, with the bond and angle

components probably being the most well defined and the remaining components being highly interdependent.

Because of their emphasis on the highly accurate representation of structures and vibrational frequencies, the MM2/MM3 force fields have a relatively complicated energy expression as compared to many of the macromolecular force fields. MM3 includes higher order terms (beyond the harmonic approximation) for both the bond and angle energies as well as cross-terms coupling bond-angle, bond-torsion, and angle-angle distortions. The Weiner *et al.* force field, on the other hand, was originally developed for modelling the energetics and structures of proteins and nucleic acids, and so there the emphasis was on the accurate representation of the electrostatic interactions. Bond and angle energies are represented in the Weiner *et al.* force field using the harmonic approximation and no cross-terms are included.

Only a limited number of force field comparisons have appeared in the literature, [16,17] and as the domains of the MM3 and macromolecular force fields begin to overlap more and more, this is perhaps an appropriate time to assess the strengths and weaknesses of the different approaches. We thus decided to follow the lead of Cui *et al.* [9] and carry out our own MD simulation of cyclopentane, so that we might compare the results obtained with our simpler force field with those obtained using MM3. Because this study is meant to serve as a comparison, we have based our analysis and most of our figures on the study by Cui *et al.* [9]

Cyclopentane is a simple molecule which lends itself well to a force field study such as this. It contains only two different atom types. Its conformational space is limited by its cyclic structure. The planar conformation is not favored because of the eclipsing of the ring dihedral angles. In fact, cyclopentane is known to have two

degenerate low energy conformations. [18] The bent conformation possesses C_s symmetry and consists of four of the carbons in a plane with the fifth carbon bent out of the plane. The twisted conformation possesses C_2 symmetry and consists of three contiguous carbons forming a plane with the remaining two carbons twisted to be situated symmetrically above and below that plane.

Because of the five-fold symmetry of the cyclopentane molecule, these two low energy conformations can be formed with different sets of atoms, and the molecule interconverts between them via a process called pseudorotation. In pseudorotation the endocyclic dihedral angles vary in a concerted fashion, with the angle of maximum pucker passing around the ring. [18] The wavelike movement of the puckering around the ring is modulated by a radial vibration which causes the degree of puckering to oscillate about an equilibrium value. The movement is not a true rotation since the motion of the atoms is perpendicular to the direction of the rotation. [18]

Methods

The calculations were carried out using the Weiner *et al.* all atom force field [1] and the AMBER package. [19] The cyclopentane molecule was first equilibrated for 50 ps at a constant temperature of 298 K using a coupling constant of 0.2. [20] Next, a constant energy run was carried out using the final coordinaes and velocities from the constant temperature run. This final set coordinates and velocities were fortuitously very close to the target temperature of 298 K, otherwise it would have been necessary to choose another set from near the end of the equilibration with the desired temperature. The system was further equilibrated at constant energy for 10 ps and then data was collected between 10 and 20 ps of the constant energy run.

A second set of data was collected at a constant temperature of 298 K. This run was started from the final set of coordinates and velocites from the 50 ps of equilibration at constant temperature. The temperature coupling constant was set to 0.2 for the 50 ps of equilibration and for the 10 ps of data collection.

A third set of data was collected at a constant temperature of 298 K but with the coupling constant relaxed to 0.5. This run was equilibrated for 50 ps starting from the final set of coordinates and velocities from the initial 50 ps of equilibration at constant temperature. Data was ten collected for 10 ps.

A fourth set of data was collected at a constant temperature of 293 K with a coupling constant of 0.2. This run was equilibrated for 20 ps starting with the final set of coordinates and velocities from the previous run. Data was then collected for 10 ps.

A fifth and sixth set of data were collected at a constant temperature of 298 K with the coupling constants relaxed to 0.9 and 2.0, respectively. These runs were equilibrated for 50 ps starting from the final set of coordinates and velocities from the initial 50 ps of equilibration at constant temperature. Each equilibration was followed by 10 ps of data collection.

A time step of 0.02 fs was employed for all simulations. The 1-4 electrostatic and van der Waals interactions were scaled by 1/2. A constant dielectric of 1.0 was employed and all non-bonded interactions were included. AMBER uses the leapfrog algorithm for MD integration. Coordinates were collected every 100 steps.

The pseudorotation phase angle, P, was calculated according to the formula of Altona et al. [21] as a function of the five endocyclic torsion angles:

$$\tan P = [(\theta_2+\theta_4) - (\theta_1+\theta_3)] / 2\theta_0 (\sin 36 + \sin 72)$$

The puckering amplitude, q, is also determined as a function of the five endocyclic torsion angles and represents the deviation from planarity of the ring:

$$q = \theta_m / 102.5$$

where

$$\theta_m = 2/5 (\Sigma \theta_j \cos \alpha_j)^2 + (\Sigma \theta_j \sin \alpha_j)^2$$

and

$$\alpha = 4\pi_j / 5$$

and the sum is taken over the five endocyclic torsion angles.

Results

I. Conformational Interconversion via Pseudorotation

The one major difference in protocol between our study and that of Cui et al. [9] was that we compared data which was obtained at constant energy with that obtained at constant temperature, whereas their study was limited to constant temperature simulations. The constant energy approach seemed more logical to us insofar as the concept of temperature is best applied to an ensemble of molecules. Furthermore, temperature coupling schemes affect the dynamics of the system under study, which is usually acceptable when one is focusing on energetic properties but is less desirable

for a dynamics study. In the event that the frequency of the temperature coupling is some multiple value of the frequency of one of the motions intrinsic to the system, anamalous effects can be introduced.

Furthermore, the temperature is well behaved for a system such as this with no non-bonded cut-off. The average temperature during the 10 ps of constant temperature MD when data was collected (data set #2) was 298.3 K and the standard deviation was 32.4 K. This compares to an average temperature oion of 38.9 K for the constant energy run (data set #1).

Figure 1 depicts one of the endocyclic torsion angles as a function of time and Figure 2 plots the pseudorotation phase angle from the constant energy simulation. Both are measures of conformational conversion and both exhibit the same frequency of about 2.5 cyc/ps. Figure 3 presents the pseudorotation amplitude as a function of time. The average value is calcualted to be 0.46 A with a standard deviation of 0.02 A. This compares well with the experimental value of 0.48 A from Raman spectroscopic studies. The value reported from the MM3 simulation is 0.5 ±0.03. [9]

II. Kinetics

The pseudorotational velocity can be calculated as the change in pseudorotational phase angle with time. The instantaneous quantity from the constant energy simulation is plotted against time in Figure 4. The average pseudorotational velocity is calculated as 910 deg/ps with a standard deviation of 495 deg/ps. Cui et al. [9] reported an RMS pseudorotational velocity, so we will follow suit although we believe the straight pseudorotational velocity to be a more appropriate quantity to calculate and compare with experiment. The RMS value equals 1036 deg/ps, much less than the value obtained with MM3 of 1700 deg/ps. Both the straight and RMS

pseudorotational velocities calculated by AMBER are in better agreement with the estimated value of 400 deg/ps from experiment, [23] which assumes that there is no significant energy barrier associated with pseudorotation.

III. Energetics

The instantaneous potential energy of the system is plotted against time in Figure 5. The average value of about 18 kcal/mol differs from the average value of 25 kcal/mol seen with MM3, however, such molecular mechanical energies are only relevant in terms of differences seen with a particular force field. The components of the potential energy over the course of 1 ps are shown in Figure 6. It is interesting to compare the contributions of the various components with those seen with MM3, where the angle bending contributed ~11-12 kcal/mol, the torsion energy ~8 kcal/mol, van der Waals ~5 kcal/mol, bond stretching ~3 kcal/mol, and the bend-bend, stretch-bend, and torsion-stretch energies were 1 kcal/mol or less. The Weiner *et al.* force field energy components were ~8 kcal/mol for the angle bending, ~6 kcal/mol for the torsion, ~4 kcal/mol for the bond stretching, ~1 kcal/mol for the van der Waals, and essentially 0 kcal/mol for the electrostatic interactions. Thus with both force fields the angle energy dominated, followed by the torsion, but the order of the bond and van der Waals contributions were reversed.

At this point it is worth pointing out one of the major differences between the Weiner *et al.* and MM2/MM3 force fields. The MM2 and MM3 force fields employ bond dipoles to determine the electrostatic interactions for a system, where as the Weiner *et al.* and Cornell *et al.* force fields use atom-centered point charges which have been optimized to reproduce the electrostatic potential surface of a molecule. The MM# simulation of cyclopentane had an electrostatic contribution to the potential energy of absolutely zero because the two different types of bonds present in this system have

bond dipole moments equal to zero. The ESP fit charges used by AMBER, however, did have finite values. The carbon charge was equal to -0.0074 and the hydrogen charge was equal to 0.0037. Because the charges were so small, however, the total electrostatic energy for the system was very nearly equal to zero, as one would expect with a hydrocarbon.

Another aspect of the conformational energetics of cyclopentane which can be investigated is the energy barrier to pass through the planar conformation. This barrier has been determined to be 5.2 kcal/mol by experiment. [24] The Weiner *et al.* force field calculates the barrier at 4.7 kcal/mol compared to the MM3 value of 4.2 kcal/mol. Using our new force field, [11] which has a higher value for the X -CT-CT-X V_3 torsional parameter of 1.4 kcal/mol compared to the previous value of 1.3 kcal/mol, we obtain a barrier of 4.9 kcal/mol. (The new torsional parameter was adjusted to fit the rotational barrier for ethane.)

