11-9-2011

Theoretical Study of Chloroperoxidase Catalyzed Chlorination of beta-Cyclopentanedione and Role of Water in the Chlorination Mechanism

Cassian D'Cunha
Florida International University, cassian.dcunha@gmail.com

Follow this and additional works at: http://digitalcommons.fiu.edu/etd

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THEORETICAL STUDY OF CHLOROPEROXIDASE CATALYZED

CHLORINATION OF BETA-CYCLOPENTANEDIONE AND ROLE OF WATER IN

THE CHLORINATION MECHANISM

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Cassian D'Cunha

2011

To: Dean Kenneth Furton
     College of Arts and Sciences

This dissertation, written by Cassian D'Cunha, and entitled Theoretical Study of Chloroperoxidase Catalyzed Chlorination of beta-Cyclopentanedione and Role of Water in the Chlorination Mechanism, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Alexander Mebel

_____
Xiaotang Wang

_____
Bernard Gerstman

_____
Kenneth Furton

_____
David Chatfield, Major Professor

Date of Defense: November 9, 2011

The dissertation of Cassian D'Cunha is approved.

_____
Dean Kenneth Furton
College of Arts and Sciences

_____
Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2011

DEDICATION

I dedicate this dissertation to my parents. To my mother, my role model, for her constant

love and support, and in loving memory of my father.

ACKNOWLEDGMENTS

I would first like to thank my committee members. Their help, encouragement and patience during the various stages of my dissertation have been most appreciated. I would like to thank Dr. Xiaotang Wang for being particularly helpful during the design phase of my dissertation. I would like to thank Dean Kenneth Furton for taking time from his extremely busy schedule to serve on my committee. I would like to thank Dr. Bernard Gerstman for taking an interest in my research and inviting me to present my research to his research group on several occasions. I would like to thank Dr. Alexander Mebel for being supportive and patient, and for answering the many questions I had regarding quantum mechanics approaches. I would especially like to thank Dr. David Chatfield, my major professor. He has been very supportive, encouraging and patient, and I sincerely appreciate the time I spent under his guidance. I cannot thank him enough for the time he took to discuss research, guide me through my dissertation and make resources available for my research. His mentorship was invaluable and much appreciated.

My experience in the graduate program in Department of Chemistry & Biochemistry at Florida International University has been very rewarding. The professors in the department have been very approachable, accommodating and always willing to help. I would also like to thank my peers in Dr. Mebel's and Dr. Chatfield's research groups. The invaluable discussions during our group meetings and the discussions in the research laboratory have been invaluable and enriched my learning experience.

I would like to thank my friends and roommates over the years. They have been my family away from home while I was pursuing my degree. I would also like to thank the Martinez family that has all but adopted me and made me feel welcome in their home.

ABSTRACT OF THE DISSERTATION

THEORETICAL STUDY OF CHLOROPEROXIDASE CATALYZED

CHLORINATION OF BETA-CYCLOPENTANEDIONE AND ROLE OF WATER IN

THE CHLORINATION MECHANISM

by

Cassian D'Cunha

Florida International University, 2011

Miami, Florida

Professor David Chatfield, Major Professor

Chloroperoxidase (CPO) is a potential biocatalyst for use in asymmetric synthesis. The mechanisms of CPO catalysis are therefore of interest. The halogenation reaction, one of several chemical reactions that CPO catalyzes, is not fully understood and is the subject of this dissertation.

The mechanism by which CPO catalyzes halogenation is disputed. It has been postulated that halogenation of substrates occurs at the active site. Alternatively, it has been proposed that hypochlorous acid, produced at the active site via oxidation of chloride, is released prior to reaction, so that halogenation occurs in solution. The free-solution mechanism is supported by the observation that halogenation of most substrates often occurs non-stereospecifically. On the other hand, the enzyme-bound mechanism is supported by the observation that some large substrates undergo halogenation stereospecifically.

The major purpose of this research is to compare chlorination of the substrate β-cyclopentanedione in the two environments. One study was of the reaction with limited

hydration because such a level of hydration is typical of the active site. For this work, a purely quantum mechanical approach was used. To model the aqueous environment, the limited hydration environment approach is not appropriate. Instead, reaction precursor conformations were obtained from a solvated molecular dynamics simulation, and reaction of potentially reactive molecular encounters was modeled with a hybrid quantum mechanical/molecular mechanical approach. Extensive work developing parameters for small molecules was pre-requisite for the molecular dynamics simulation.

It is observed that a limited and optimized (active-site-like) hydration environment leads to a lower energetic barrier than the fully solvated model representative of the aqueous environment at room temperature, suggesting that the stable water network near the active site is likely to facilitate the chlorination mechanism.

The influence of the solvent environment on the reaction barrier is critical. It is observed that stabilization of the catalytic water by other solvent molecules lowers the barrier for keto-enol tautomerization. Placement of water molecules is more important than the number of water molecules in such studies. The fully-solvated model demonstrates that reaction proceeds when the instantaneous dynamical water environment is close to optimal for stabilizing the transition state.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ACRONYM	FULL NAME
ABNR	Adopted Basis Newton Raphson
ACM	Adiabatic Connection Method
AGLC	Alpha-D-Glucose
ANEB	Adaptive Nudged Elastic Band
AXYS	Alpha-D-Xylopyranos
BXYS	Beta-L-Xylopyranos
BARB	Beta-L-Arabinose
BNAG	N-Acetyl-Glucosamine
CC	Coupled Cluster
CI-NEB	Climbing Image-Nudged Elastic Band
CPO	Chloroperoxidase
CPD	1,3 Cyclopentanedione
DFT	Density Functional Theory
DHP	Dehaloperoxidase
ESP	Electrostatic Potential
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbitals
GRMS	Root Mean Square Gradient
HF	Hartree-Fock
IC	Internal Coordinate
IRC	Intrinsic Reaction Coordinate

LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
MCD	Monochlorodimedone
MD	Molecular Dynamics
MEP	Minimum Energy Path
MK	Merz-Kollman
MM	Molecular Mechanics
MO	Molecular Orbitals
MP	Moller-Plesset
NEB	Nudged Elastic Band
NMR	Nuclear Magnetic Resonance
PCA	Pyroglutamic Acid
PCM	Polarizable Continuum Model
PES	Potential Energy Surface
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
RMSD	Root Mean Square Deviation
RPATH	Replica Path
SD	Steepest Descent
STO	Slater Type Orbitals
TS	Transition State
VDW	van der Waals

1. INTRODUCTION

1.1. Chloroperoxidase

Shaw and Hager [1] reported chlorination activity when extracts of a marine fungus, *Caldariomyces fumago*, were used to react with β-ketoadipic acid to give δ-chlorolevulinic acid. Another observation made was that the main metabolite of this organism was a chlorinated species, 2,2-dichloro-1,3-cyclopentanediol, clearly indicating that there was an enzyme or group of enzymes in the organism that facilitated chlorination. It was later discovered by the same group that there were at least four enzymes in the mold extract. The four enzymes identified were an amyloltic enzyme, a catalase, glucose oxidase and another enzyme that they termed Chloroperoxidase (CPO) to denote the activity of the enzyme [2].
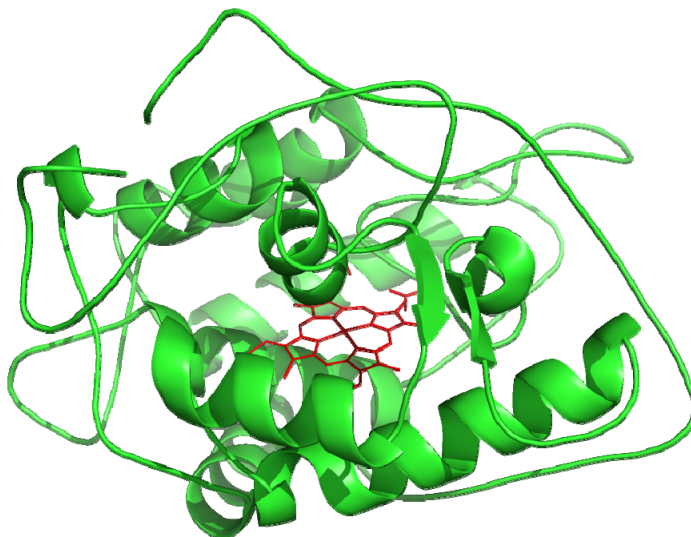
Figure 1: Chloroperoxidase (CPO)

1

Chloroperoxidase is a heme containing metalloprotein enzyme. It shares this feature with several other globular enzymes like myoglobin, haemoglobin, cytochrome P-450 and horseradish peroxidase. One of the main characteristics distinguishing CPO from most heme enzymes is that the proximal ligand bound to the heme is a cysteine residue [3]. It shares this feature with cytochrome P-450, but a main distinguishing feature between the two is that the distal site of the porphyrin group is charged or polar in CPO and hydrophobic in cytochrome P-450.

Since being discovered more than 50 years ago, CPO has been studied extensively because of its versatility, ecological significance and potential applicability in the pharmaceutical industry [4].

[5]
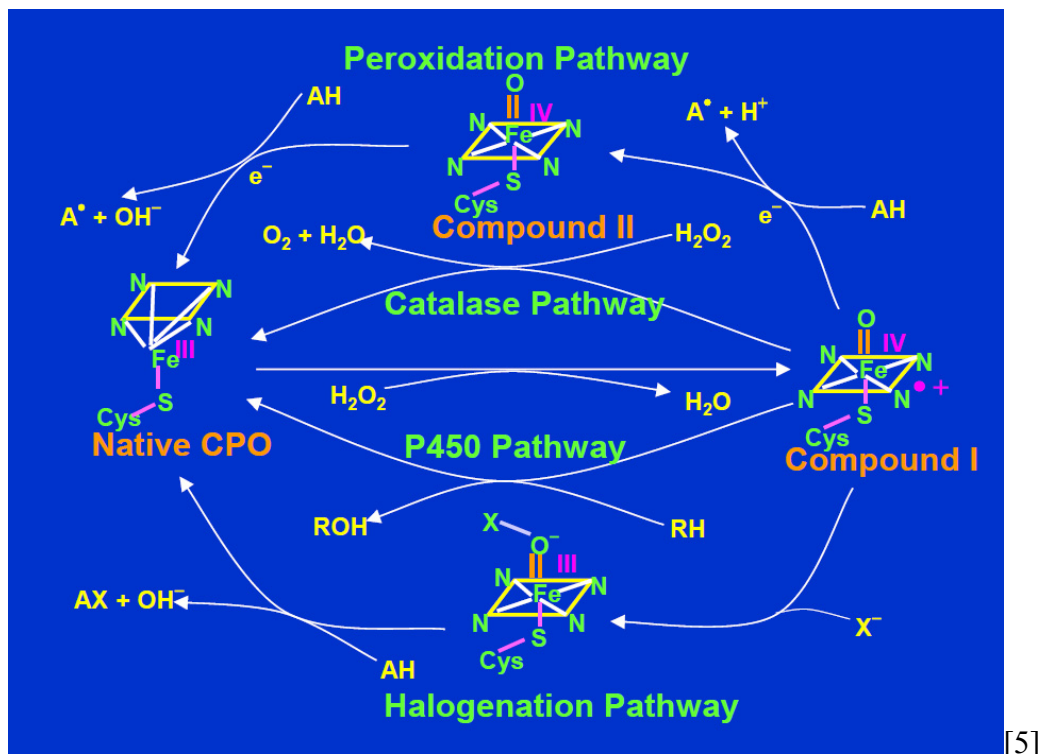
Figure 2: Reaction pathways for different reactions CPO catalyzes

Chloroperoxidase is known to catalyze several different types of oxidation reactions, summarized in Figure 2. All of these reactions begin with activation of CPO by an oxygen donor. The most commonly used reagent for activating the enzyme is hydrogen peroxide. The unreactive form of the enzyme, called the resting state, has the heme iron in the +3 oxidation state (ferric). Peroxide activates the resting state to a highly reactive state with the iron formally in the +5 oxidation state (perferryl). The active form of the enzyme is called compound I. In fact, it has been shown that a better description of compound I has the iron in the +4 oxidation state (ferryl) and a delocalized radical cation on the porphyrin ring.

1.1.1. Catalytic Reactions of CPO

Perhaps the most important of CPO's catalytic reactions is the oxygen insertion reaction, one form of which is epoxidation. Chloroperoxidase has been shown to catalyze the epoxidation of a number of organic substrates. These epoxidation reactions are highly enantiospecific, which makes CPO a pharmaceutically significant enzyme with substantial biotechnological potential. It is this potential combined with the ease of manufacture and availability of CPO that has driven the study of this reaction over the past several years. Much has been learned about the reaction mechanism of this reaction. Modification of the histidine residue at the distal site inhibits the epoxidation reaction, suggesting that the reaction occurs near the distal site of the heme. Other catalytic reactions compete with the epoxidation reaction. The most important competitor is the

catalase reaction, which involves the dismutation of hydrogen peroxide to give water and an oxygen atom. The catalase reaction is much faster than the epoxidation reaction and depletes the hydrogen peroxide. The active site near the heme is only accessible via two narrow channels. One possible reason that the catalase reaction can out-compete the epoxidation reaction is that the substrates for the epoxidation reaction have to travel through the narrow channel to reach the active site. One line of research is aimed at controlling or inhibiting the catalase activity. It was discovered that there are different optimal pH ranges for different substrates and that pH $5.5 - 6.0$ diminishes the catalase activity while not affecting the epoxidation reaction greatly. There is, however, a risk of deactivating CPO at the high end of this pH range. Using an alkyl peroxidase instead of hydrogen peroxidase as the oxygen donor slows down the catalatic activity, but the alkyl peroxide tends to lower the stability of Chloroperoxidase. Other mutation studies are being performed to discover motifs that would reduce catalase activity while maintaining or improving the epoxidation turnover rates.

Chloroperoxidase also catalyzes the dehydrogenation of substrates that is characteristic of heme peroxidase enzymes [6, 7]. The mechanism for this reaction is not completely understood. It is thought that for most, though not necessarily all, substrates the peroxidase reaction occurs at the surface of the enzyme and not near the heme active site. Like the epoxidation reaction, the peroxidase reaction also competes with the catalase reaction.

Another reason CPO is extensively studied is that it shares structural and functional features with several important enzymes, most importantly cytochrome P-450 and the thyroid enzyme involved in the iodonization process. Chloroperoxidase is more

easily produced and activated than these enzymes. Structure and reactivity can be studied using CPO, and the findings can be extrapolated to these other enzymes to better understand how they function.

The most prominent of CPO's catalytic reactions is the chlorination reaction, which has ecological implications [8-14]. In addition to the chlorination reaction, CPO also catalyzes other halogenation reactions including those of iodine and bromine, but not fluorine. The enzymatic chlorination reaction occurs at low pH. There was early concern [9] that the global prevalence of organohalogens was mainly due to anthropogenic contributions. However, several experiments have been able to show that there is significant contribution from biological activity to the global organohalogen burden. Given that CPO is easily accessible and found in nature, it was the enzyme of choice used by many research groups to demonstrate the possibility of enzymatic catalysis of chlorination of organic compounds found in plant litter, fluvic acid, lignin structures and humic substances found in soil samples. There are also examples of plant matter still attached to trees that contains organohalogens originating from degradation, presumably from microbial or fungal attacks [13]. The source of the halogen and oxidizer could be the plant itself, which tends to produce hydrogen peroxidase in cell walls [13]. There are several organisms like fungi, algae and bacteria that produce similar enzymes that would explain the presence of organohalogens in soils, surface water and volatile halogenated anisoles found on surface of oceans [14].

1.1.2. Previous Work on Mechanism of CPO-catalyzed Halogenation Reaction

Given the versatility and potential economic significance of CPO, there is little wonder that it has been studied extensively over the years. However, its most prominent reaction, chlorination, is still not completely understood. One postulated mechanism involves catalytic formation of HOCl at the active site of CPO followed by release into solution, where it halogenates substrates [4, 15-21]. The other postulated mechanism involves the chlorination of substrates at the heme active site [22-30]. Studies have been performed to support either of these two postulates, but it has been difficult for a consensus to be reached.

The non-active site hypothesis is based on the observation that most of the chlorinated substrate products show no enantiomeric excesses; that some substrates are too large to access the active site; and that HOCl and Cl_2, which can chlorinate any of the substrates, are observed in solution in chlorination reactions involving CPO.

The active site hypothesis is based on the following experiments. (1) There is a preference for certain substrates over others when the reaction is carried out using a mixture of substrates. For example, the preference for thiourea and methionine over monochlorodimedone (MCD) is 50:1 and 30:1, respectively, for the enzymatic reaction. By comparison, the preference for thiourea over MCD is 2:1 when HOCl is directly added to the reaction mixture in the absence of CPO, and 3:1 when Cl_2 is directly added. There is no preference for methionine over MCD using either HOCl or Cl_2 directly [24]. (2) The rates of chlorination of some substrates are higher for the reactions in the presence of CPO than for reactions involving chemical controls [23]. (3) Woggon et al.

developed synthetic active-site analogs of CPO for which the formation of stable intermediates with OCl$^-$ and HOCl bound to the heme was demonstrated. These were presumed to be the intermediates through which the enzyme chlorinates the substrates [27-31]. (4) Some steady state kinetic studies suggest that chlorination via the free HOCl mechanism is unlikely. The studies show that the production of HOCl and Cl$_2$ at the enzyme active site and the subsequent release into solution is a slow process that may be considered as competing with the chlorinating reaction [22, 23]. (5) Some substrates give different product distributions in the enzyme catalyzed reaction and the chemical control reaction. For example, while no appreciable stereoselectivity is observed for most chlorinated substrates, for 2methyl-4-propyl-cyclopentane-1,3-dione it was reported that the enzyme catalyzed reaction results in a product, 2-chloro-2-methyl-4-propyl-cyclopentane-1,3-dione, with a 40:60 ratio of the racemic diastereomers, while the chemical control gives a 55:45 ratio of the racemic diastereomers [26]. (6) There is also a reported case of stereoselective halohydration of glycals and steroids using CPO [32-35].

Thus the experimental evidence regarding the mechanism of enzymatic halogenation by CPO is inconclusive. Some consider the general lack of enantiospecificity to be the most persuasive evidence and conclude that halogenation occurs outside the active site. However, the many observations favoring active-site halogenation leave the case far from clear. The bulk of the work presented in this dissertation is aimed at contributing to the resolution of this dispute.

1.1.3. Substrate Used to Study Reaction

In the research presented here, I employ a variety of theoretical approaches to study the chlorination mechanism of CPO, and I choose a particular model substrate, 1,3 cyclopentanedione (CPD), for consideration. This substrate was chosen for several reasons. First, CPD is a cyclic β-diketone, and it is known that cyclic β-diketones, β-keto acids, and substituted phenols make good halogen acceptors. Second, 2,2-dichloro-1,3-cyclopentanediol was found to be the main metabolite of CPO [1]. CPD has been shown to be catalyzed to 2,2-dichloro-1,3-cyclopentanedione [36]. CPD Third, CPO and another cyclic β-diketone, monochlorodimedone (MCD), are the compounds most widely used for monitoring the halogenation activity of CPO spectroscopically. The intermediate species in the reaction of CPD to the dichloro metabolite is a monochloro species, 2-chloro-1,3-cyclopentanedione [36]. It is easy to observe the conversion of CPD to the metabolite because CPD and the intermediate have absorption maxima at 242 and 248 nm, respectively, with a high extinction coefficient while the metabolite has an absorption maximum at 277 nm with a very low extinction coefficient [37]. Monochlorodimedone is more commonly used because it is readily available, and it is easy to observe the conversion from MCD to dichlorodimedone [37]. For computational studies, though CPD is preferable to MCD because it contains fewer heavy atoms and is hence less demanding of computational resources. Finally, a crystal structure of CPO with CPD at its active site has been reported [38]. This structure will be useful for future studies of halogenation at the active site.
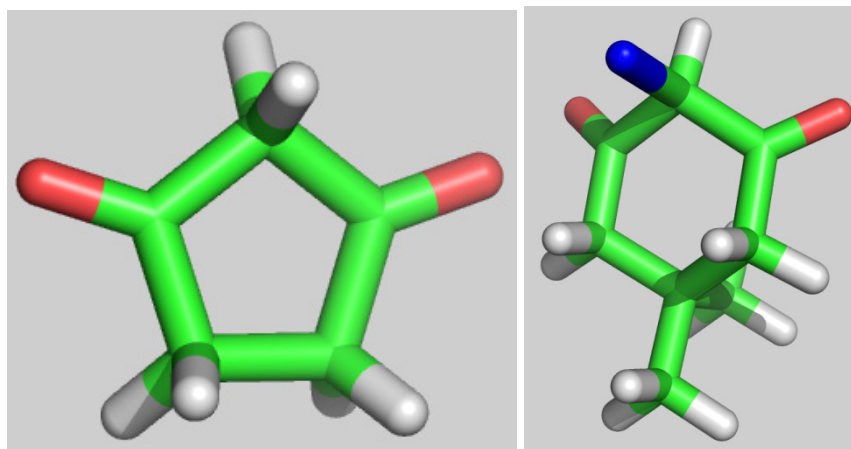
Figure 3: Substrates – CPD and MCD
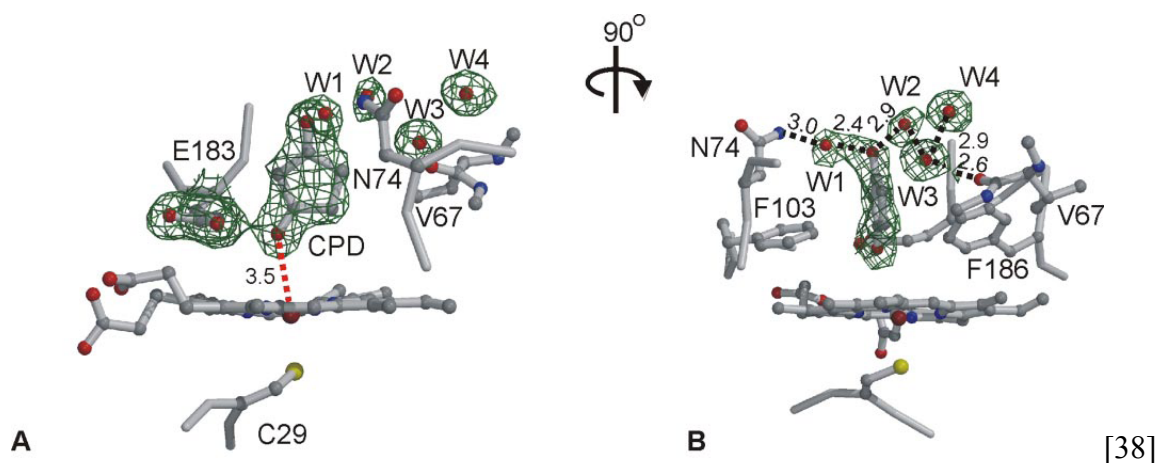
Figure 4: CPO active-site showing CPD and water molecules

1.2. Chlorination Mechanism of CPO

It is well known that ketones in aqueous solution can undergo an internal rearrangement in which a hydrogen is transferred from the α-carbon to the keto oxygen and the double bond shifts from C=O to C=C. The resulting species is called an enol because it contains both –ene (carbon-carbon double bond) and alcohol (OH) groups.

The two forms are called tautomers, and the process of interconversion is tautomerization. In aqueous solution, both tautomers are present in equilibrium. An alternative nomenclature for the tautomers of β-diketones is diketo and keto-enol.

The overall process of the chlorination reaction, with keto-enol tautomerization as the first step, is depicted in Figure 5. The mechanisms of the individual steps are shown in Figure 6 and Figure 7. The chlorination (or more generally, the halogenation) of ketones begins with tautomerization to the enol form. Electrophilic attack of a chloronium (Cl^+) on the carbon-carbon double bond of the enol form, followed by transfer of the hydroxy proton to a solvent water, yields the chlorinated product. The tautomerization step can be either acid- or base-catalyzed, as shown in Figure 6. I consider acid catalysis because CPO-catalyzed chlorination is carried out at low pH (pH 3-4).
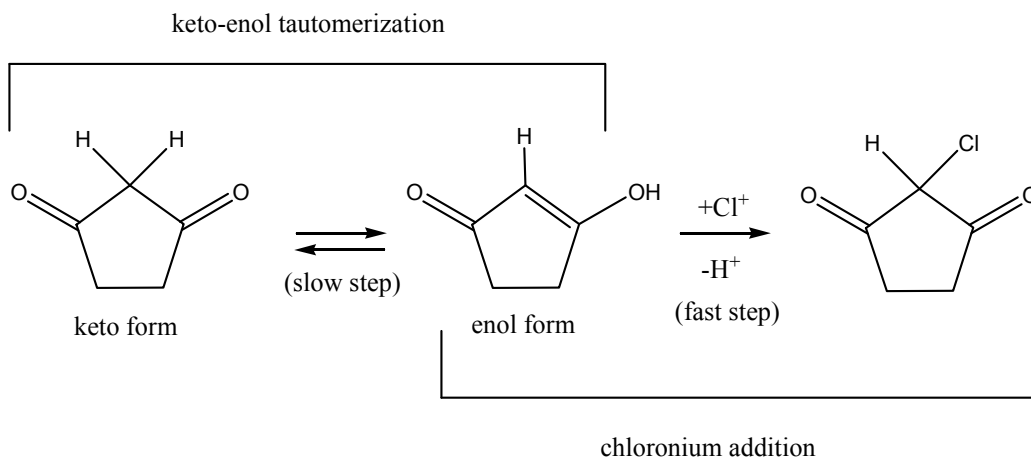
Figure 5: Chlorination of CPD
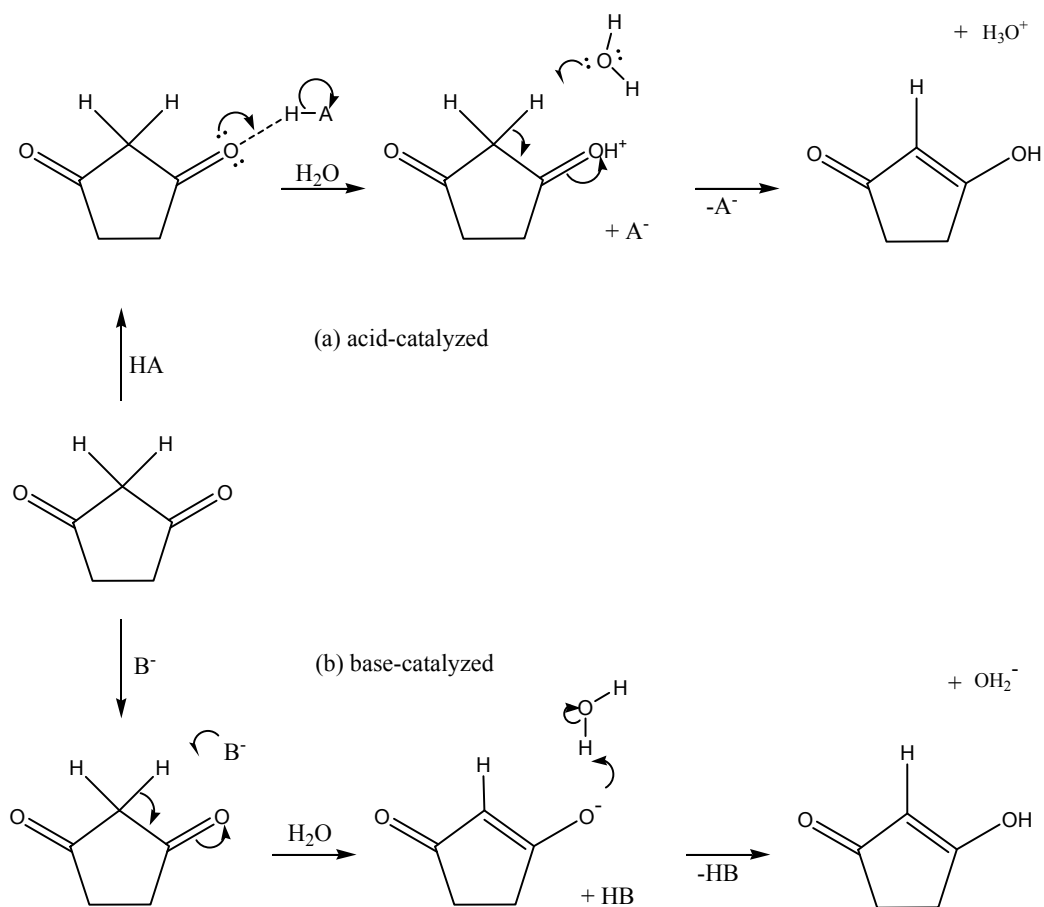
Figure 6: Mechanism of keto-enol tautomerization: (a) acid catalyzed, (b) base catalyzed
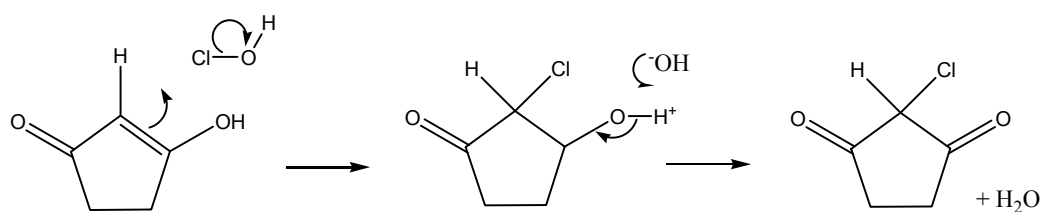
Figure 7: Mechanism of chlorination step