IV. Cooperative Movement in Pseudorotation

The pseudorotation process requires a tight coupling of the five endocyclic torsion angles due to the cyclic nature of the system. This coupling is illustrated in Figure 7 where each of the first four endocyclic torsion angles of a given conformation is plotted separately against the fifth. Torsions 2 and 3 appear in the ellipse which runs from the second to the fourth quadrant and torsions 1 and 4 appear in the ellipse which runs from the first to the third quadrant. This is consistent with the fact that each pair is symmetrically disposed with respect to torsion 5. Our ellipses are tighter than the ones shown in the MM3 paper, however their data set also includes conformations from the equilibration period, when the coupling was not yet so tight.

The valence angles can also be shown to be coupled to the pseudorotation process. In Figure 8 the values of one of the C-C-C valence angles are plotted against the pseudorotationl phase angle. The data was derived from all of the conformations sampled during the second 10 ps of constant energy MD. The positioning and range of the curve are about the same as seen with MM3, reflecting distortions about the equilibrium C-C-C angle value of 110.0° for the MM3 force field (specific for a five-membered ring) and the generic value of 109.5° for the Weiner *et al.* force field.

The coupling of the C-H bond lengths to the pseudorotation process can also be explored. The MM3 study produced a plot of C-H bond length versus time where the bond length varied about two distinct average values during one cycle of pseudorotation. In the AMBER simulation, however, the bond varied about one average value. The length of a particular (arbitrary) C-H bond is plotted against time in Figure 9. The magnitude of te bond length variation was modulated by a lower frequency motion (Figure 10). In Figure 11 the average C-H bond length for a particular bond is plotted against the pseudorotation phase angle for the entire ensemble of conformations from the second 10 ps of constant energy MD. Again, the bond lengths vary about a single average in contrast to the MM3 data where there was a transition between two average values. Cui *et al.* attributed this bimodal preference to the existence of two different equilibrium values for the C-H bond, depending on whether the hydrogen was in an axial or equatorial position, as supported by experimental evidence. [22,25-26]

The inability of the Weiner *et al.* force field to reproduce this effect has at least two possible causes. The first is the absence of anharmonicity in the representation of the bond energy. A second possibility is the use of the 6-12 function for van der Waals interactions, which is known to overestimate the repulsion at short distances. The

MM3 force field employs a 6-exp function, where the repulsive interaction is represented with an exponential function. The C-H bond length is linger when the hydrogen is in an axial position, due to transannular repulsion of axial hydrogen atoms. It is possible that the 6-12 function is unable to reproduce this fairly subtle effect (the two equilibrium bond lengths differ by 0.005 A in the MM3 study).

V. Comparison of Constant Energy and Constant Temperature Simulations

We chose to carry out our simulation at constant energy in order to avoid perturbing the trajectory of the molecule through the scaling of atomic velocities during a constant temperature simulation. As noted above, the temperature remained very close to the desired value of 298 K during the constant energy simulation, and the standard deviation of the temperature was not much greater than was seen with a constant temperature simulation. In Figure 12 we present the variation in endocyclic torsion angle as a function of time for a constant temperature simulation carried out with a coupling constant of 0.2. Berendsen recommened the use of coupling constants in the range of 0.1-0.4, however, this recommendation was based on simulations carried out on liquid water with a non-bonded cut-off. [20] This cyclopentane trajectory has a pseudorotational velocity of 3.0 cyc/ps, in contrast with the velocity of 2.5 cyc/ps observed during the constant energy simulation. The pseudorotation amplitude exhibited an average value of 0.47 ±0.02 A as compared to the value of 0.46 ±0.02 A observed during the constant energy run. The variation of C-H bond length with time is plotted in Figures 13 and 14 for this constant temperature simulation. The variation is more uneven than was seen in Figures 9 and 10.

A second constant temperature simulation was run with a coupling constant of 0.5 for the constant temperature bath. The velocity of pseudorotation is not reduced by much

with the looser coupling constant, having a value of about 2.9 cyc/ps. Since the average temperature of he constant energy run was actually 293 K rather than 298, we ran a third constant temperature simulation with a coupling constant of 0.2 and a target temperature of 293 K. Under these conditions, the pseudorotation velocities seen with the constant temperature and constant energy simulations are due to the constant temperature protocol.

We tested this hypothesis further by carrying out three additional constant temperature runs with successively looser coupking constants. The fourth constant temperature simulation was carried out with a target temperature of 298 K and a coupling constant of 0.9 This simulation exhibited a pseudorotation velocity of 2.9 cyc/ps as well. The fifth and sixth constant temperature runs were carried out with target temperatures of 298 K and coupling constants of 2.0 and 20.0, respectively. The pseudorotation velocity calculated for these simulations was 2.7 cyc/ps, reflecting a return to conditions which are similar to constant energy, due to the weak coupling to the bath.

It is therefore clear that carrying out MD simulations of cyclopentane under conditions of constant temperature can affect the calculated pseudorotation velocity by up to 20% when the coupling constant is within the range recommended by Berendsen. [20] Unfortunately, Cui et al. [9] did not address this issue or report the coupling constant which they employed, and it is possible that the higher pseudorotation velocity which they observed resulted from the constant temperature protocol.

Conclusion

We have carried out MD simulations of the cyclopentane molecule under conditions of both constant energy and constant temperature using the Weiner et al. [1] force field and compared the results to those reported by Cui et al. [9] using MM3. The pseudorotation process was found to be well reproduced using the Weiner et al. [1] force field in terms of the velocity and puckering amplitude. In addition, the varrier to passing through the planar conformation was calculated at 4.7 kcal/mol as compared to the experimental value of 5.2 kcal/mol and the value of 4.2 kcal/mol calculated with the MM3 force field. The Weiner et al. [1] force field was unable, however, to reproduce the subtle differences in C-H bond length which exist for axial and equatorial positions. We found that carrying out the simulation under conditions of constant temperature, even with a fairly loose coupling constant, increased the observed pseudorotation velocity. We conclude that the Weiner et al. [1] force field, although simpler in form than the MM3 force field, performs quite well at reproducing the conformational dynamics and energetics of this small molecule in the gas phase.

Acknowledgements

References

(1) (a) Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. (b) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comput. Chem.* **1986**, *7*, 230.

(2) (a) Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551. (b) Allinger, N.L. *J. Am. Chem. Soc.* **1977**, *99*, 8127. (c) Allinger, N.L.; Burkert, U. *Molecular Mechanics*, American Chemical Society Monograph 177, American Chemical Society: Washington, D.C.; 1982.

(3) Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(4) Jorgensen, W.L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(5) Dauber-Osguthorpe, P.; Roberts, V.A.; Osguthorpe, D.; Wolff, J.; Genest, M.; Hagler, A.T. *Proteins* **1988**, *4*, 31.

(6) Lii, J.-H.; Allinger, N.L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.

(7) MM2(87), QCPE, Bloomington, IN.

(8) MM3-MD User Manual, Chemistry Department, University of Georgia.

(9) Cui, W.; Li, F.; Allinger, N.L. *J. Am. Chem. Soc.* **1993**, *115*, 2943.

(10) Lii, J.-H.; Allinger, N.L. *J. Comput. Chem.* **1991**, *12*, 186.

(11) Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M., Jr.; Spellmeyer, D.C.; Ferguson, D.M.; Fox, T.; Caldwell, J.W.; Kollman, P.A., submitted.

(12) Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. *J. Phys. Chem.* **1993**, *97*, 10269.

(13) Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Kollman, P.A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

Cieplak, P.; Cornell, W.D.; Bayly, C.I.; Kollman, P.A. *J. Comput. Chem.*, accepted.

(14) Howard, A.E.; Cieplak, P.; Kollman, P.A. *J. Comput. Chem.*, in press.

(15) (a) Singh, U.C.; Kollman, P.A. *J. Comput. Chem.* **1984**, *5*, 129. (b) Cox, S.; Williams, D. *J. Comput. Chem.* **1981**, *2*, 304.

(16) Hall, D.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 441.

(17) (a) Roterman, I.K.; Lambert, M.H.; Gibson, K.D.; Scheraga, H.A. *J. Biomol. Struct. Dynam.*. **1989**, *7*, 421. (b) Kollman, P.A.; Dill, K.A. *Ibid.* **1991**, *8*, 1103.

(18) Kilpatrick, J.E.; Pitzer, K.S.; Spitzer, R. *J. Am. Chem. Soc.* **1947**, *69*, 2483.

(19) Pearlman, D.A.; Case; D.A.; Caldwell, J.W.; Seibel, G.L.; Singh, U.C.; Weiner, P.A.; Kollman, P.A. AMBER 4.0 (UCSF), 1991, Department of Pharmaceutical Chemistry, University of California, San Francisco.

(20) Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.; DiNola, A.; Haak, J.R. *J. Chem. Phys.* **1984**, *81*, 3684.

(21) (a) Altona, C.; Sundaralingham, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205. (b) Rao, S.T.; Westhof, E.; Sundaralingham, M. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1981**, *37*, 421.

(22) Bauman, L.E.; Laane, J. *J. Phys. Chem.* **1988**, *92*, 1040.

(23) Henry, P.R.; Hung, I.F.; MacPhail, R.A.; Strauss, H.R. *J. Am. Chem. Soc.* **1980**, *102*, 515.

(24) Carreira, L.A.; Jiang, G.J.; Person, W.B.; Willis, J.N. *J. Chem. Phys.* **1972**, *56*, 1440.

(25) Variyar, J.E.; MacPhail, R.A. *J. Phys. Chem.* **1992**, *96*, 576.

(26) MacPhail, R.A.; Variyar, J.E. *Chem. Phys. Lett.* **1989**, *161*, 239.

Figure Captions

Figure 1. The value of a particular (arbitrary) torsion angle is plotted against time for the constant energy simulation. One cycle of rotation for the torsion is equivalent to one complete cycle of pseudorotation.

Figure 2. The pseudorotation phase angle, P, is plotted agains time for the constant energy simulation. The phase angle specifies the position of maximum puckering in the ring during the pseudorotation process.

Figure 3. The pseudorotation amplitude, q, is plotted against time for the constant energy simulation. This parameter specifies the magnitude of the ring puckering seen at each point in the pseudorotation process. The average value is 0.46 ±0.02 A.

Figure 4. The instantaneous pseudorotation velocity is plotted against time for the constant energy simulation. The average velocity of 910 ±495 deg/ps. The instances when the velocity is negative correspond to the teeth seen in the curves in Figures 1 and 2. At those times the pseudorotation of the molecule is momentarily reversed.

Figure 5. The potential energy is plotted against time for the constant energy simulation. The energy varies about an average value, indicating that the system is equilibrated.

Figure 6. Comparison of the potential energy for the constant energy simulation plotted over 1 ps. The angle and torsion components are the first and second largest, as was seen in the MM3 simulation. The third largest component is from the bond energy, in contrast to the MM3 simulation where was the van der Waals energy.

Figure 7. Endocyclic torsion angles 1-4 (assigned arbitrarily but in order around the ring) are plotted against the fifth endocyclic torsion angle. Torsion 1 = solid circle. Torsion 2 = solid square. Torsion 3 = open square. Torsion 4 = open circle. The high degree of correlation demonstrates the concerted nature of the pseudorotation process.

Figure 8. An arbitrary C-C-C valence bond angle is plotted against the pseudorotation phase angle, P, with which it is associated. As a given pseudorotation phase angle occurs many times during the simulation, the graph is not single-valued. Although each pseudorotation phase angle has corresponding valence angles which vary by 5°, the average values are clearly periodic, following the pseudorotation.

Figure 9. An arbitrary C-H bond length is plotted against time for the constant energy simulation. The bond stretching motion is clearly modulated by a lower frequency motion, thus demonstrating the coupling between the bond stretching and the pseudorotation process.

Figure 10. The C-H bond stretching shown in Figure 9 is displayed in greater detail for a smaller period of time.

Figure 11. The C-H bond length is plotted against the pseudorotation phase angle for the constant energy simulation. The coupling between the C-H bond stretching and the pseudorotation process is not readily evident in this representation.

Figure 12. An arbitrary endocyclic torsion angle is plotted against time for the constant temperature simulation with a target emperature of 298 K and a coupling

constant of 0.2. The pseudorotation velocity is 30 cyc/ps as compared to the velocity of 25 cyc/ps seen in the constant energy simulation.

Figure 13. An arbitrary C-H bond length is plotted against time for the constant temperature simulation with a target temperature of 298 K and a coupling constant of 0.2 The trajectory appears to be more distorted than the one from the constant energy simulation plotted in Figure 9.

Figure 14. The C-H bond stretching shown in Figure 13 is displayed in greater detail for a smaller period of time. The distortion is greater than that seen in the constant energy plot in Figure 10.
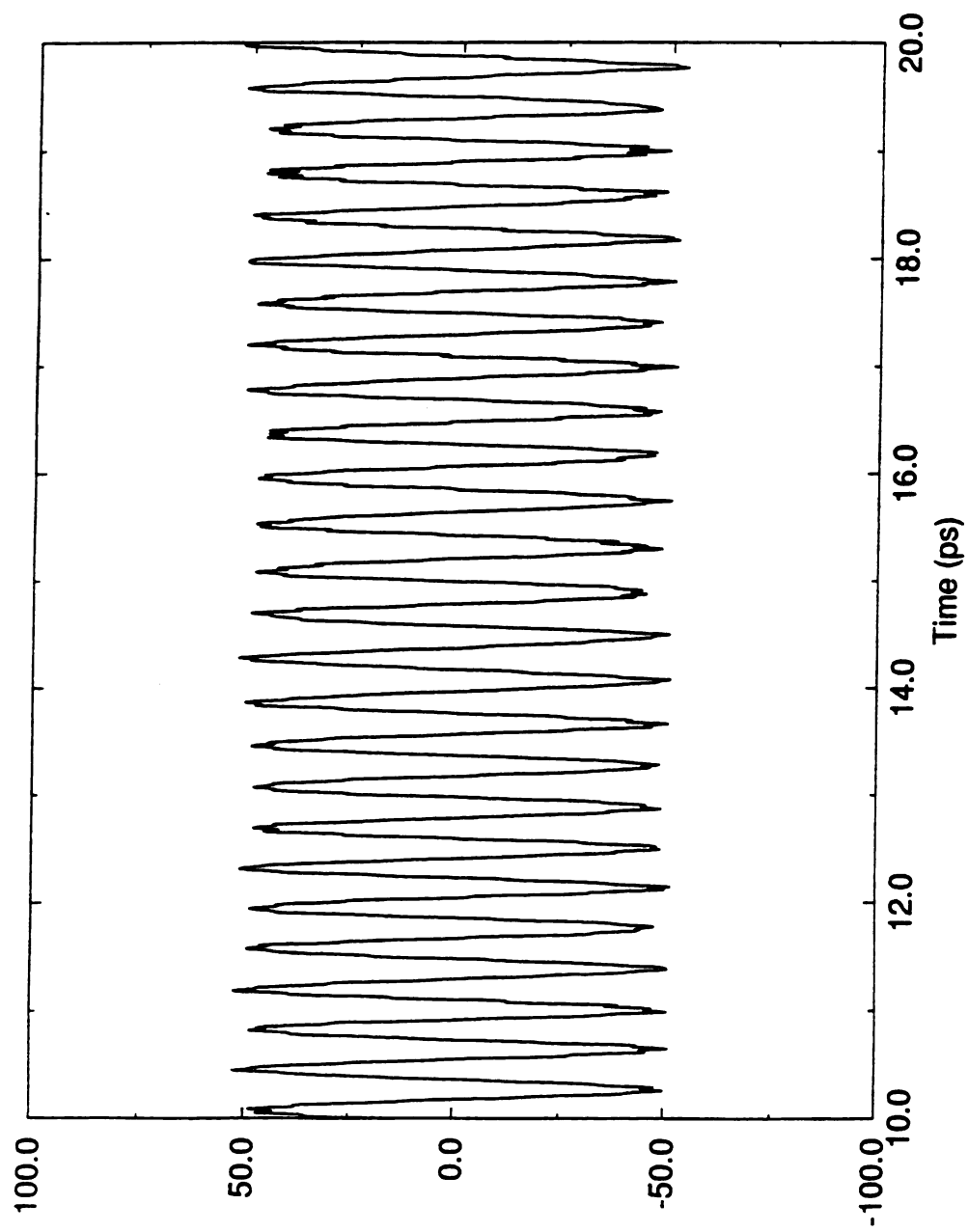
Figure 1

Endocyclic Torsion Angle (deg) at Constant Energy

Figure 2

Pseudorotation Phase Angle (degree)