For the overall process of the chlorination reaction shown in Figure 5, the second step is expected to be fast and barrierless. The mechanism of this step is depicted in Figure 7. I modeled the chlorination step in a limited hydration environment, with 14

water molecules, using the quantum mechanics approach explained in section 2.4. The data in Figure 9 verifies that the reaction shown in Figure 8 is nearly barrierless and hence a kinetically fast reaction. Therefore, I focus the study on the rate-limiting step which is the acid catalyzed keto-enol tautomerization.
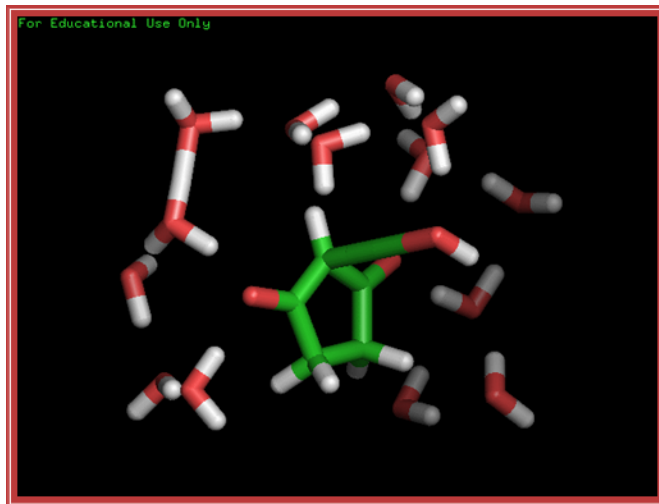
Figure 8: Second step of the chlorination mechanism in a limited hydrated environment
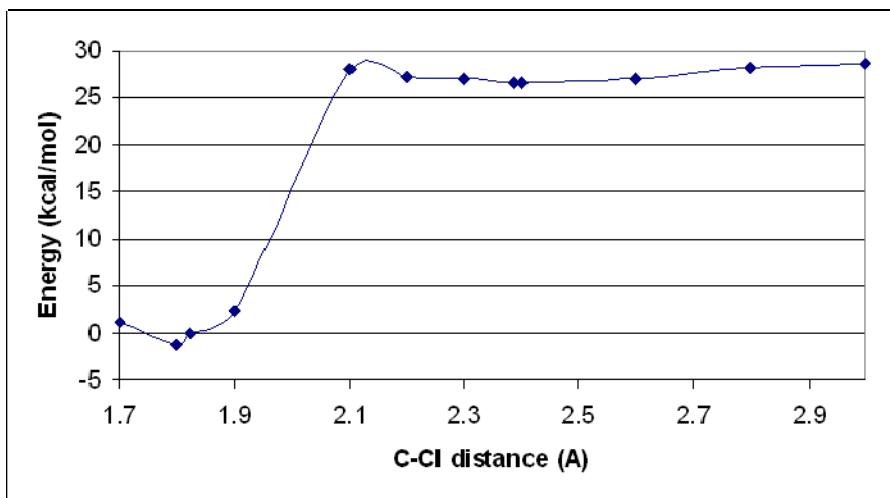
Figure 9: Second step of the chlorination mechanism depicting barrierless reaction

1.3. *Previous theoretical studies of keto–enol tautomerization*

Solvation plays a critical role in many reactions including keto-enol tautomerization and will be the focus of my studies on the chlorination mechanism of CPD. Considerable prior theoretical work has been done on the role of water in keto-enol tautomerization. Here this prior work is summarized.

The equilibrium between the keto and enol tautomers in solution has been studied with implicit solvent methods, with reasonable success, reproducing logK, where K is the equilibrium constant, to within ~±0.2 in many cases [39-41]. However, such methods are not sufficient for modeling the kinetics of tautomerization in aqueous solution because the mechanism explicitly involves water molecules.

Calculations determine the barrier for keto-enol tautomerization via intramolecular proton transfer to be generally in the range 60-85 kcal/mol, depending on the compound. This is much larger than the experimental reaction barriers [42-44]. Inclusion of a single explicit catalytic water lowers the barrier by 30-40 kcal/mol for typical molecules. The mechanism is a two-step process with the water first abstracting a proton from the alpha carbon, and subsequently donating a proton to the keto oxygen. Ma et al. determined that a second catalytic water lowers the tautomerization barrier for 3-methyl-5-hydroxyisoxazole a further 8 kcal/mol [43]. Freitag et al. showed that for malonaldehyde, a third explicit water with a solvating but not catalytic role can lower the barrier another 7 kcal/mol [44]. Further solvating waters had decreasing incremental effects. These studies demonstrate that inclusion of explicit water molecules is crucial

for modeling the kinetics of keto-enol tautomerization, but the degree of explicit solvation required for an accurate model is not known.

Cucinotta et al. applied an ab initio metadynamics treatment to keto-enol tautomerization of acetone in aqueous solution [45]. They found that a single water lowers the free energy barrier from 57.7 to 46.1 kcal/mol, and a solution environment of 28 waters lowers the barrier further to 38.5 kcal/mol. In solution, a Grotthuss mechanism involving four water molecules in a proton transfer chain was observed.

With the exception of the metadynamics work, all the studies focused on optimized structures with up to two catalytic waters, and in one case with up to three further waters that were solvating but not catalytic. It is unknown whether this degree of hydration is sufficient. More fundamentally, the assumption is made that barriers for fully optimized structures involving multiple waters can accurately predict experimental activation energies. However, such structures might represent reaction events so extremely rare that they do not predict the observed kinetics well. I address these issues with two studies on CPD. In the first study, I systematically increase the number of explicit waters in fully optimized structures and determine the barrier and the reaction energy as functions of the number of waters. I find that the barrier to reaction continues to decrease as the number of waters increases beyond five. In the second study, I use classical molecular dynamics to generate solution environments representative of 300 K. I then determine reaction barriers from snapshot structures using a hybrid QM/MM method, allowing one or two water molecules to be explicitly active. I find that the instantaneous environment has a significant effect on the barriers. This result, ultimately,

is consistent with Cucinotta's metadynamics work [45] but provides complementary detail on the nature of the influence of water on this keto-enol tautomeric reaction.

The energies of the stationary points (reactant, product and transition state) were calculated in different environments as described above. I start with the simplest system consisting of only the CPD and HOCl molecules, with or without a single catalytic water, in the gas phase. I then focus most of the study on systems consisting of CPD, HOCl and one reactive/catalytic water molecule in different environments.

1.4. Solvation

Based on previous theoretical work on keto-enol tautomerization it is clear that solvation effects are quite substantial and critical, as they influence the reaction barrier strongly. Both at the active site and in aqueous solution, water molecules are present and will interact with the substrate during reaction. Water itself, in fact, plays a catalytic role in the reaction. Furthermore, water molecules not directly involved in the breaking and forming of chemical bonds may also influence the reaction strongly through hydrogen bonding, in addition to less directional dielectric effects that water exerts. Thus we will find that the way in which water is incorporated into the reaction model is critical. The most basic way to represent solvent in theoretical calculations, as a dielectric continuum, turns out to be entirely inadequate in this case.

In reaction at the active site, a limited number of waters are available to interact with the substrate. This is apparent from the crystal structure shown in Figure 4, which

identifies four waters quite near a CPD molecule bound at the active site. These four waters are resolved in the X-ray structures because they have strong stabilizing interactions with neighboring moieties. In reality, there may be other waters in addition to these four, but the total number is limited by the size and shape of the active site pocket. Thus in the context of the active site, it is misleading to speak of solvation because the number of waters is limited. In this work, I will refer to such a water environment as "limited hydration." In contrast I will refer to the water environment of CPD dissolved directly in aqueous solution as "full solvation."

The central work in this dissertation reports on two investigations, one of the chlorination reactions in "limited hydration" conditions and the other chlorination reaction in "full solvation" conditions. Quite different approaches are used for the two investigations. The limited hydration work connects with previous theoretical work reported on keto-enol tautomerization, which is a critical step in the reaction (see below). Keto-enol tautomerization is a fundamental concept in organic chemistry, but the influence of the water environment on the reaction is still not fully understood. My work on limited hydration suggests that previous work on keto-enol tautomerization needs to be understood in a more restricted way than that work made clear. The energetics worked out in the previous work is representative of the particular limited hydration structures used, but not representative of limited hydration in general and certainly not representative of full solvation.

My work on the reaction in "full solvation" at near-physiological temperature leads to a paradoxical observation. The dominant reaction events have energetic barriers very similar to those encountered under extremely limited hydration conditions, namely a

single water that assumes a catalytic role. This similarity in barriers turns out to be a consequence of equivalent degrees of reactant- and transition-state stabilization by other waters in the full-solvation calculations. The full-solvation barriers are quite different from the barriers for limited hydration with several waters present.

The work on the chlorination reaction presented here is restricted to the various water environments just described. This turned out to be rich and difficult undertaking in itself. Reaction of CPD situated within the enzyme CPO at the active site is a subject for future work. The work presented here demonstrates that inclusion of hydrating waters will be critical for modeling the reaction at the active site. Implications of this work extend beyond CPO, as many chemical reactions are likely to be sensitive to the nature of the limited-hydration environment found at an active site.

The rest of this dissertation is organized as follows. Chapter 2 contains an outline of theoretical methods used (content on basic quantum mechanics and molecular mechanics is from standard computational chemistry texts [46, 47]). Part of the work presented is based on molecular mechanics models. This necessitated the development of such models for CPD and other molecules used. This is an involved process of parameterization, and it is detailed in Chapter 3. Chapter 4 presents the research on the chlorination reaction in limited hydration conditions representative of the active site environment. Chapter 5 presents the research on the chlorination reaction in fully solvated conditions representative of aqueous solution. Chapter 6 presents the major conclusions regarding the halogenation mechanism. Chapter 7 briefly describes contributions to three other research projects. The appendix contains the most important scripts used for running the calculations.

2. METHODS

2.1. *Molecular Mechanics*

The term Molecular Mechanics (MM), which was first introduced in 1970, describes the application of classical mechanics in the determination of molecular structures at equilibrium. In the simplest MM treatment of molecules, the atoms are treated as balls and the bonds as springs connecting them. These springs follow Hooke's law, which is a representation of a harmonic oscillator. The application of classical mechanics to the motion of these balls (or atoms) bound by springs (or bonds) is the basis of molecular mechanics. Thinking of molecules as being made up of balls and springs dates back to as early as 1930 and was first suggested by Andrews [48, 49].

It might seem as if the laws of quantum mechanics are ignored with this treatment of molecules. In reality, though, the force fields are based on experimental and quantum mechanical calculations, so quantum mechanics is incorporated in the force field. The MM treatment of molecules works quite well as long as the structures do not deviate too much from their respective equilibrium structures. The harmonic oscillator approximation of the bonds often provides sufficiently accurate results. However, bond breaking or forming cannot be modeled using molecular mechanics.

Quantum Mechanics (QM) cannot be used to model large molecules such as proteins, which number in the thousands of atoms due to computational costs. The MM models use a simplistic representation of atoms in comparison to quantum mechanics to allow one to

18

study systems with large numbers of atoms with reasonable accuracy. Molecular mechanics is used extensively in the study of biomolecular systems like docking for drug design, protein folding, binding constants and reaction precursors.

One of the main objectives of a good force field is transferability from molecule to molecule. For example, we would like to be able to apply the force field derived from one molecule to another similar molecule to correctly predict the minimum energy structure for the new molecule. This key development in MM, where it was possible to transfer force constants between alkanes was first made by Snyder and Schachtschneider [50].

The functional forms used for developing force fields are described below. The various types of interactions in a macromolecule can be classified as bonded interactions and non-bonded interactions. The potential energy contributions from the various interactions are combined to give the total potential energy of the macromolecule. Each contribution is described separately below.
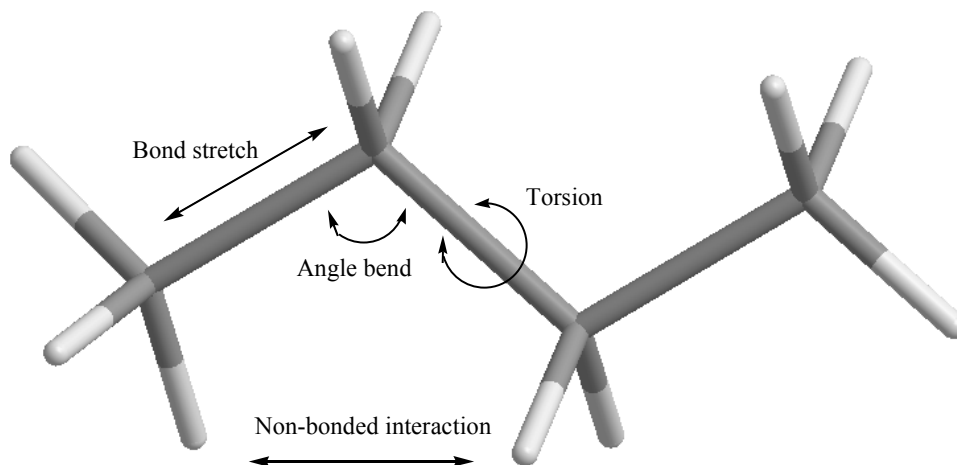
Figure 10: Bonded and non-bonded interactions in MM

The program I use for the MM model is CHARMM [51]. CHARMM is a molecular simulation program developed to study large, many-atom structures usually encountered in biological systems, for example proteins, peptides, nucleic acids, carbohydrates and organic drug-like molecules or ligands in solution, crystal or membrane environments. CHARMM provides computational tools to perform molecular energy minimization, molecular dynamics, analysis and path sampling. It can be interfaced with several QM software packages to study systems that require the hybrid QM/MM method explained in section 2.5.

2.1.1. Bonded Interactions

Bond Stretching

To determine the potential energy contribution from bond stretching we consider the ball and spring model. The force due to the spring is given by Hooke's Law

$$F_s = -k_s(R - R_e)$$

The potential energy is given by

$$U(r_{AB}) = -\int F(r_{AB})dr$$

Applying a Taylor expansion we obtain

$$U(r) = U(r_{eq}) + \frac{dU}{dr}\bigg|_{r=r_{eq}}(r - r_{eq}) + \frac{1}{2!}\frac{d^2U}{dr^2}\bigg|_{r=r_{eq}}(r - r_{eq})^2 + \frac{1}{3!}\frac{d^3U}{dr^3}\bigg|_{r=r_{eq}}(r - r_{eq})^3 + ...$$

We label the equilibrium bond distance r_{eq}. We can, for convenience, consider the minimum of the functional form to be zero. Therefore, the first and second terms in the Taylor's expansion are zero. The simplest contribution then is from the first non-zero term which is the quadratic term. This approximation may be sufficient for small stretching, but as the bond stretches far from the equilibrium values the potential energy contribution becomes infinitely positive, which is unrealistic, and the harmonic approximation fails. This failure is inherent in truncating the Taylor's expansion, so the logical thing would be to include more terms. When the next cubic term is included, it causes an unintended divergence to negative infinity as the bond is stretched further. Therefore, the quartic term also has to be included.

Alternately, the functional form of the Morse potential can be used which describes the bond stretch rather accurately.

$$U(r_{AB}) = D_{AB}[1 - e^{-\alpha_{AB}(r_{AB} - r_{AB,eq})}]^2$$

However, the exponential term in the Morse potential is computationally less efficient for calculation than the truncated polynomial terms from the harmonic oscillator. In most force fields some variation of the following bond stretching term is used

21

$$U(r_{AB}) = \frac{1}{2}[k_{AB} + k_{AB}^{(3)}(r_{AB} - r_{AB,eq}) + k_{AB}^{(4)}(r_{AB} - r_{AB,eq})^2 + ...](r_{AB} - r_{AB,eq})^2$$
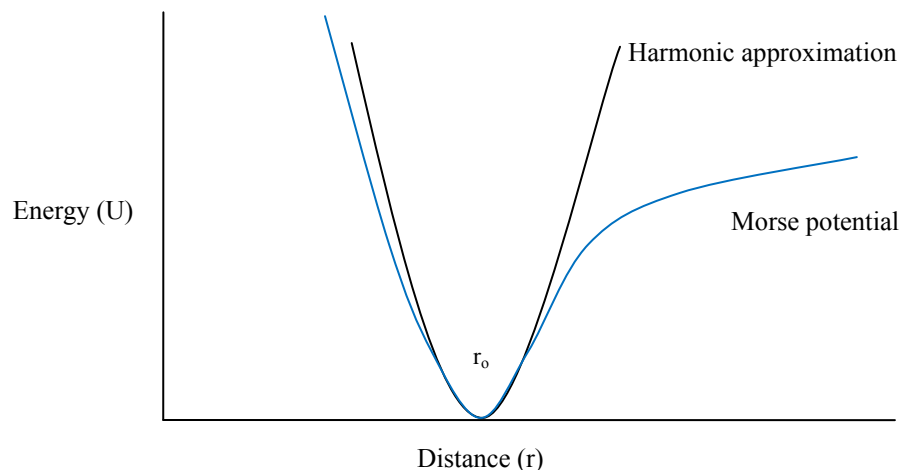
Figure 11: Bond stretch energy functions

Angle Bending

As with bond stretching, angle bending or bond angle deformation can be modeled using the harmonic approximation about the equilibrium bond angle. It has an important contribution to the total potential energy as evidenced by vibrational spectroscopy data that show energy variation attributed to bond bending. The same considerations as for bond stretching apply to the functional form, so a generally useful function form is

$$U(\theta_{ABC}) = \frac{1}{2}[k_{ABC} + k_{ABC}^{(3)}(\theta_{ABC} - \theta_{ABC,eq}) + k_{ABC}^{(4)}(\theta_{ABC} - \theta_{ABC,eq})^2 + ...](\theta_{ABC} - \theta_{ABC,eq})^2$$
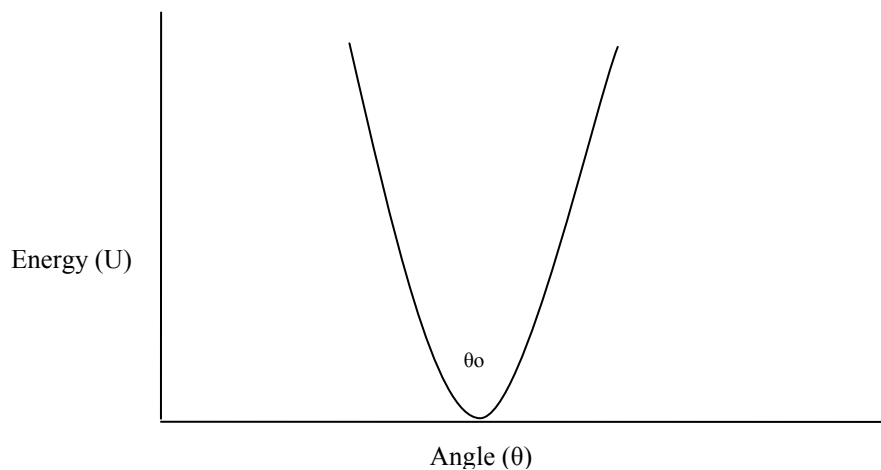
Figure 12: Angle bending energy function

Torsions

Torsional motion is periodic, so its contribution to the potential energy is also periodic. It is conveniently modeled using a cosine function, or more generally a sum of cosines to allow for irregularity in the periodic function. Energy contribution due to dihedral (torsional) motion involving four atoms is given by the following expression.

$$U(\omega_{ABCD}) = \frac{1}{2} \sum_{\{j\}_{ABCD}} V_{j,ABCD}[+(-1)^{j+1} \cos(j\omega_{ABCD} + \psi_{j,ABCD})]$$
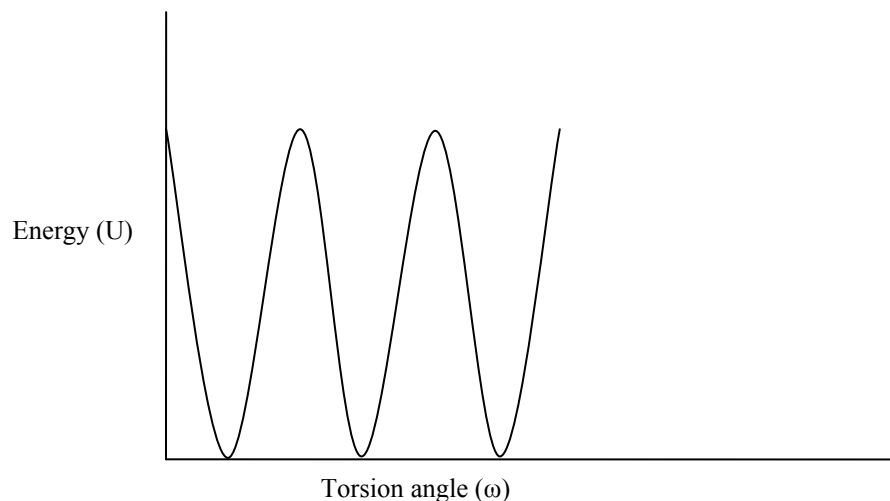
Figure 13: Torsion energy function

Improper torsions
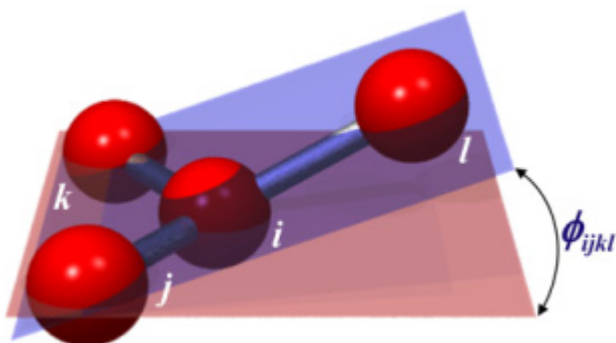
[52]

Figure 14: Improper torsion angle

Improper torsions are a special situation that involves four atoms, but unlike the case with regular torsions, for improper torsions there is a central atom bonded to the three other atoms. If we define a plane by the central atom and two other atoms that are bonded to it, then the angle the fourth atom creates to that plane is the improper torsion angle. Improper torsion terms are included only when the other energy terms are

24

insufficient to maintain correct geometries. Specifically, improper torsion terms are used to maintain puckering in structures. Improper torsional energy contributions are only accounted for by CHARMM if there are explicit parameters for them for a specific molecule. Improper torsion terms are not included in the energy by default like the other bonded interaction terms. The CHARMM topology file must specify that a particular improper torsion term be included.

2.1.2. Non-bonded Interactions

As stated earlier, one of the main objectives in developing a force field is the transferability of the force field to similar groups of molecules. During the initial development of force fields it was observed that 'transferability' was attainable by including non-bonded interactions, at least of the nearest neighbors. This is why non-bonded interaction contributions are critical and have to be included. The following two types of non-bonded interactions are included in most force fields, while some force fields include additional interactions.

Van der Waals Interactions

Van der Waals force is the attractive forces experienced by non-bonded neutral atoms. As neutral atoms gets in close vicinity to each other they experience attractive forces due to the induced dipoles. These short distance intermolecular forces are weaker

than a chemical bond, but are important when atoms get close enough like in liquids and solids and are responsible for properties like surface tension and capillary action.

In the simplest case we have two atoms, represented as hard spheres, each with a characteristic radius. We also assume they have no charge. At infinite separation, there is no interaction and hence the contribution to the potential energy is zero. As they approach each other the two spheres eventually touch. At this point they cannot approach each other and the energy contribution tends to infinity. Realistically, we know from QM that this is a simplistic case and in fact as the atoms approach each other there is an attractive force due to the electronic correlation, also called dispersion. As they come closer, at some very short distance, the energy contribution tends to rise exponentially to infinity. The attractive energy is inversely proportional to r^6. This dispersion interaction between atoms is usually modeled by the Lennard-Jones potential and is given below.

$$U(r_{AB}) = 4\varepsilon_{AB}[(\frac{\sigma_{AB}}{r_{AB}})^{12} - (\frac{\sigma_{AB}}{r_{AB}})^6]$$

It would be more accurate to represent the repulsion with an exponential term instead of the inverse r^{12} term. But, exponential calculations are more difficult on the computer in comparison to a polynomial term, which is why the inverse r^{12} term is used instead. This is an empirical result and used for computational convenience.
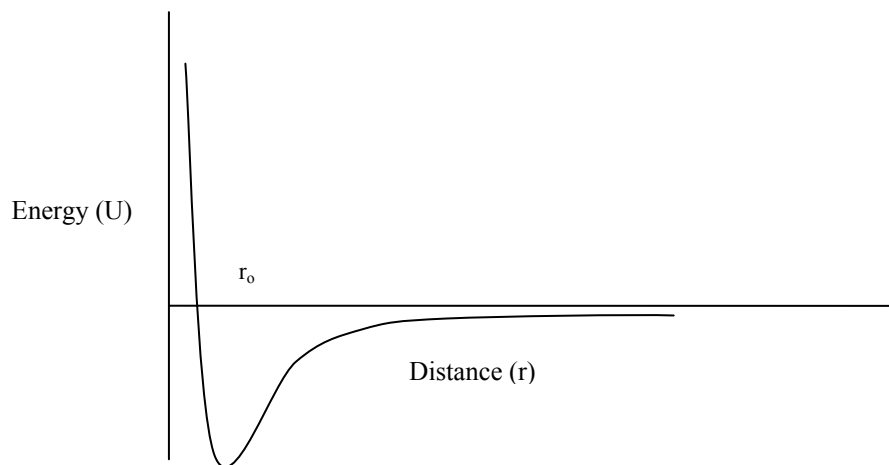
Figure 15: Lennard-Jones potential for van der Waals interaction

Electrostatic Interactions

Electrostatic interactions are the energy contribution that arises from non-bonded interactions between atoms that have a charge associated with them. Again, if we consider the simplest case, we assume a fixed point charge associated with an atom. We then apply Coulomb's Law to determine the energy contribution which is given by the following equation.

$$U_{AB} = \frac{q_A q_B}{\varepsilon_{AB} r_{AB}}$$

We know that as atoms move the effective atomic charges will respond to the environment, that is to say, atomic systems are polarizable. When permanent charges are assigned, the polarizability is neglected. However, the fixed permanent charges used are generally capable of reproducing several experimental observable or QM computed

properties quite accurately. Development of polarizable force fields is an active topic of research and there are several groups working on developing polarizable force fields. However, the community has not yet embraced any particular polarizable force field.

The calculations of non-bonded interactions are the most time consuming part of the energy evaluations since it would mean calculation for every pair of atoms in the system ($N^2/2$ interactions, where N is the number of atoms). These interactions are distant dependent and two atoms that are quite far from each other may have minimal contribution to the total energy. An implementation consideration then is how far we want to calculate non-bonded interactions. In interest of saving computation time a truncation scheme is used. A cutoff distance is set so that non-bonded interactions are only calculated if an atom is within a certain distance from another atom. However, this introduces a discontinuity in the energy evaluations. Consider for example, two atoms that are exactly at the cutoff distance when the energy contribution is calculated. At the next step, they move just beyond the cutoff distance from each other, thus the new energy contribution calculated is precisely zero. The small step would cause a drastic difference in the energy evaluation. To avoid this abrupt discontinuity, a switching function is used to gradually exclude the non-bonded interactions instead of abruptly. For example, if the cutoff was set at 12 Å, one can use a switching function to gradually exclude the non-bonded interactions from 12 Å to 14 Å instead of abruptly setting it to zero at 12 Å. The preferred method of treating electrostatic interactions is the Particle Mesh Ewald summation method. It is a method to calculate the electrostatic interactions in an infinite lattice or under periodic boundary conditions (explained in section 2.3.3).

2.1.3. Solvation

Modeling gas phase reactions are relatively simple. However, all biochemical processes occur in the condensed phase and the chemical reaction we study occurs in solution. Therefore, it is critical to model a solvation environment around the molecule of interest. A solvation environment model can be broadly classified as an implicit model or explicit model.

In an implicit solvent model, also called the continuum solvation model, the environment is like a continuous medium with a dielectric constant representing the solvent and properties consistent with those of the solvent. To represent water as the extended solvent environment we use the dielectric constant of water ($\varepsilon = 78.4$). This kind of representation of the solvent has the advantage that the computation overhead is low and used with QM model. I used the Polarizable Continuum Model (PCM) method as implemented in Gaussian98 [53]. The PCM model is a widely used implicit solvent model that represents the environment of the system by an dielectric continuum having the dielectric constant of the solvent. However, an implicit model is only used when the condensed phase does not play a critical role in the reaction environment.

An explicit solvent model includes explicit solvent molecules. Adding explicit water molecules can get prohibitively expensive in the quantum mechanics model. At best it is feasible to include a single solvation shell. However including explicit water molecules is quite feasible in the molecular mechanics model which is routinely done while studying large biomolecular systems. I use both types of explicit solvated environments.

2.2. *Geometry optimization*

One of the key motivations of MM is to find minimum energy structures or equilibrium structures that are representative of realistic molecules. Locating minimum energy structures constitutes a mathematical problem and the general field of study is called optimization theory. One of the priorities is to facilitate geometry optimization by simplifying the functional forms without compromising accuracy. One such example is using polynomial terms instead of exponential terms to facilitate computation. There are several algorithms that have been developed that are used in achieving this goal. Obviously, some are more efficient that others and in most cases it is best to use a combination of different algorithms.

Consider a molecule with N atoms. This molecule has 3N degrees of freedom, of which six are translational and rotational (or five if it is a linear molecule). The remaining $3N - 6$ (or $3N - 5$) degrees of freedom are vibrational degrees of freedom and therefore, there are $3 N - 6$ independent variables that describe them which are also called internal coordinates.

All these variables can be written as a matrix

$$q = \begin{pmatrix} q_1 \\ q_1 \\ q_3 \\ ... \\ q_P \end{pmatrix}$$

The simplest way to optimize would be to optimize the energy for one variable at a time. This is called a line search, but when we have several variables and if they are correlated this approach becomes inefficient.