Time (ps)
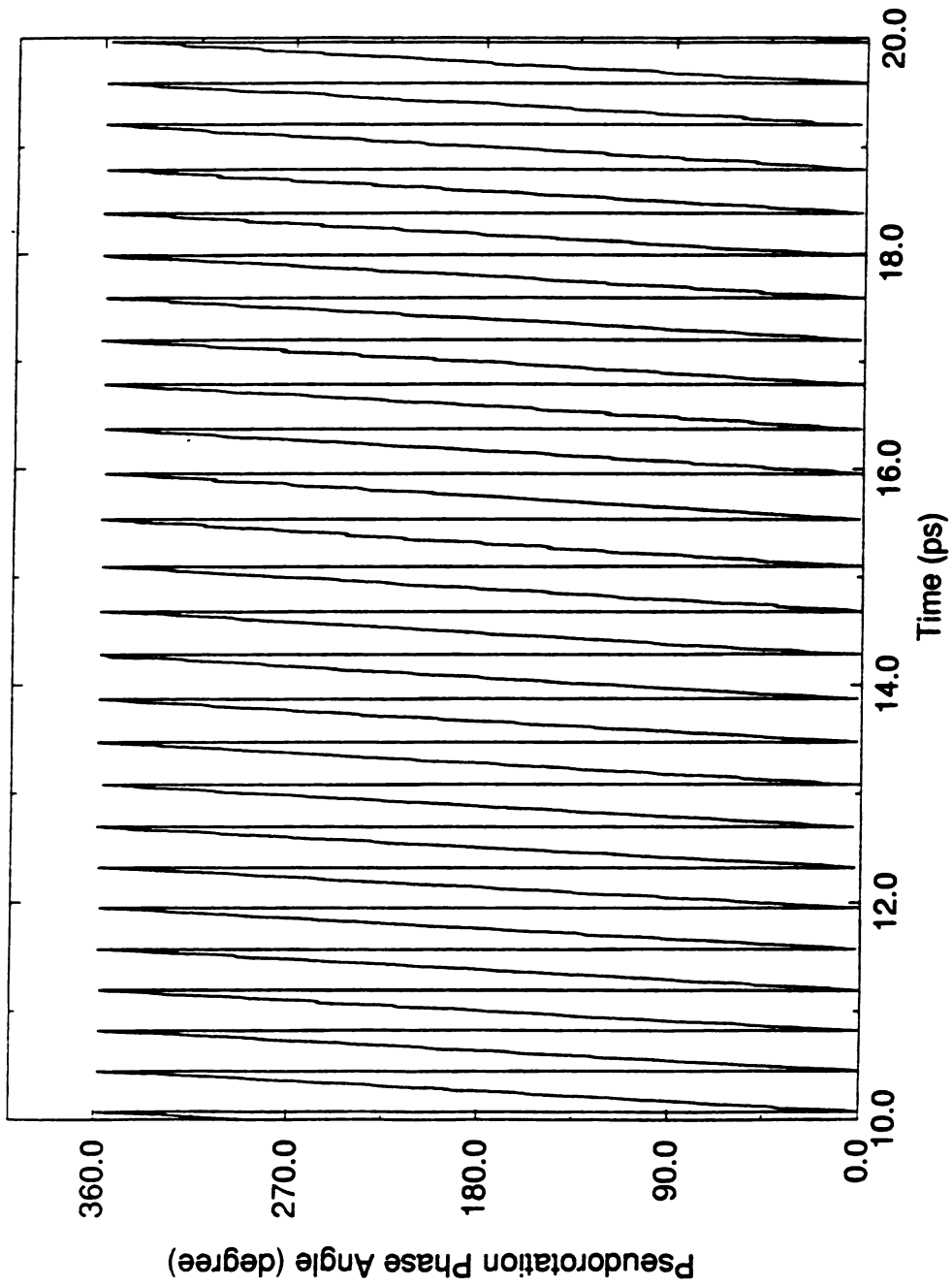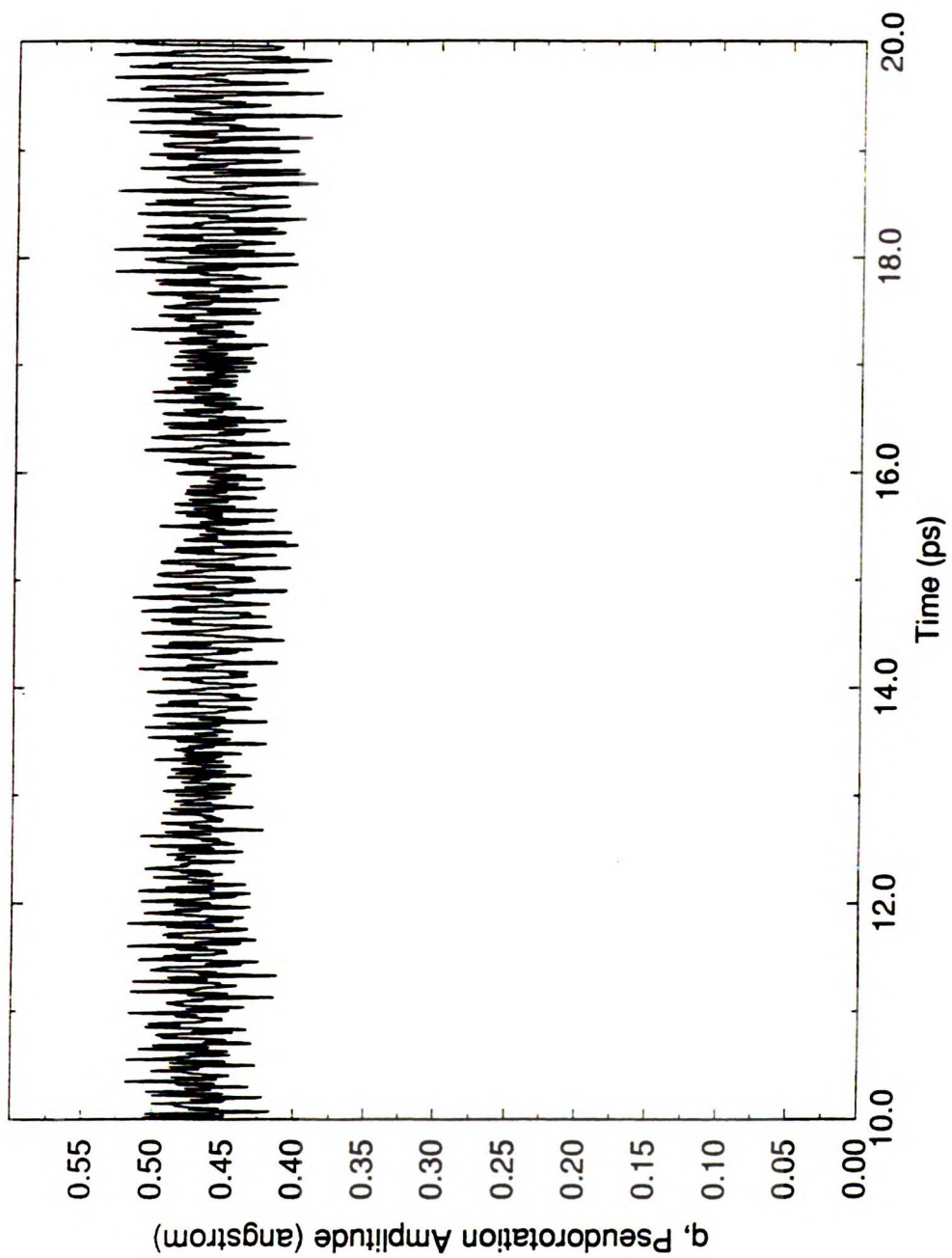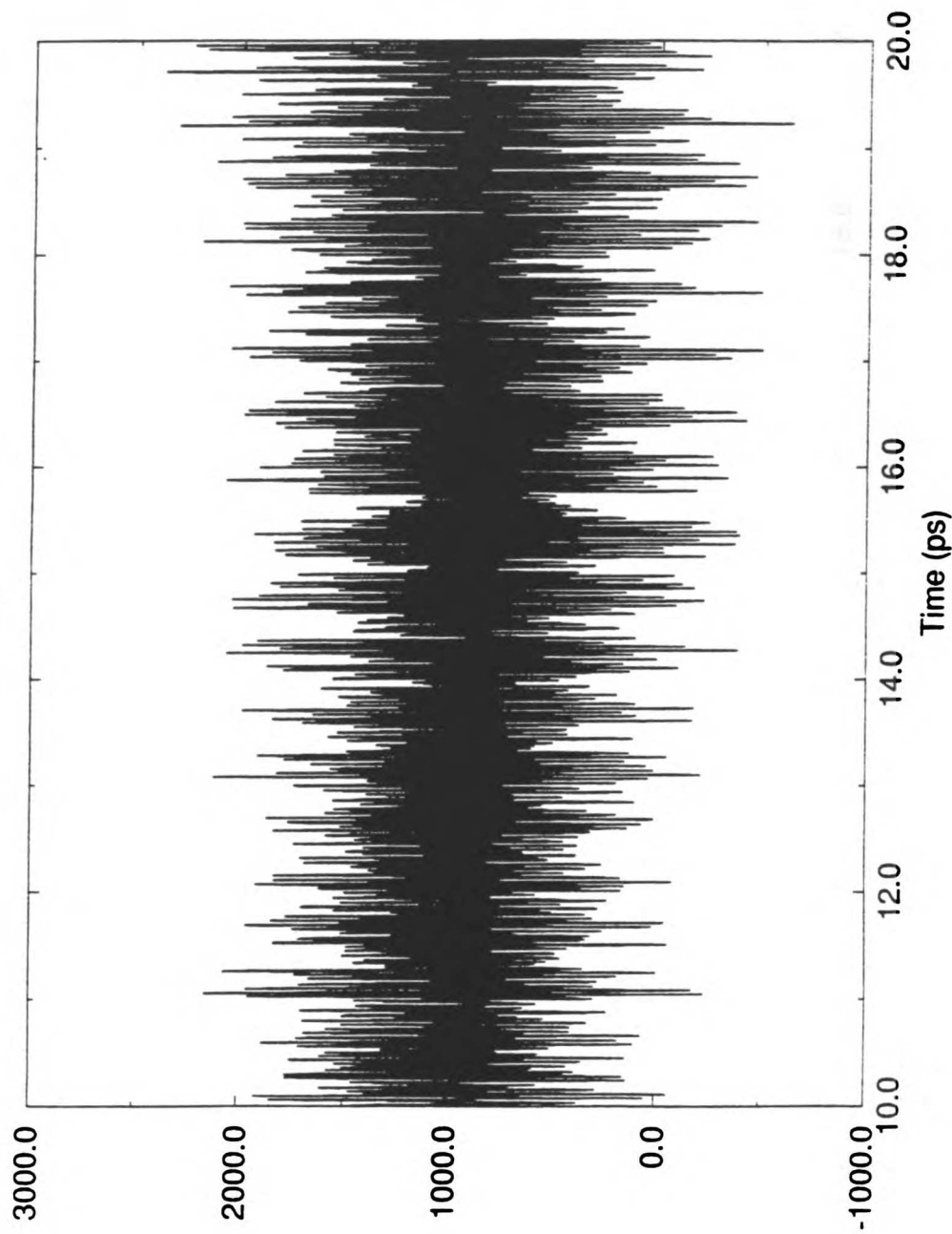
Figure 3

Figure 4

Pseudorotation Velocity at Constant Energy (degrees/ps)
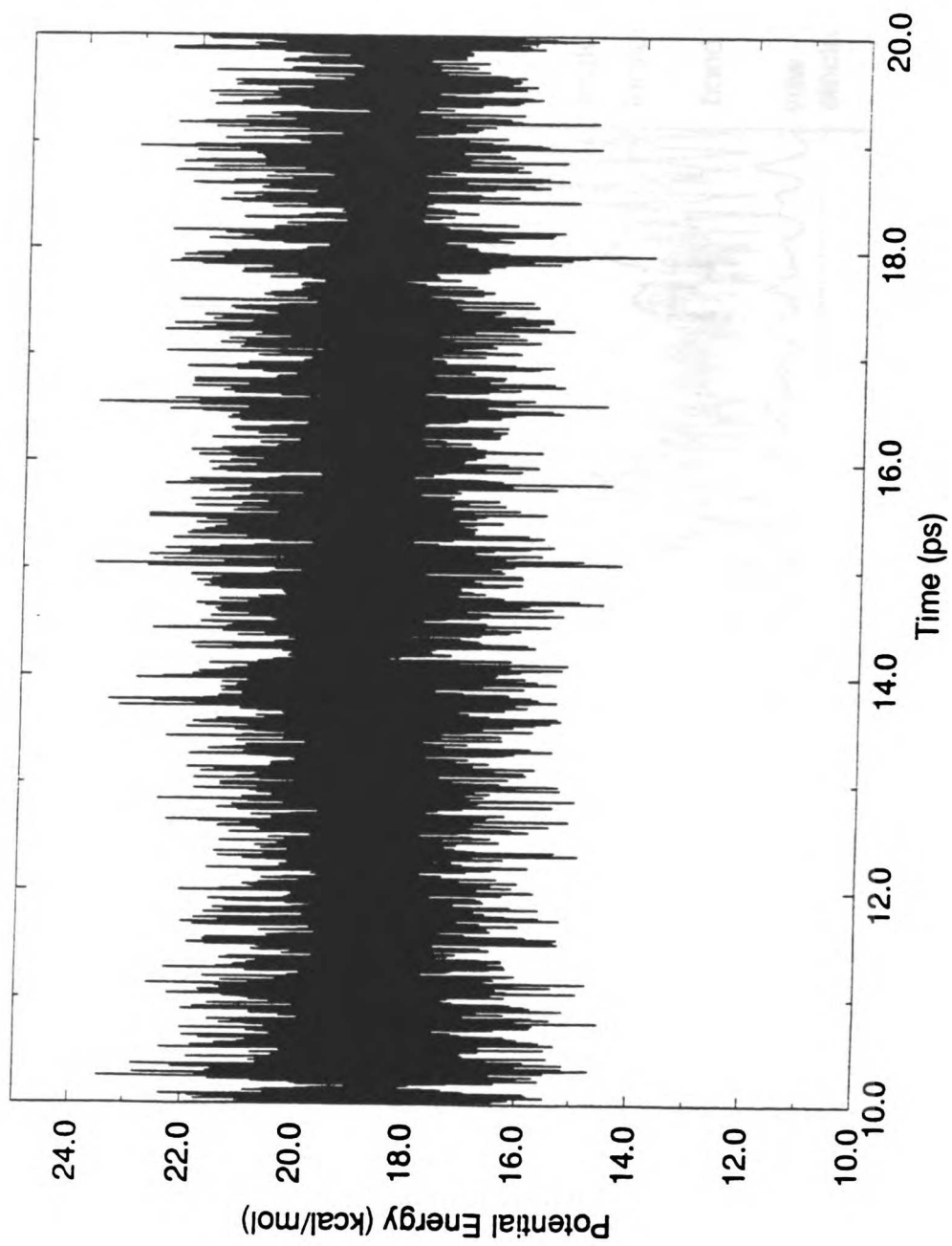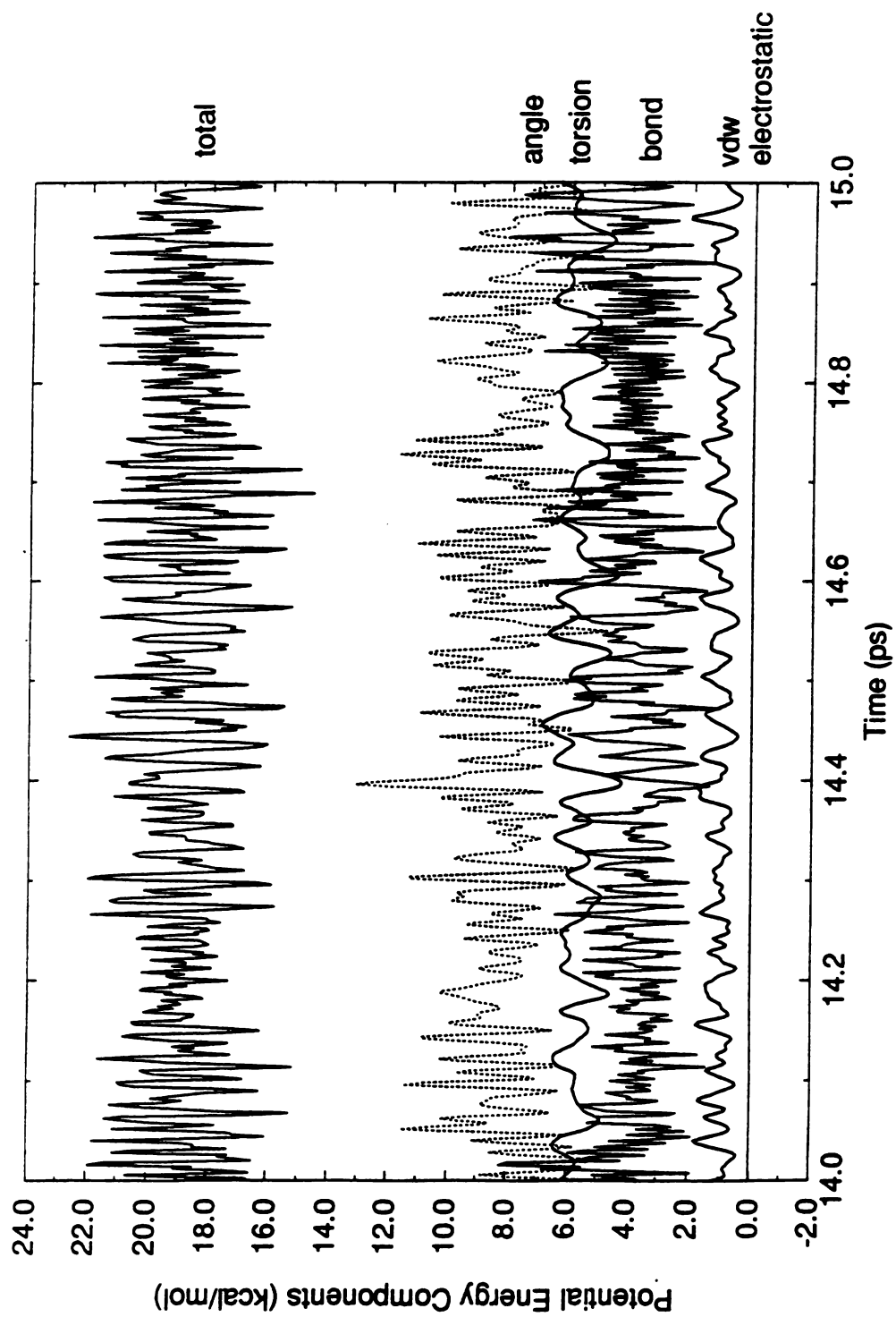