Another approach would be to proceed in a direction that minimizes the total energy with respect to all the variables at the same time. The first derivative matrix, also called the gradient and shown below, is needed. The optimization process proceeds by altering the coordinates so as to modify the structure in the negative direction of **g(q)** until a local minimum is reached. This is a first order approach.

$$g(q) = \begin{pmatrix} \dfrac{\partial U}{\partial q_1} \\ \dfrac{\partial U}{\partial q_2} \\ ... \\ \dfrac{\partial U}{\partial q_p} \end{pmatrix}$$

Steepest Decent methods, which was introduced by Wiberg in 1965 [54], and Conjugate Gradient methods, which were introduced by Fletcher and Reeves [55], are examples of first derivative methods. Such methods get us close to the minimum quickly, but an inherent drawback of such an approach is the convergence. It can sometimes be hard for such algorithms to converge to the minimum. One usually starts an optimization process with such an algorithm to get close to the minimum. Then one uses a more sophisticated method, usually a second order method, to finally converge to the minimum.

Second order methods make use of the second derivative of the energy function, also called the Hessian.

$$H(q) = \begin{pmatrix} \dfrac{\partial^2 U}{\partial^2 q_1} & \dfrac{\partial^2 U}{\partial q_1 \partial q_2} & \cdots & \dfrac{\partial^2 U}{\partial q_1 \partial q_p} \\[2ex] \dfrac{\partial^2 U}{\partial q_2 \partial q_1} & \dfrac{\partial^2 U}{\partial^2 q_2} & \cdots & \dfrac{\partial^2 U}{\partial q_2 \partial q_p} \\[2ex] \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\[2ex] \dfrac{\partial^2 U}{\partial q_p \partial q_1} & \dfrac{\partial^2 U}{\partial q_p \partial q_2} & \cdots & \dfrac{\partial^2 U}{\partial^2 q_p} \end{pmatrix}$$

If we substitute in the Taylor's expansion series of energy we obtain

$$U(q) - U(q_{eq}) = \sum_{i=1}^{p} \xi_i \left(\frac{\partial U}{\partial q_i} \right)_{\xi_i = 0} + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \xi_i \xi_j \left(\frac{\partial^2 U}{\partial q_i \partial q_j} \right)_{\xi_i = 0, \xi_j = 0} + \ldots\ldots$$

$$U(q) - U(q_{eq}) = \xi^T g + \frac{1}{2} \xi^T H \xi + \ldots\ldots$$

Second order methods tend to converge to the minimum quickly if they are close to the minimum to begin with. However, they are computationally more expensive because of the necessity of calculating the Hessian, especially if needed at every step. In practice, it is necessary to calculate the Hessian for the first step when using a second order method, but some algorithms use an approximate Hessian for subsequent steps to make the

calculation computational efficient. However, in some sensitive cases it is essential to calculate the Hessian for every step, which is the basis for the Newton-Raphson method.

The minimization algorithms used in this project were a combination of first order and second order methods. Minimization was initiated using Steepest Descent followed by the Adopted Basis Newton Raphson (ABNR) method. ABNR is a quasi-Newton-Raphson method.

2.3. *Molecular Dynamics*

The most popular application of Molecular Mechanics is optimization of a structure to find its minimum energy or an equilibrium structure using the laws of classical mechanics. While this is quite useful, it is even more interesting to predict molecular properties computationally that may or may not be feasible to measure experimentally. One technique that is used to do this computationally is called Molecular Dynamics (MD).

2.3.1. Phase Space

In order to describe MD, I will first discuss the concept of phase, and then show that MD is a way to sample phase space in order to compute average properties. The state of a macromolecule can be completely described using classical laws by specifying the position and momenta of all the particles that make up the macromolecule. A system

made up of N particles would need 6N coordinates to describe the macromolecule, 3N coordinate to describe position along the x, y and z axes and 3N coordinates to describe their corresponding momenta along the respective axes. The resulting 6N dimensional space is called the 'phase space' of the macromolecule. At any given instant in time, the macromolecule occupies one point in phase space. Over time, a system maps out a trajectory in phase space. Molecular dynamics is a technique used to construct the phase-space trajectory of a macromolecule, given a set of initial conditions.

Phase space point

$$X = (q, p)$$

$$q = (x_1, y_1, z_1, x_2, y_2, z_2, ...)$$

$$p = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, ...)$$

Properties measured are averages over the states of the system. Phase-space is extremely large and encompasses every possible state of the system, even those that the system is unlikely to sample, like unrealistic high energy states. In other words, phase space includes states that have a very low probability of being visited in reality. Thus phase-space is essentially a large wasteland that mostly contains states that are unlikely to contribute much to the expectation or average value of a particular property. Hence, it

makes sense to use available computational resources efficiently by only mapping out trajectories that concentrate on regions of phase space the system would be most likely to sample. These regions would likely be near local minima or equilibrium structures. For this reason, most MD simulations are preceded by short energy minimization simulations. (Another reason MD simulations are preceded by a short energy minimization is that most structure coordinates come from x-ray crystallography. The crystal structures usually contain bad contacts between atoms with artificially high forces. If we start the MD simulation with these high forces, the trajectory may be pushed away from the local minima into high-energy, improbable regions of phase space.)

Experimental measurements of properties are averages over states of a system, as mentioned above. Often the system of interest is microscopic or submicroscopic, yet the experimental conditions include a macroscopic sample containing many "copies" of the system. For example in a solution containing a mole of enzyme molecules in which case Avogadro's number of enzyme molecules is present, each capable of sampling a different state of the enzyme. In this case, an experimental measurement of some property of the enzyme will represent an ensemble average over many copies. (Experimental measurements will also be time averaged if the measurement takes longer than the characteristic time associated with the property in question, but the important point here is the ensemble-average nature of most experiments). An MD trajectory provides a sampling over time of a single copy (or small number of copies) of the system. The connection between simulation and experiment rests on the assumption that the time-average over the trajectory of a single copy of the system will give the same results as the ensemble average over many copies of the . This assumption is the ergodic theorem. It

allows us to make a connection between simulation and experiment, provided the simulation is long enough for sufficient sampling of phase space.

An interesting aspect of a phase-space point is that the location of the next point can easily be determined from the current position and momentum. The relationship between two positions and the momentum over some time interval is given by

$$q(t_2) = q(t_1) + \int_{t_1}^{t_2} \frac{p(t)}{m} dt$$

The relationship between momenta and acceleration is given by

$$p(t_2) = p(t_1) + m \int_{t_1}^{t_2} a(t) dt$$

The acceleration $a(t)$ is given by

$$a = \frac{F}{m}$$

The acceleration can be calculated from the forces acting on each particle via the force field. Hence, we can map out the trajectory in phase-space in the forward and reverse directions once we have a single phase-space point.

For non-analytical systems the above expression can be re-written using Euler's approximation in the limit $\Delta t \to 0$. We get the same result if we use the Taylor's expansion truncated at second order.

$$q(t + \Delta t) = q(t) + \frac{p(t)}{m} \Delta t$$

$$p(t + \Delta t) = p(t) + ma(t)\Delta t$$

2.3.2. Verlet-leapfrog algorithm

On the basis of the above general discussion of phase space we now consider the algorithm used to map the trajectory of the system through phase space. One way to map out the trajectory is by using the Verlet algorithm. Consider time steps of Δt in the forward and reverse direction.

The expression for position in the forward direction

$$q(t + \Delta t) = q(t) + \left(\frac{dq}{dt}\right)_t \Delta t + \frac{1}{2!}\left(\frac{d^2 q}{dt^2}\right)_t (\Delta t)^2 + \ldots$$

The expression for position in the reverse direction

$$q(t - \Delta t) = q(t) - (\frac{dq}{dt})_t \Delta t + \frac{1}{2!}(\frac{d^2q}{dt^2})_t (\Delta t)^2 + \ .....$$

If we add the two expressions we get

$$q(t + \Delta t) = 2q(t) - q(t - \Delta t) + \frac{1}{2}a_t (\Delta t)^2$$

$$a_t = (\frac{d^2q}{dt^2})_t$$

We can see that to get the next position, after a time interval of Δt, all we need are the current and previous position and the acceleration, which we can get from the force field.

$$a = \frac{F}{m}$$

$$F = -\frac{dV}{dq}$$

The Verlet algorithm allows us to propagate through phase-space with no explicit reference to velocity or momentum. However, in most cases it is useful to run simulations at a particular temperature or gradually to be able to change temperature. Temperature is directly related to random velocity, so one needs to be able to manipulate velocity directly in order to control the temperature of the simulation. The leapfrog algorithm, a

modification to the Verlet algorithm, was developed to couple the propagation of position

and velocity directly.

$$q(t+\frac{\Delta t}{2}+\frac{\Delta t}{2}) = q(t+\frac{\Delta t}{2}) + \frac{dq(t+\frac{\Delta t}{2})}{dt}\frac{\Delta t}{2} + \frac{1}{2!}\frac{d^2q(t+\frac{\Delta t}{2})}{dt^2}(\frac{\Delta t}{2})^2 +.....$$

$$q(t+\frac{\Delta t}{2}-\frac{\Delta t}{2}) = q(t+\frac{\Delta t}{2}) - \frac{dq(t+\frac{\Delta t}{2})}{dt}\frac{\Delta t}{2} + \frac{1}{2!}\frac{d^2q(t+\frac{\Delta t}{2})}{dt^2}(\frac{\Delta t}{2})^2 +.....$$

Subtracting we get

$$q(t+\Delta t) = q(t) + v(t+\frac{\Delta t}{2})\Delta t$$

$$v(t+\frac{\Delta t}{2}) = \frac{dq(t+\frac{\Delta t}{2})}{dt}$$

For $v(t)$ we obtain

$$v(t+\frac{\Delta t}{2}) = v(t-\frac{\Delta t}{2}) + a(t)\Delta t$$

From the above two expressions we see that for half a time step ($\Delta t/2$) we

evaluate velocity and then for the second half (Δt) we evaluate the new position. The new

position is calculated based on the current position at time t and velocity at time step Δt/2. The velocity at time step Δt/2 in turn is based on the previous velocity at time step −Δt/2 of the particle and the current acceleration of the particle at time t.

An important consideration that has to be made is the time interval Δt used for computation. If the time step is too large we will deviate from the true phase-space trajectory, since we take a step in the direction tangent to the true trajectory i.e. constant acceleration. However, if we take time steps that are too small, we will end up with very short trajectories through time that will sample too small a region of phase-space to be useful. Practically, a time step one or two orders of magnitude smaller than the highest frequency motion generally yields a sufficiently accurate trajectory. The highest frequency motion in atomic systems is the vibrational motion of bonds between a heavy atom and a hydrogen atom, which has a period of about 10^{-14} s. In practice, time steps of 0.5-1.0 fs are recommended for most purposes.
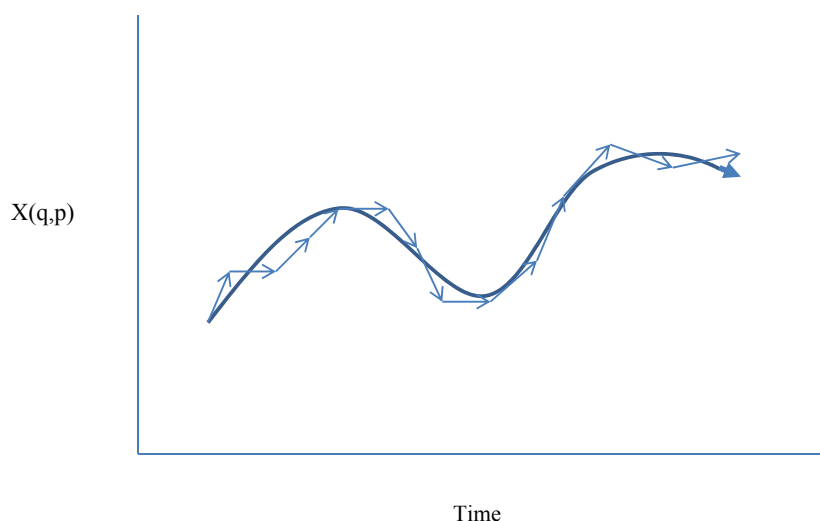
Figure 16: Trajectory in phase space

Another option to allow the use of larger time steps is to freeze the highest frequency motion. This can be done by constraining the heavy atom – hydrogen bond length. The keyword in CHARMM to constrain the heavy atom – hydrogen atom bond is called SHAKE. The next highest frequency motion would then be a heavy atom – heavy atom vibration with a frequency typically a factor of 2 – 5 smaller in magnitude. This approach can only be used if the property being studied is independent of those degrees of freedom which are removed. In addition to increasing the time step by a factor of two 2 (time step of 2 fs), reduction in the number of degree of freedom provides some saving in time and integration stability.

2.3.3. Periodic Boundary Conditions

A molecule being modeled is placed in a box of finite size. This model presents a technical difficulty. The size of the box cannot be very large because of the computational cost of including a large number of atoms is great, especially if the box is filled with water. If the box is too small then the surface effects of the box may affect the chemistry of the system. Beyond the edge of the box is essentially vacuum. Both these effects are unrealistic.