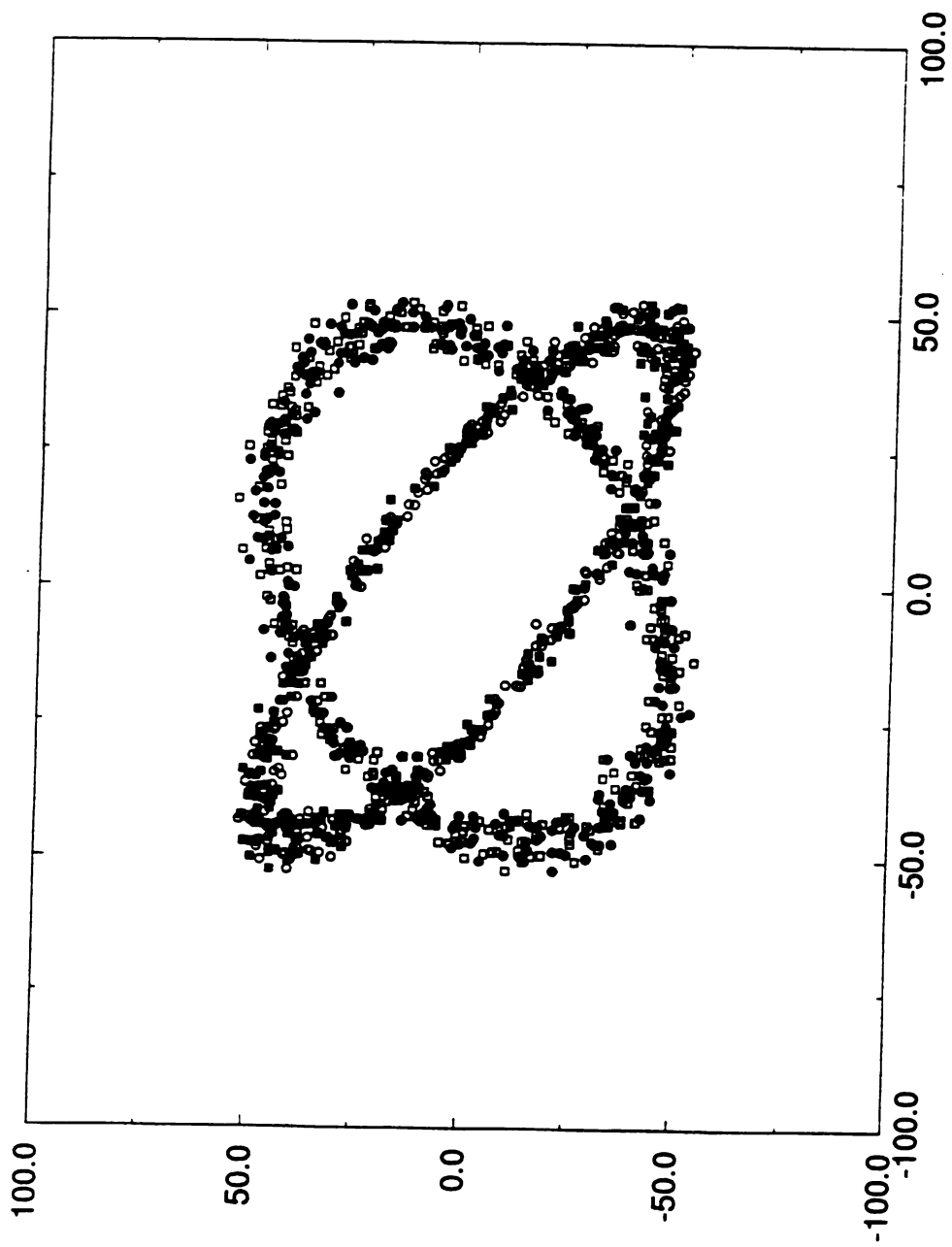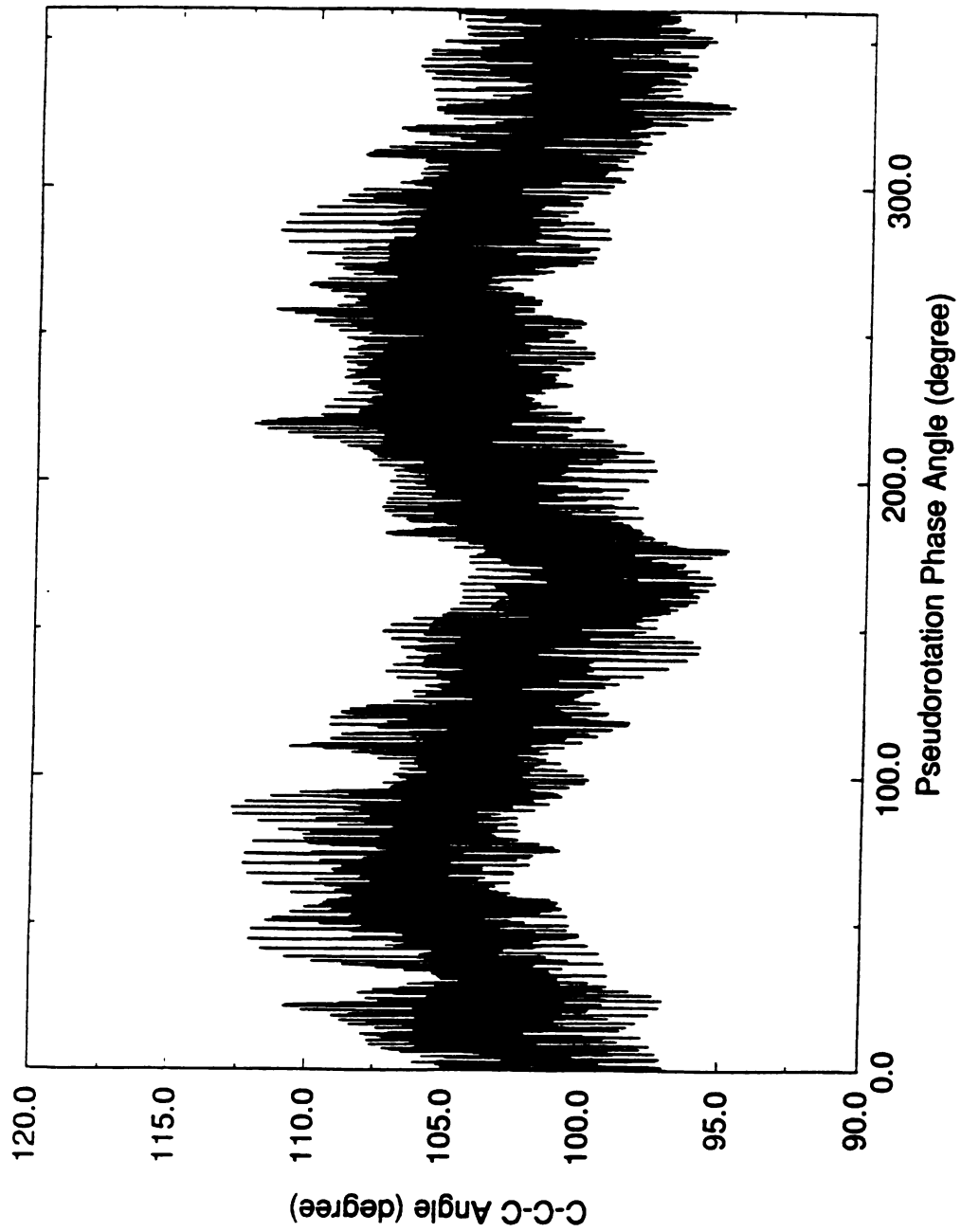
Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

C-H Bond (angstrom) at Constant Energy

Time (ps)

Figure 10

Figure 11

Endocyclic Torsion Angle (deg) at Constant Temp. (T=298,tau=0.2)
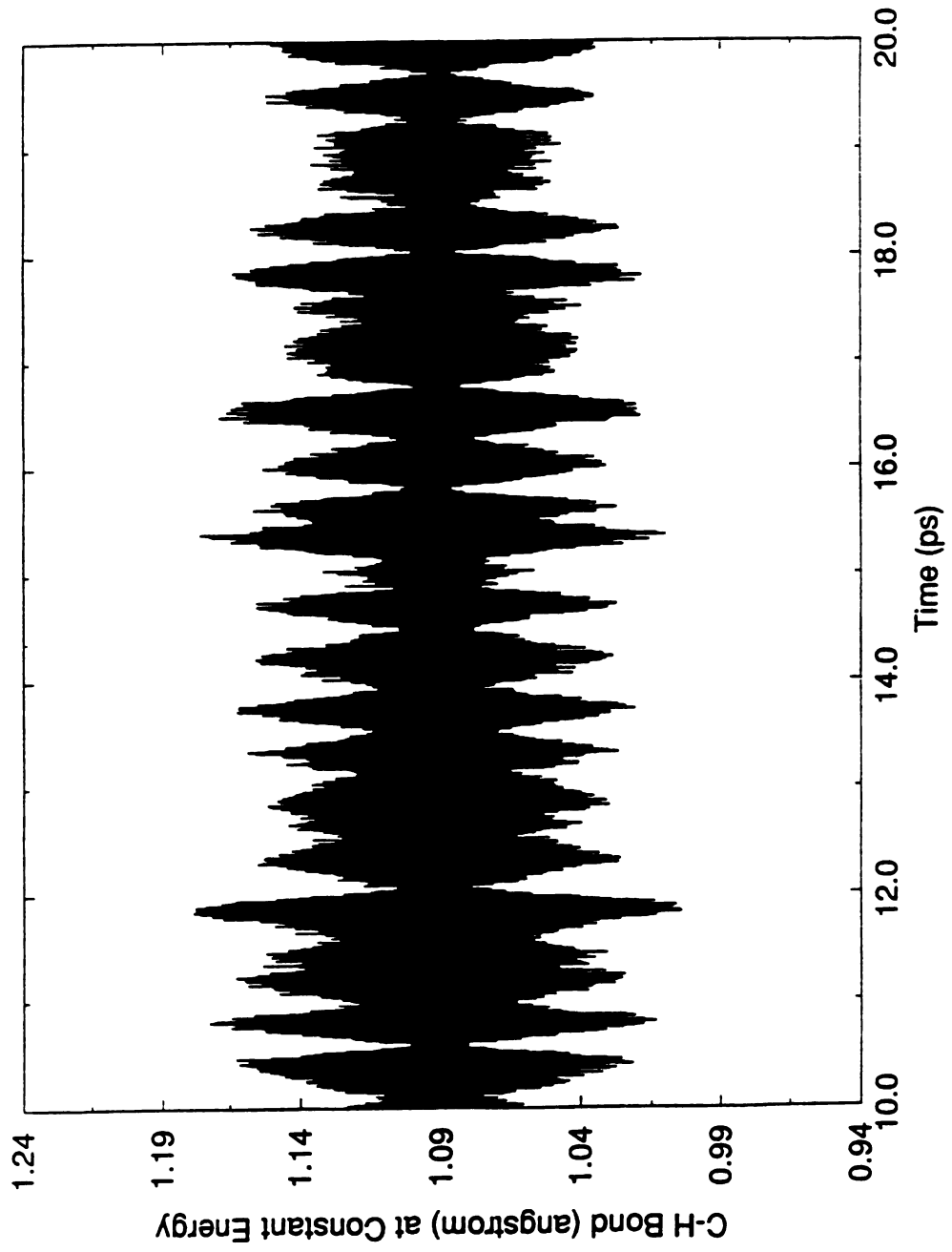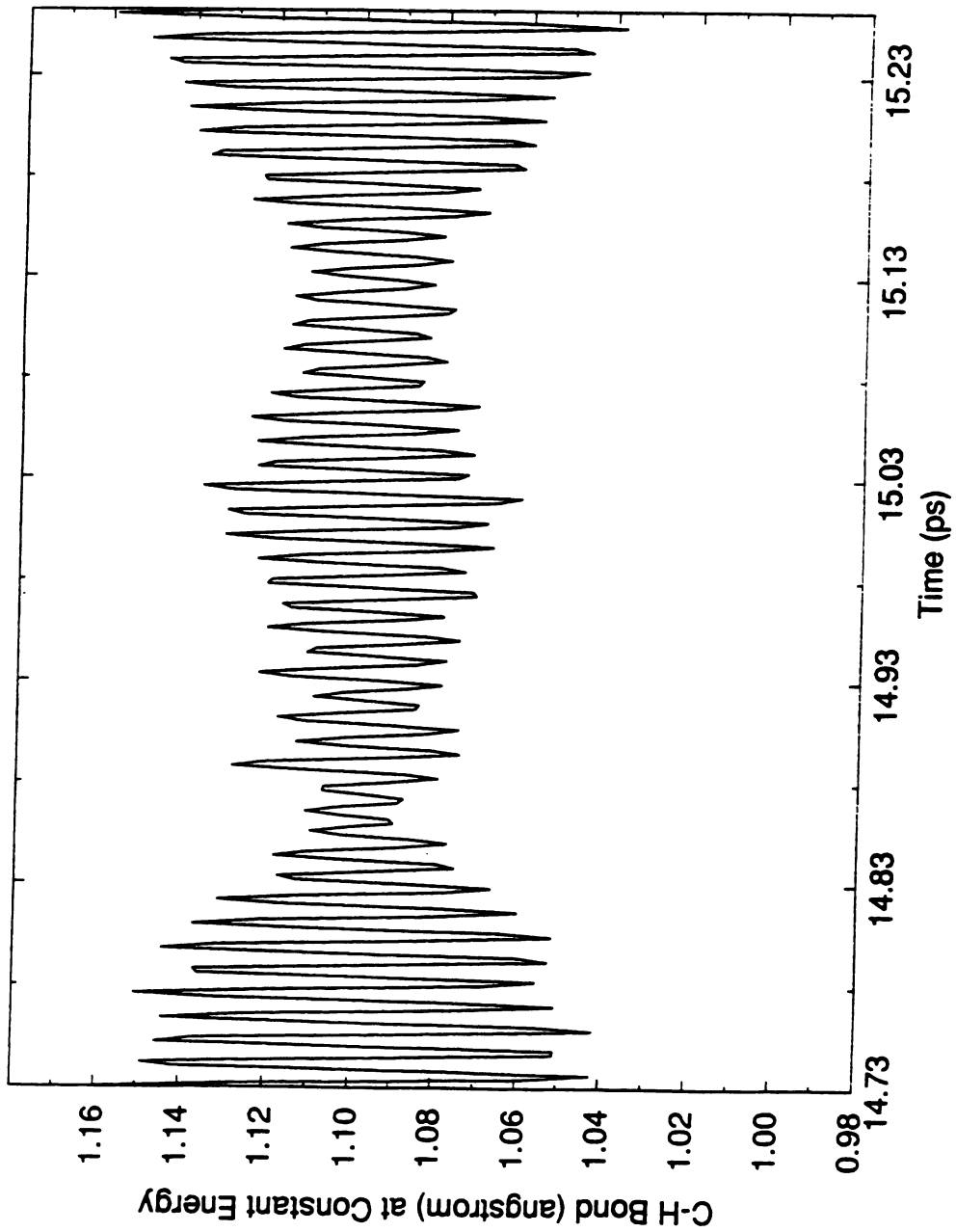