One way to handle this situation is to use Periodic Boundary Conditions (PBC). The system is placed in a unit cell, usually a cubic or truncated octahedral unit cell. The main unit cell is then surrounded by identical images. The long-distance non-bonded interaction energies are calculated between the unit cell and the images that are within the

cutoff distance. Each atom in the main unit cell has identical images surrounding the unit cell that follows the trajectory of the main unit cell particle. During the course of the MD simulation, if for example, a water molecule from the main unit cell moves out of the unit cell an identical image water molecule enters the main unit cell. This image water molecule then becomes part of the main unit cell when the atom list is updated. Therefore, PBC conserves mass, the number of particles and total energy. An important consideration is the size of the main unit cell. Ideally, the size of the main unit cell should be the total length of the molecule being modeled plus two times the largest cutoff distance used for non-bonded interactions. When the size of the unit cell is defined in this way, no two solute atoms will interact with each other directly and no solvent molecule will see two copies of a given solute atom.
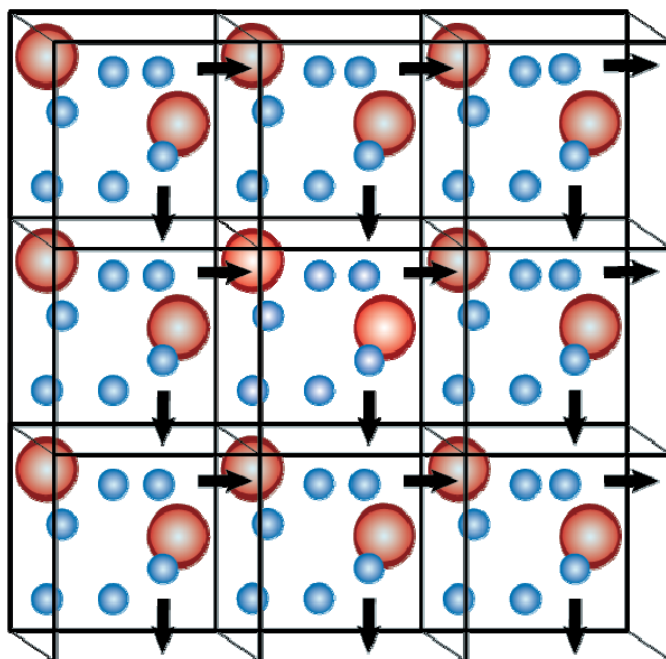
Figure 17: Schematic representation of periodic boundary condition
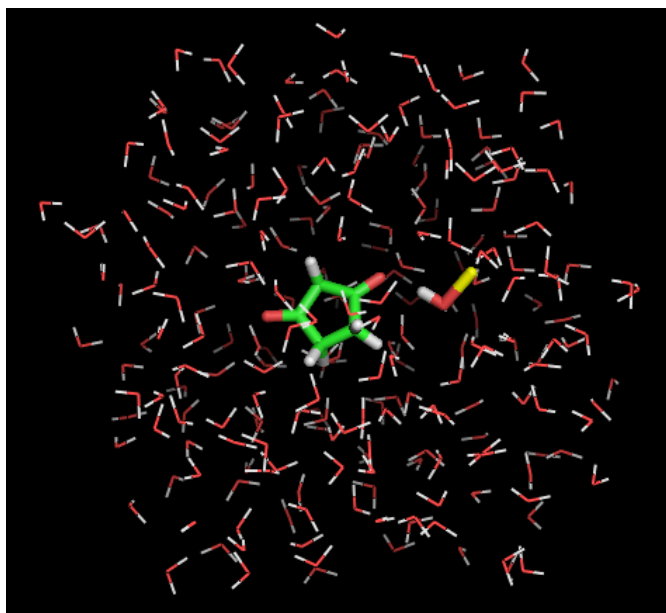
Figure 18: Solvated CPD and HOCl in a truncated octahedron box of water

2.4. Quantum Mechanics

To study reactions that involve the formation of new bonds or breaking of bonds, one needs to solve the equations of Quantum Mechanics (QM) applied to the electronic structure of molecules. The QM model is described by the time-independent Schrödinger equation:

$$\hat{H}\psi = E\psi$$

where ψ is the wavefunction describing the system, $\hat{H}$ is the Hamiltonian operator and E is the total energy of the system. The Hamiltonian operator for the time-independent Schrödinger equation is given by the following:

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} \nabla_i^2 + \sum_{i<j} \frac{e^e}{r_{ij}} + \sum_{k<l} \frac{e^e Z_k Z_l}{r_{il}}$$

where i and j label electrons, and k and l label nuclei. The first two terms describe the kinetic energy of electrons and nuclei, respectively, and the last three terms describe the potential energy of interaction electrons with nuclei, of electrons with each other, and of nuclei with each other, respectively. To simplify the above Hamiltonian the Born-Oppenheimer approximation is applied, which decouples the correlated motion of electrons with nuclei. The nuclei are assumed fixed in space while the energy of the electrons surrounding the nuclei is calculated. The above Hamiltonian is then simplified to the following (nuclear-nuclear repulsion energy can be added back in after the electronic energy is calculated):

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} \nabla_i^2 + \sum_{i<j} \frac{e^e}{r_{ij}}$$

Most approaches to solving the Schrödinger equation begin by expressing the wavefunction as a sum over basis functions:

$$\psi = \sum_{i=1}^N a_i \phi_i$$

where ϕ_i is a basis function and a_i is a coefficient. When the basis functions are appropriately chosen and spin is correctly accounted for, the correct solution for ψ can in principle be determined with an appropriate methodology.

2.4.1. Theoretical Methods

It is not practical to solve the Schrödinger equation exactly for any but the simplest of systems which include only one electron, for example a hydrogen molecule ion, because of the electron-electron repulsion term. Hence, there are several approximations made to solve systems of interest. First, the basis set expansion above must be truncated at a finite number of orbitals. In most approaches, the basis functions in the expansion above are themselves expanded further:

$$\phi_i = \sum_{j=1}^{M} a_j \varphi_j$$

The functions φ_j are termed atomic orbitals, and the functions ϕ_i are termed molecular orbitals. When either expansion is finite (which must be the case for practical applications), the solution will in principle be approximate. A molecular orbital constructed as a linear combination of atomic orbitals is called a linear combination of atomic orbitals (LCAO).

One of the most basic methods for solving the Schrödinger equation, the Hartree-Fock (HF) method, makes the approximation that each electron is influenced by an average

static potential of the remaining electrons in the system and an iterative 'self-consistent field' (SCF) method is used to determine the wavefunction. The HF molecular orbitals consist of Slater Determinant wavefunctions that include the exchange effects of the electrons. However, the HF method neglects most aspects of electron correlation, and this approximation may yield large errors in properties calculated from the wavefunctions. More sophisticated and robust methods need to be used, but HF often provides a practical stepping stone to a more exact solution.

There are a large number of methods for including more aspects of electron correlation than HF theory does. Some of the most widely used methods are Moller-Plesset (MPn) perturbation theory and Coupled Cluster (CC) theory. The MPn method depends on the perturbation order (n) correction used. However, these methods do not scale too well with the number of basis set and are computationally very expensive. For example, the CCSD (CC with single and double excitations) scales at the order of N^6 and MP2 (with second-order perturbation correction) scales at the order of N^5, where N is the number of basis sets. All of the methods described thus far are called ab initio methods because they do not make reference to any empirical parameters, and thus construct a solution from first principles.

Another alternative method is Density Functional Theory (DFT), which is different from the wavefunction theory discussed so far. The DFT approach accounts for electron correlation and offers equal or comparable accuracy to MP2 methods with much better computational efficiency. Scaling of DFT is of the order N^3, but for practical purposes (discussed later in this section) it scales to the order of N^4. Density Functional

Theory scales substantially better than the other methods., For these reasons, I use DFT for my calculations.

Density functional theory was first introduced in 1927 by Thomas and Fermi, but the approximations made were inaccurate. The theory was later legitimatized by the proof of two theorems, the Hohenberg-Kohn existence theorem and the Hohenberg-Kohn variational theorem in 1964. The most difficult part of DFT is not having an analytical form for the exchange correlation energy, and for this reason some approximations are made. One is to use parameters that are empirically optimized. Hence, DFT is sometimes referred to as a semi-empirical method, but in fact the functionals used are not parameterized to the extent that semi-empirical molecular orbital theory is. Density functional theory makes an approximation in order to calculate the energy density, a component of the exchange correlation energy. One form of the approximation, called local density approximation (LDA), treats the electron density as a uniform electron gas and is single valued at every position. Because of this approximation we can expect limitations in accuracy. One way to improve the exchange correlation functional is to allow it to depend not only on the local electron density but also on the extent on the gradient of the local density. This approximation is called the generalized gradient approximation (GGA).

The most popular GGA exchange functional is due to Becke (B); it incorporates a single empirical parameter. The most popular GGA correlation functional is LYP. The functional I use, B3LYP, is a hybrid functional that uses B3 with a three parameter scheme for the exchange functional and an adiabatic connection method (ACM) and LYP

for the correlation functional. ACM methods are called hybrid methods because they incorporate HF and DFT exchange terms.

I chose the B3LYP/6-31+G** level of theory and basis set because it has been shown to be sufficiently accurate for cases like my system, and it is sufficiently fast. The software packages I used for the QM calculations were Gaussian98 and Gaussian09 [56].

2.4.2. Basis sets

Choosing an appropriate basis set is very important for achieving an accurate solution to the electronic structure problem. In practice, the atomic orbitals in the second expansion above are themselves further expanded in terms of other functions. I will briefly discuss these. It is computationally efficient to represent atomic orbitals in terms of Gaussian type orbitals (GTOs) [46]. While GTOs are computationally convenient, they have issues with representing the radial portion of the orbital. The best way to deal with this drawback of GTOs is to form linear combination of GTOs. Basis sets are made up of linear combinations of primitive GTOs called contracted orbitals. Using primitive and contracted orbitals gives more flexibility in manipulating the coefficients so as to have a proper radial representation of the orbital. It is found that the most efficient (accuracy vs. computational efficiency) way to create a minimally realistic representation of an atomic orbital is in terms of three contracted orbitals. The nomenclature used to describe this kind of representation is "STO-3G basis set."

For the accuracy that I need, larger basis sets than the STO-3G are needed. The most important considerations for increasing the basis set size are described next. It is known that the valence orbitals play a much more major role in reactions than the core orbitals. Hence, there is more to be gained by having flexibility in the valence basis functions than in the core basis functions. This led to the concept of a 'split-valence' basis set, in which the core orbitals are represented by a single contracted basis function and the valence orbitals are represented by several contracted basis functions. I use the 6-31G basis set in my QM calculations. The '6' in the basis set representation indicates that the core orbitals are represented by one contracted orbital consisting of six primitive orbitals. The other two numbers, '31', indicate that the valence orbitals are represented by two contracted basis functions each. The first contracted basis function includes three primitive functions and the second contracted function includes one normalized primitive function. In addition polarization and diffuse functions can be augmented to the basis set, which are described below.

Polarization functions

Several molecular properties and especially geometries are sensitive to polarization effects. Atomic orbitals can be accurately represented with s, p functions, etc. This is because atomic orbitals, like these functions, are centered about the atom nucleus. Molecular orbitals, on the other hand, are often extended in space and are expanded in terms of LCAOs with several atom nuclei. For this reason it is necessary for the molecular orbital to have more flexibility than atomic orbitals.

One way to provide this flexibility is to include an additional quantum number of higher angular momentum than the valence orbitals in the form of a GTO basis function. This means that elements in row two like carbon and oxygen. have polarization functions that include d-type GTOs; for hydrogen atoms the polarization functions includes p-type GTOs. The nomenclature used in this case is either a '*' or, equivalently, (d) at the end of the basis set specification for row two heavy atoms; and '**' or, equivalently, (d,p) when polarization functions are to be used for both second row atoms and hydrogen atoms. In my calculations I use the following basis set specification and polarization functions: 6-31G**.

Diffuse functions

Electrons in some molecular complexes tend to be more spatially diffuse than in atomic orbitals. To provide the flexibility needed to represent this spatial diffuseness, the basis function is augmented with diffuse functions when needed. The nomenclature used to indicate the addition of diffuse functions is a '+' added to the basis set specification for additional heavy atom s and p functions and a '++' to indicate diffusion functions of s type for hydrogen atoms in addition to heavy atom diffuse functions. In my calculations I use diffuse function for only heavy atoms; 6-31+G**.

2.5. *Quantum Mechanics / Molecular Mechanics*

As described in previous chapters MM is useful for modeling large systems. However, it cannot be used to model reactions that involve bond-making or bond-forming. Instead we use a QM model to model chemical reactions. The disadvantage to the QM model is that it does not scale well as the number of heavy atoms increases and therefore the limiting factor is the size of the system. We are restricted to using the QM model for only small systems, or for reactions of large systems, one has to ignore atoms not near the reaction site. In many cases, though, these atoms may significantly influence the reaction even though their role may not be as important as the ones near the reaction site. These atoms may influence the reaction via long range non-bonded interactions (electrostatic or van der Waals) with the reactive atoms or, in the case of macromolecules, by influencing the conformation of the system as a whole.

MM

QM
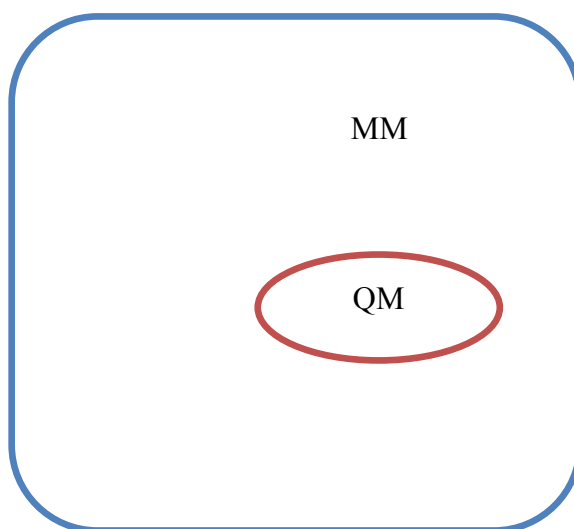
Figure 19: Schematic of QM and MM regions

Fortunately, larger systems can often be treated by using a combination of the MM and QM approaches. This kind of a hybrid method is denoted QM/MM. One can model reactive site atoms that are deemed crucial to the reaction using QM. The other atoms of the large system can then be modeled using MM instead of ignoring them. The development and application of QM/MM methods have grown rapidly over the past few years.

For a QM/MM model, the total Hamiltonian is written as

$$H_{Total} = H_{QM} + H_{MM} + H_{QM/MM}$$

This equation has a deceptively simple appearance. The details of $H_{QM/MM}$ can become quite complicated, depending on the system. The range of factors that can arise is described below, but as in earlier chapters emphasis is given to factors arising in the research reported here.

The first thing that comes to mind when we consider QM/MM systems is the boundary between the QM section and MM section and whether the boundary intersects a bond or not. It is a little simpler if the boundary does not intersect a bond. If it does, there are a few ways it is handled. One way is to use a link atom, usually hydrogen, to cap off the dangling bond in the QM section. Another way is to use a frozen orbital. Sometimes a different methodology, ONIOMM, is used. In my system, the boundary does not intersect any bond, so I did not have to deal with any of these issues.

Even though my system does not have to deal with the issue of the QM/MM boundary severing bonds, there are still other matters that need to be considered and

choices to be made. One is the coupling or polarization between the two regions. There are three possible interactions. Both regions can be unpolarized, only the QM region is polarized and the third option is having both regions polarized.

My model, incorporates polarization of the QM region by the MM region, but not polarization of the MM region. The MM atoms are represented as fixed point charges. The Hamiltonian for such a system is

$$H_{QM/MM} = \sum_i^{solute\ electrons} \sum_m^{MM\ atoms} \frac{q_m}{r_{im}} + \sum_k^{solute\ nuclei} \sum_m^{MM\ atoms} [\frac{Z_k q_m}{r_{km}} + 4\varepsilon_{km}(\frac{\sigma_{km}^{12}}{r_{km}^{12}} - \frac{\sigma_{km}^6}{r_{km}^6})]$$

Another issue that needs to be considered is the interaction of atoms in both regions that are very close to the boundary. MM model represents atoms as point charges, while the QM atoms have electronic and nuclear representation of atoms. The MM point charges near the boundary can have large unrealistic influences on the electronic representation of the QM atoms near the boundary. They can distort the QM electron density and create forces acting on QM nuclei. One way to deal with this situation is to blur the point charges in the MM region that are close to the boundary by Gaussian functions so that the charges are spread or 'blurred' and do not have such drastic, unphysical effects on QM atoms close to the boundary. The CHARMM program includes such an option, and I used it in the work reported here.

The QM software I use to interface with CHARMM to implement the hybrid QM/MM model is Q-Chem 3.2 [57].

2.6. Chain-of-replica Methods

A reaction mechanism is better understood by studying its reaction path. A reaction path is the relative motion taken by atoms during a chemical reaction. We usually think of a reaction path as beginning with reactants, proceeding through a high-energy structure called a transition state, and terminating in products. In complicated cases, there may also be significant intervening structures called intermediates. For any chemical reaction, there are many possible reaction paths. Particular significance is given to that reaction path that passes through the lowest-energy transition state and for which the path from the transition state toward either reactants or products follows the direction of steepest descent; this is called the Minimum Energy Path (MEP) or Steepest Descent (SD) path. The significance of the MEP is that in a macroscopic system of many reacting molecules, the vast majority of reactive events follow a trajectory close to the MEP, and consequently the energetics of the MEP dominates the kinetics. Thus the MEP is often predictive of rate constants and reaction mechanisms, and so most analyses are based on the MEP. In mathematical terms, reactants, products and transition states (TS), and also intermediates if they exist, are stationary points because for them the gradient of the potential energy surface is zero. Clearly, the development of methods to identify MEPs, and especially the stationary points, is very important in understanding reaction mechanisms.

There are broadly three classifications of methods used to identify the reaction path. The first are eigenvector following methods, of which the most popular method is the Intrinsic Reaction Coordinate (IRC) method [58, 59]. This method relies on first

identifying the transition state, for which by definition the Hessian possesses exactly one negative eigenvalue, and then following the eigenvector corresponding to the negative eigenvalue in the forward and reverse directions. This process eventually leads to the reactant and product structures. While it is robust, there are several drawbacks to this method. It is sometimes very difficult to find the TS structure if the potential energy surface is complex or the saddle point is quite flat. Locating the TS is also a very expensive method in terms of scaling. Memory scales as N^2 and time scales as N^3, where N is the number of degrees of freedom. The size of the system restricts the use of this method. It is not practical to use in large systems such as the one studied here.

Another method which is quite intuitive is the reaction coordinate following method. If we know the reaction coordinate that drives a particular reaction, we just follow it and it should connect the three stationary points. The reaction coordinate following method, however, is impractical for all but the simplest of systems, since it is very difficult to know the reaction coordinate in advance, especially for reaction mechanisms that involve large numbers of atoms. Another issue with this method is that one needs to check the hysteresis to make certain that the same path is mapped out in the forward and reverse directions.

The method I use is from a group of methods called chain-of-replica methods [60-71] that are fairly new compared to the other methods. Its development is quite significant because it allows one to study reaction mechanisms of large systems, and in most cases only first derivatives need to be calculated. In some cases the results can be refined using a second derivative method for geometry optimization, but even then, chain-of-replica methods are more feasible for large systems than are other methods.

Another significant advantage of chain-of-replica methods is that one only needs the reactant and product structures as input, whereas the IRC method also requires a fully optimized transition state. In fact, chain-of-replica methods are efficient for locating transition states in high-dimensional systems where the usual transition-state geometry optimization methods may be impractical. There is also no need to know the reaction coordinate in advance which can be a very difficult task with a system having many degrees of freedom. A further benefit of this method is that it can take advantage of today's parallel computers. Since individual points on the path are only weakly coupled, they can be refined on different processors. The other methods are sequential and cannot make use of parallel computers as efficiently.

The chain-of-replica method was originally proposed by Elber and Karplus [67]. It was based on work by Empedocles who in 1969 had proposed using a line integral for searching a reaction coordinate.

$$S = \frac{1}{L} \int_{R_A}^{R_B} V(R) dl(R)$$

where S is the function to be optimized, R_A and R_B are position vectors representing the coordinates of reactant and product structures, $V(R)$ is the potential energy and $dl(R)$ is the line element along the length of the path, L, between R_A and R_B. $V(R)$ and $dl(R)$ are functions of the N-dimensional position vector representing coordinates, where N is the number of degrees of freedom.

Elber and Karplus approximated the above line integral by a discretized version.

$$S = \frac{1}{L}\sum_{j=0}^{M} V(R_j)\Delta l_j + \text{ constraints}$$

where M is the number of intermediate discretized grid points and Δl_j is the interval between each point. Constraints were included to keep the individual points equidistant from each other. This is because, during the process of minimization, if the Δl_j were widely different the equation above would no longer be a good approximation of the line integral. Later additional repulsion constraints were added so that the points did not all pool into a local minimum. In the methods I used that are described below, the distance Δl between coordinate sets representing points on the path is defined as the root mean square difference (RMSD) between the atomic coordinates of the two sets. The definition of Δl includes overall translation and rotation of the coordinate sets to achieve a best fit in a least squares sense. The total length of the path, L, is the sum of the distances Δl between points on the path so defined. Thus, L is a function defined in RMSD space, $L(r_1, r_2, \ldots, r_{M-1})$, where r_i is an RMSD of coordinates just defined..

The formalism described above forms the basis for all chain-of-replica methods. Individual methods differ mainly in the way the points are generated, the minimization protocol used, and most importantly the kind of constraints imposed to control the optimization.

The methods I used for include the Replica Path method and the Nudged Elastic Band (NEB) method with the Climbing Image algorithm to obtain the TS structure (CI-NEB). I also used the Adaptive Nudged Elastic Band (ANEB) method for a single test

case. All these methods used are implemented in CHARMM, the software I used for my calculations.

2.6.1. Replica Path Method

The easiest and most common way to create initial replicas or images is by linear interpolation of the reactant and product structures. The replica path method, like most modern chain-of replica methods, only replicates certain atoms that are deemed important in the reaction being studied instead of the entire system. The un-replicated part of the system is still present and identical across all replicas. The un-replicated part can and will affect the replicas and respond, if needed, but it will do so the same way for all replicas along the MEP. This, quite conveniently, sets up for use with the QM/MM model. It allows the focus of the study and resources on the active site of the molecule. In the geometry optimization calculation of the replicas the un-replicated atoms were kept fixed.

The main aspect of replica path that sets it apart from the other chain-of-replica methods is that the method uses a harmonic penalty function to constrain/restrain the points on the reaction path. The penalty function applied is directly proportional to the root mean square deviation (RMSD) in the atomic coordinates of two neighboring replicas or images. During the evaluation of the RMSD the deviation because of rotation and translation are avoided. An advantage of using this RMSD space to define the restraints is that there is no need to deal with effects of rigid body rotation and translation which would otherwise affect the minimization process.

Below are the penalty functions and other keywords used with replica path as implemented in CHARMM.

$$E_{rms} = \frac{1}{2} \sum_{i=1}^{N} K_{rms} (r_i - \bar{r})^2$$

where E_{rms} is the penalty function used to force the replicas or images to be approximately equidistant from each other. N is the number of replicas or images along the pathway. The optimal value of the force constant K_{rms} should result in a path step variation of no greater than 15 – 20 %. The distance between path points i and i+1 in RMSD-space is denoted r_i, and the average over all pairs of consecutive path points is $\bar{r}$:

$$r_i = RMSD_{best\ fit}(i, i+1)$$

$$\bar{r} = \sum_{i=1}^{N} \frac{r_i}{N}$$

One artifactual difficulty sometimes encountered is a tendency for the reaction path to become highly curved so that points fall into each other, resulting in small values of the constraint energy but giving unrealistic paths. To avoid this problem, an additional restraint, E_{angle}, is used to keep the reaction path realistically linear:

$$E_{angle} = \frac{1}{2}\sum_{i=1}^{N} K_{angle}[COSMAX - \cos(\Theta)_i]^2 \text{ if } COSMAX > \cos(\Theta)_i$$

$$E_{angle} = 0 \qquad\qquad\qquad \text{if } COSMAX \leq \cos(\Theta)_i$$

where Θ is the $(i-1) - i - (i+1)$ angle in RMSD-space. COSMAX is a value between 0.9 and 0.98, with the value 0.98 making the pathway more linear and rigid. I use the value 0.9 in my calculations. K_{angle} is the force constant used for the restraint. The recommended range is $100 - 500$ kcal/mol. I used the value 100 kcal/mol in the calculations reported here. Below are the definitions of the most important keywords used in the CHARMM's replica path facility.

RMAX: Maximum distance the replicas are allowed to be separated before switching on the force constant KMAX instead of K_{rms} in the penalty function E_{rms}.

KMAX: The force constant to be used when the distance between replicas exceeds RMAX and is usually greater than or equal to K_{rms}.

WEIGHT: This keyword causes the values in the main weighting array are to be used. This allows one to place more emphasis on atoms that play a major role in the reaction, so that when the RMSD is calculated to constraint the replicas more emphasis is placed on those atoms.

ROTATION and TRANSLATION: Keywords that turn on the RMSD best-fit procedure between neighboring replicas, thus preventing the rigid-body rotations and translations from effecting the minimization. Without the RMSD best-fit procedure the atoms would adjust unnecessarily by rotating and/or translating in order to meet the constraint requirement.

BLOCK: The utility in CHARMM that calculates the total potential energy at each point along the minimum energy path. The total potential energy is the sum of the interaction energy between atoms in the un-replicated region plus the interaction energy between atoms in an individual replica plus the interaction energy between atoms in a single replica and the atoms in the un-replicated region. The BLOCK utility has to be appropriately setup so that the inter-replica interactions are ignored and the interaction matrix is appropriately scaled. The scaling factor for the interaction energy between atoms in the un-replicated region is set at 1.0. The scaling factor for the interaction energy between atoms in an individual replica is the inverse of the number of replicas created. The scaling factor for the interaction energy between atoms in a single replica and the atoms in the un-replicated region is the inverse of the number of replicas created.

2.6.2. Nudged Elastic Band Method

The NEB method is similar to the replica path method and uses the same keywords in CHARMM. The implementation, however, is different. The NEB method

was originally proposed to address a concern about replica path. In replica path, the physical interatomic interactions and the extra spring restraints can interact, potentially causing deviations from the true MEP. Such deviations are especially likely to occur in steep regions of the reaction path, such as are often encountered near a transition state. While replica path method implements the penalty function based on the RMSD as described above, NEB uses a force projection scheme in which the physical force due to interatomic interactions acts only perpendicular to the path, and spring-constraint forces act only along the path. The spring forces keep the replicas separated from each other and the potential forces acting perpendicular to the path 'nudge' the pathway closer to the MEP. This is accomplished with the following set of equations.

$$F_i^{NEB} = F_i^{\perp} + F_i^{S\|}$$

[60]
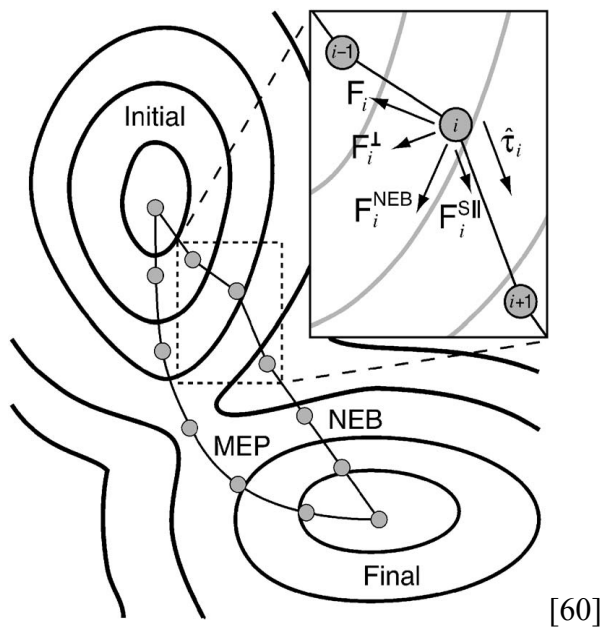
Figure 20: NEB

where $F_i^\perp$ is the component of the force due to the potential perpendicular to the band given by

$$F_i^\perp = -\nabla(R_i) + \nabla(R_i) \bullet \hat{\tau}_i \hat{\tau}_i$$

where $\hat{\tau}_i$ is the unit vector in the tangential direction of the higher energy neighboring replica, and $F_i^{S\|}$ is the spring force parallel to the band and is given by

$$F_i^{S\|} = k(|R_{i+1} - R_i| - |R_i - R_{i-1}|)\hat{\tau}_i$$

2.6.3. Climbing Image – Nudged Elastic Band Method

A problem often encountered in the methods described above is that the actual transition state may be difficult to reach, and the resolution of replicas near the transitions state may be poor. Poor resolution is particularly problematical if the energy barrier is narrow in comparison to the length of the minimum energy path. What is usually noticed in such situations is that the replicas tend to slide away from the transition state and cause 'corner-cutting' around the transition state.

The climbing image algorithm (CI-NEB) is a modification of the NEB method. The main difference is the treatment of the replica that has the highest energy in the pathway after a few minimization steps. That point is not subjected to the spring forces

and is driven towards the transition state instead of being restricted by the spring forces. The total force acting on that particular point is given by

$$F_{i_{max}} = -\nabla E(R_{i_{max}}) + 2\nabla E(R_{i_{max}})|_{\parallel}$$

$$F_{i_{max}} = -\nabla E(R_{i_{max}}) + 2\nabla E(R_{i_{max}}) \bullet \hat{\tau}_{i_{max}} \hat{\tau}_{i_{max}}$$

This modification to the NEB does not add any additional computational time. The replica moves along the elastic band up the potential energy surface to the transition state.

2.6.4. Adaptive Nudged Elastic Band Method

The ANEB method [61] is a variation of the NEB method. The traditional NEB method can sometimes have difficulty approaching the transition state structure due to poor convergence on difficult and flat potential energy surfaces. The focus of the ANEB method is to increase the resolution of points or replicas near the transition state and to acquire a good estimate of the reaction barrier. Just like the regular NEB method the initial setup requires two local minima representing the reactant and product structures. The initial replicas or images may be created by linear interpolation as in the NEB method and minimized until a reasonable convergence criterion is met. At this point ANEB deviates from the regular NEB method. New fixed points are selected instead of the reactant and product structures in the previous step. The new fixed points selected are

the ones on either side of the point or replica with the maximum energy. Presumably, these points encompass the actual transition state. New images are then created between these points as in regular NEB and minimized. This procedure is continued until the transition state is found. Hence, the resolution of points or replicas is adaptively increased near the transition state. Another benefit of the ANEB method is that fewer replicas are needed at each iteration.