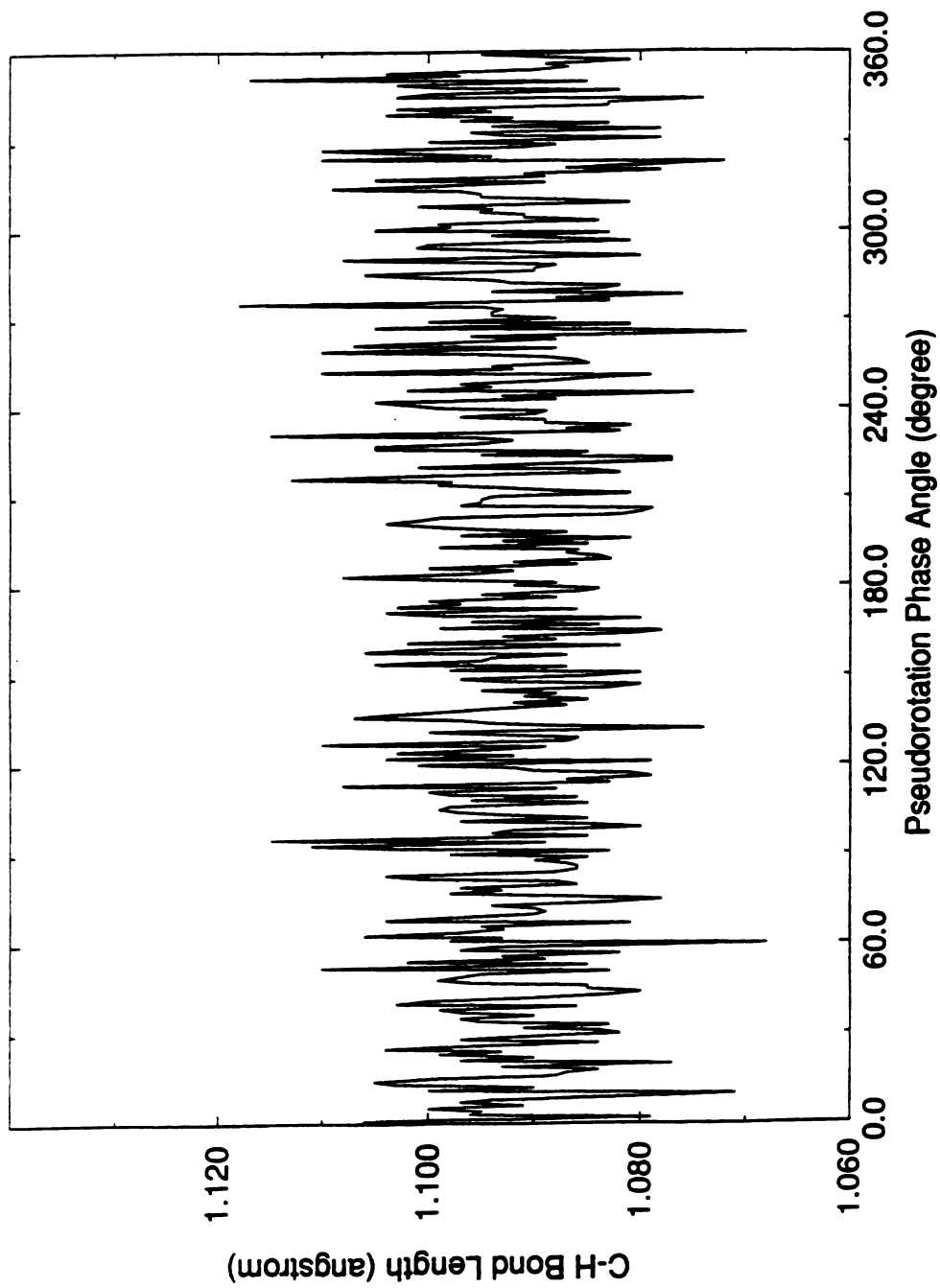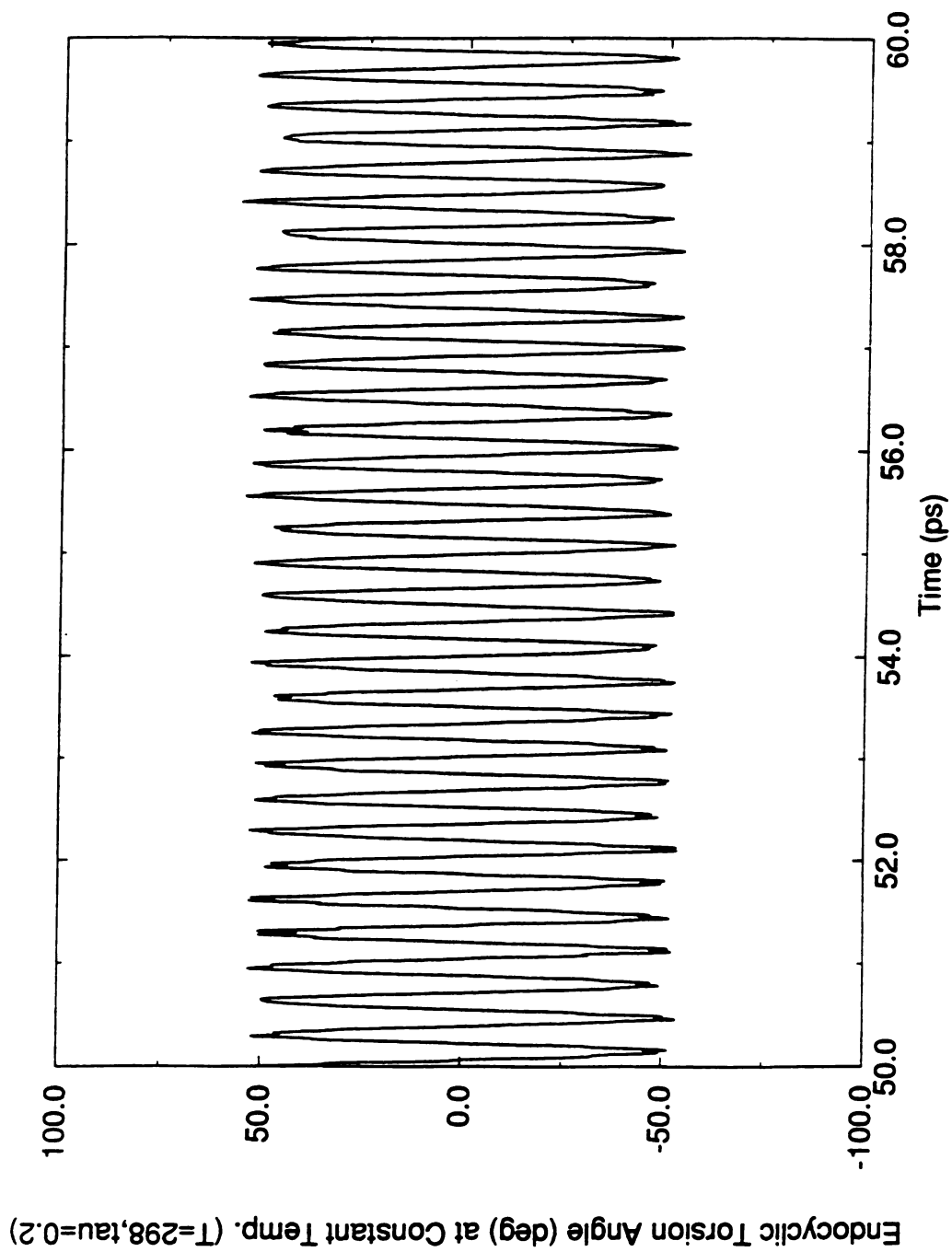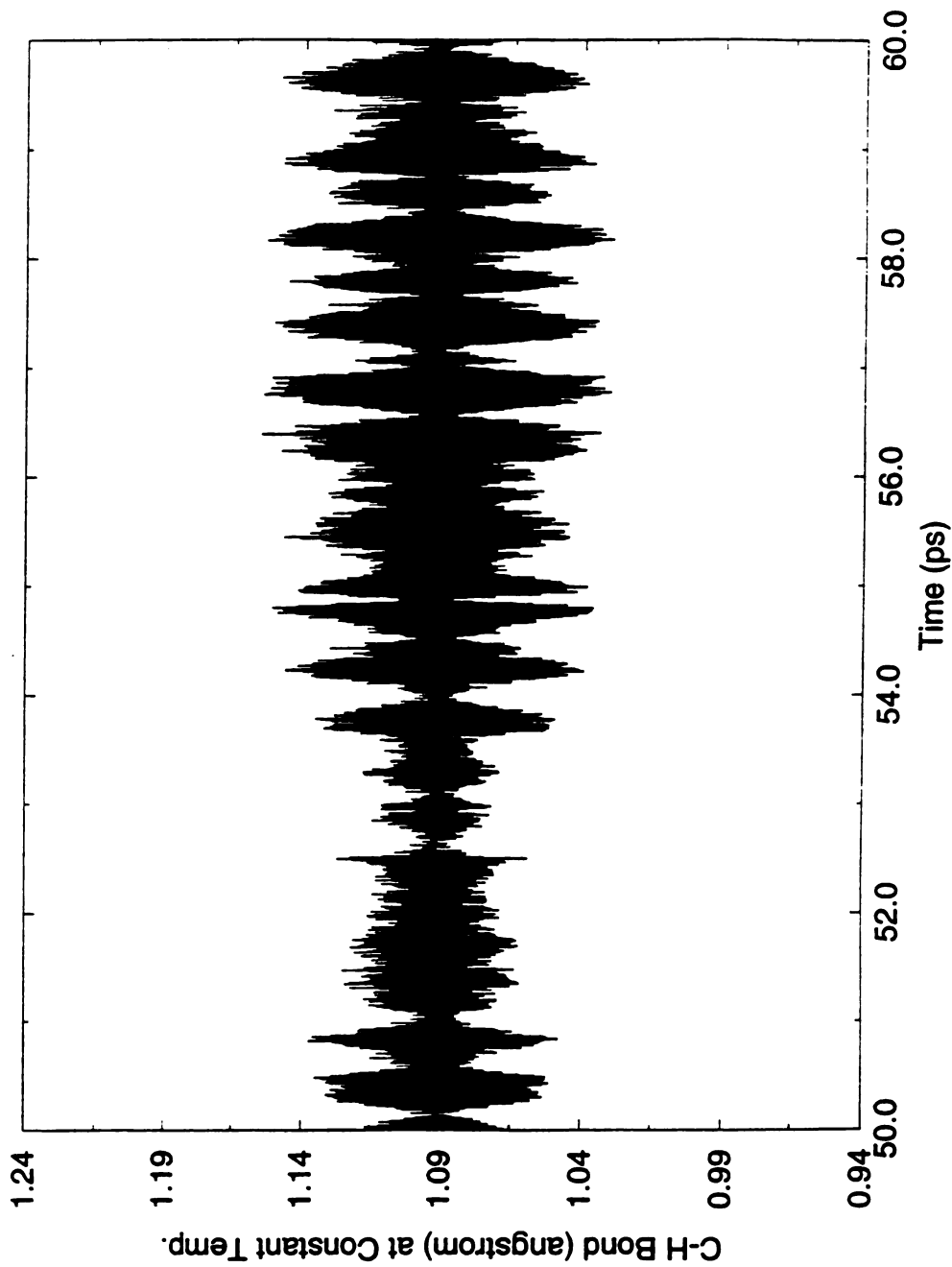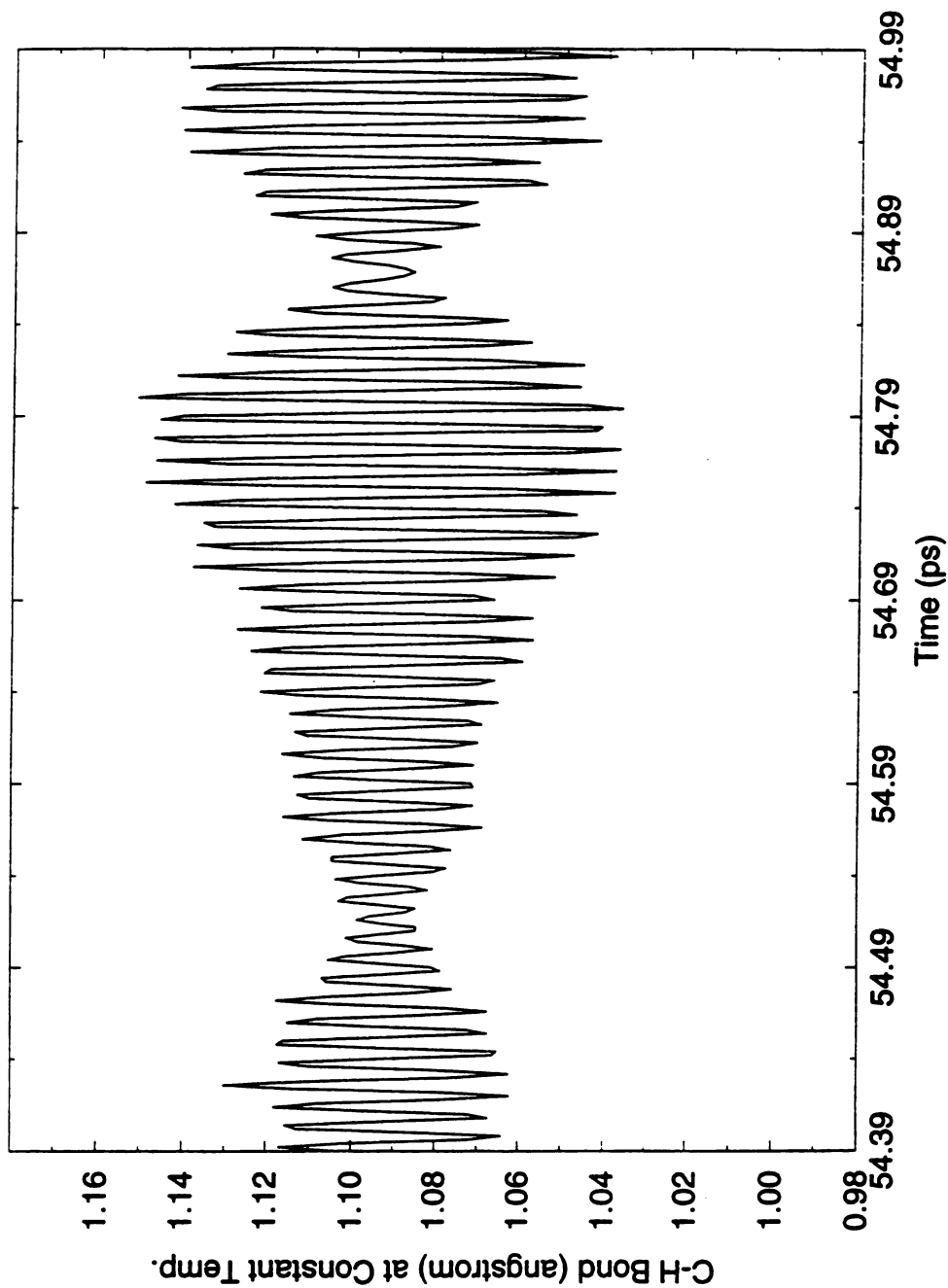
Figure 12

Figure 13

C-H Bond (angstrom) at Constant Temp.

Time (ps)

Figure 14