[61]
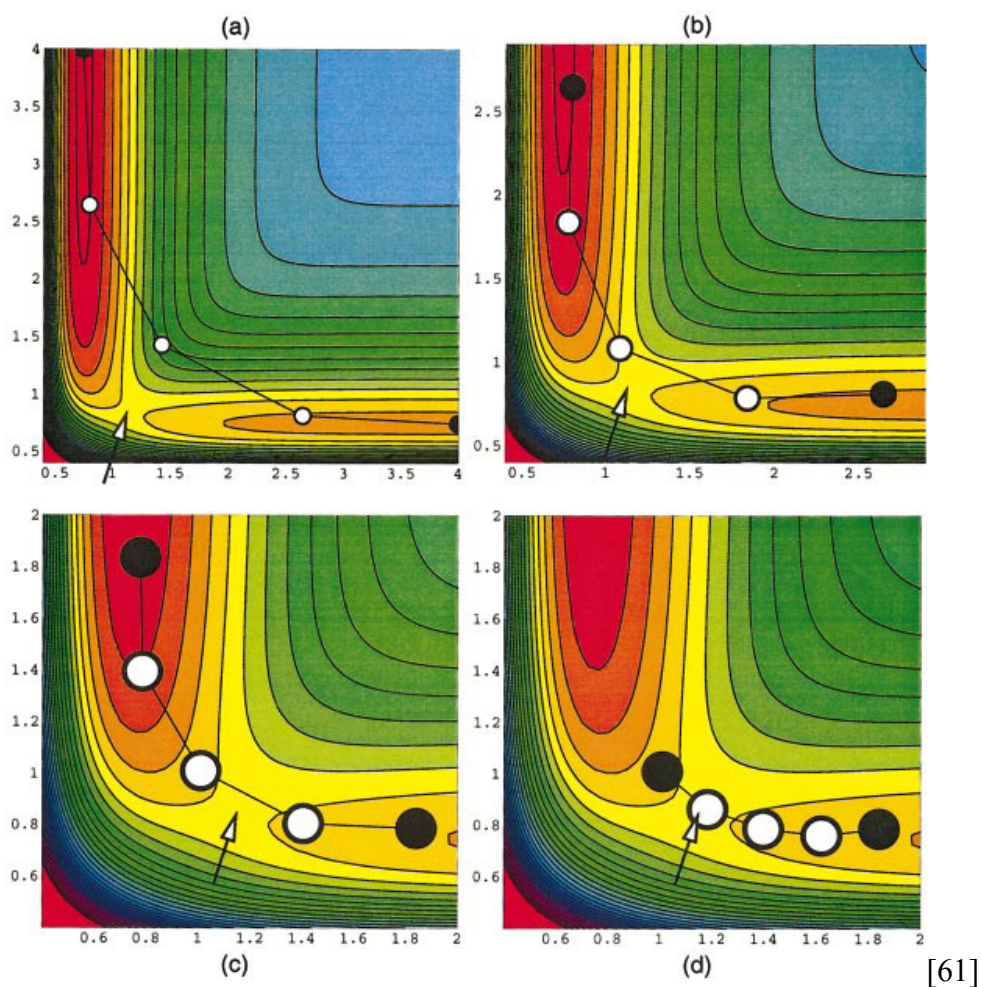
Figure 21: ANEB

3. PARAMETERIZATION

3.1. *Principles*

As discussed earlier in the Molecular Dynamics and QM/MM sections, programs running MD calculations require parameters that need to be read at the beginning of the script. The default parameters provided by the software are adequate for most systems or molecules. However, in some rare cases one has to develop one's own set of parameters if the software does not provide parameters for the molecule being studied.

Parameters are provided in two files called the topology file and the parameter file. The topology file contains information such as the names of molecules, atom names and atom types (defined below) partial charges on individual atoms, atom connectivity (bonds), and if present, improper dihedral angles, acceptor or donor atoms, internal coordinates and patch information (primarily need for protein molecules). The molecule name is used by the script to build or make reference to the molecule and hence has to be unique with a maximum of four characters. The connectivity information lets the software know which atoms are connected and if they share a single, double or triple bond. Acceptor or donor atom information is used for hydrogen bond analysis. A data structure for internal coordinates is set up to reserve memory in case internal coordinate manipulations are needed or in order to build Cartesian coordinates from internal coordinates. Patch information is used to connect fragments or segments and in some cases it tells the program how to handle a residue at the ends of a segment.

The parameter file contains information such as the atom type, Van der Waals constants, force constants for bonds, angles, dihedrals and impropers and their respective equilibrium values. This information is used in energy calculations as described earlier. In cases where there are no parameters available for the system or molecule(s) being studied one has to develop one's own parameters. Depending on the software being used there are different protocols. One cannot mix and match parameters from different types of software. For example, one cannot use some parameters from Amber and some from CHARMM. All the parameters have to be from either Amber or all from CHARMM for internal consistency. Hence, it is very important to be consistent in using the correct protocol. Alexander MacKerell has become the primary developer of parameters for the CHARMM suite of force fields. I use the CHARMM parameter set and therefore follow the protocol outlined by MacKerell et al. to develop the parameters for my system [72].

There are two ways to go about developing parameters for atoms or molecules for use in CHARMM. The first and simplest way is called "by analogy." One identifies fragments of the molecule being parameterized that are similar to fragments of molecules already present in the default CHARMM parameter files. If it is possible to identify a complete set of fragments that when combined yield the complete molecule, only a few parameters need to be developed. These are for connecting the individual fragments. The analogy method cannot be used if the default parameter set lacks the needed fragments. In that case, one must develop all parameters for the entire molecule from the ground up. If the parameter set contains only a few of the needed fragments, it is generally observed that developing all parameters from the ground up yields the best parameterization.

Below is a flowchart showing the process by which parameters are developed.
Following it, the procedures used to develop parameters for the systems of interest to us
are described in detail.
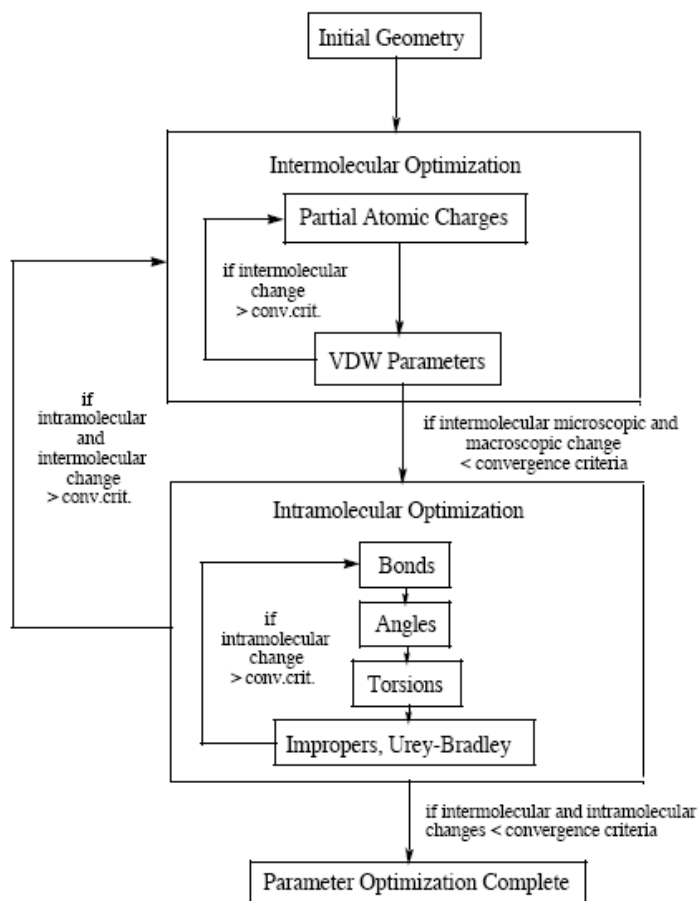
[72]

Figure 22: Parameterization protocol flow chart
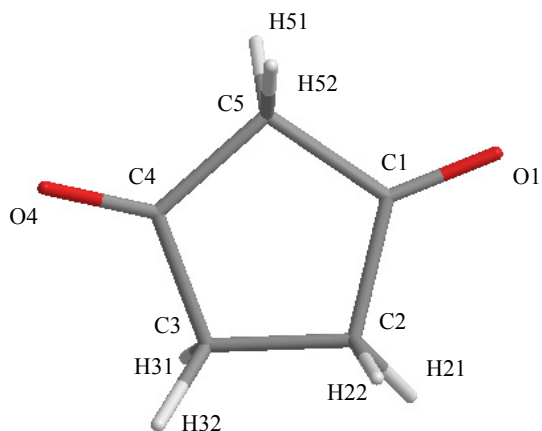
Figure 23: Cyclopentanedione (CPD). Atom names are shown

The molecule CPD is directly involved in the chlorination reaction being studied. Therefore highly accurate parameters are needed and I choose to develop all of them from the ground up. I only need to develop parameters for the diketo form of CPD and not the ketoenol form, since I only need the parameters to perform MD simulations involving the diketo form to identify "unique encounter reaction precursors," as explained in chapter 5. I will actually begin with a set of parameters determined by analogy, but I subsequently refine them. First, though, I must define a few terms.

3.1.1. Atom type and atom name

In Figure 23, CPD is shown with a unique "atom name" for each atom. In force fields, in addition to a unique "atom name," each atom is also assigned an "atom type." "Atom type" is a technical term that groups atoms according to properties. Parameters are defined in terms of atom types, not in terms of elements or individual atoms (i.e.

"atom names"). Thus the first step in parameter development is to decide which set of atoms will be assigned to the same atom type. A given element may, in different contexts, have quite different properties. Therefore, for example, an oxygen in a keto group is assigned to one atom type, while an oxygen in a hydroxyl group is assigned to another. In principle, almost every environment is unique, and so one could define a very large number of atom types. This would make force fields extraordinarily complicated and costly. In practice, a middle ground is sought in which atoms in quite different environments are assigned different atom types, while those in similar environments are assigned the same.

In the case of CPD, it is reasonable to identify two carbon atom types, one oxygen atom type and one hydrogen atom type. I look for similar atom types in the CHARMM general force field parameter file. The carbon, oxygen and hydrogen atom types are similar to their respective atoms type in 2-pyrrolidinone - C1 and C4 of type CG2R53; C2, C3 and C5 of type CG3C52; O1 and O4 of type OG2D1; H21, H22, H31, H32, H51, H52 of type HGA2. The VDW parameters are not expected to change significantly and so I use the same values without refinement, as recommended by Mackerel.[72] The reason I look for similar atom types already present in the default parameter files is to avoid creating new atom types in the original topology and parameter files if possible. I could just as easily create completely new atom types since I modify all the parameters (except the VDW parameters since it is not expected it to change significantly) for CPD. It is worth noting that some thought must be given while deciding whether it is best to create new atom types or using existing atom types. If one uses existing atom types, like I did, then the new parameters will overwrite old parameters, if they already exist, in the

parameter file. So, care must be taken that one does not intend using the old parameters for other molecules. In this case, some parameters are in fact overwritten, but in the context of my work I do not require the old parameters so it is safe for us to use the old atom types.

Atom Name	Atom Type	Description
C1	CG2R53	5-membered ring carbon bound to a keto oxygen, e.g. in 2-pyrrolidinone
O1	OG2D1	carbonyl oxygen; e.g. carboxylic acids
C2	CG3C52	5-membered ring carbon bound to aliphatic hydrogen e.g. in proline
H21	HGA2	aliphatic hydrogen
H22	HGA2	aliphatic hydrogen
C3	CG3C52	5-membered ring carbon bound to aliphatic hydrogen e.g. in proline
H31	HGA2	aliphatic hydrogen
H32	HGA2	aliphatic hydrogen
C4	CG2R53	5-membered ring carbon bound to a keto oxygen, e.g. in 2-pyrrolidinone
O4	OG2D1	carbonyl oxygen; e.g. in carboxylic acids

C5	CG3C52	5-membered ring carbon bound to aliphatic hydrogen e.g. in proline
H51	HGA2	aliphatic hydrogen
H52	HGA2	aliphatic hydrogen

Table 1: CPD Atom Names and Atom Types

3.1.2. Partial atomic charges

The next step in the protocol is to assign initial guess partial charges to the different atoms. Each atom may in principle be assigned a different partial charge (i.e. the atom type does not determine the partial charge), but in general atoms in similar environments are assigned the same charge. Following the protocol of MacKerell et al. I start by assigning a +0.09 charge to all hydrogen atoms. The next step is to assign charges to the other atoms so that the total charge is zero, since the system is neutral. I determine an initial guess by calculating the charges using a quantum mechanical method. MacKerell et al. recommend using the HF/6-31G* level of theory to optimize the geometries and electronic wave functions of all molecules. This provides internal consistency for CHARMM parameter development. They recommend using the Merz-Kollman method [73]for determining the partial atomic charges. The Merz-Kollman method derives atomic charges from a least-squares fit to the electrostatic potential (ESP) calculated at a large number of points around the molecule of interest. The Merz-Kollman method differs from similar methods in the choice of points where the ESP is calculated.

The MK scheme uses grid points on several layers around the molecule with different scaling factors for each layer. Keeping in mind the symmetry of the molecule and the fact that aliphatic hydrogen atoms need to have a +0.09 charge, I assign initial guess charges for all the atoms.

The next step is calculating target data for use in optimizing the partial atomic charges. These are obtained from quantum mechanical calculations. Two types of target data are calculated. One target datum is the dipole moment. The others pertain to the interaction between water molecules and CPD. In particular, the interaction energy and the optimal interaction distance are evaluated. An evaluation is made for each of the solvent-exposed atoms of CPD. This is accomplished by placing a single water molecule in close proximity to the atom under consideration, and performing the calculations described below. The partial atomic charges are then adjusted iteratively until the MM calculations well reproduce the dipole moment and water interactions obtained from QM calculations.
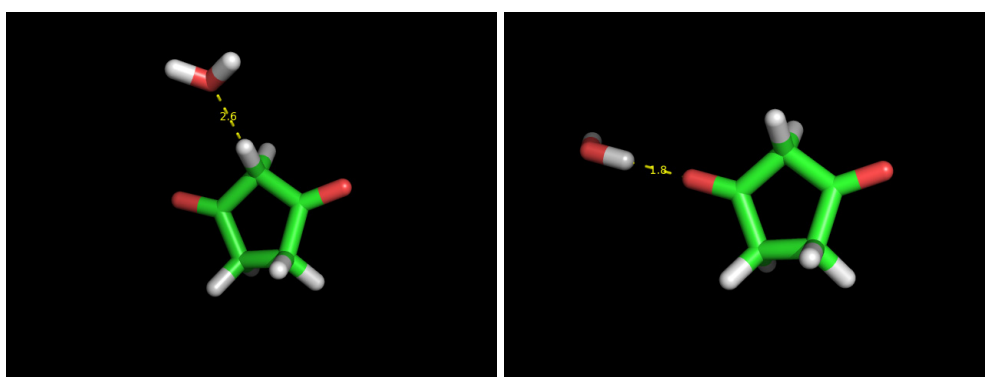
Figure 24: Two example representations of water interacting with atoms in CPD

I begin with two structures; one optimized using QM and the other optimized using the CHARMM program for 200 minimization steps using the Conjugant Gradient

algorithm and another 50 minimization steps using the Newton-Raphson algorithm. The QM optimized structure is copied in the comparison coordinate set and then the CHARMM optimized structure is aligned or oriented so that the dipole moments can be calculated and compared.

To get the other target data for water interaction energies and optimal hydrogen bond distances I first read in the QM optimized structure and build a TIP3P water molecule 1.6 Å from one of the atoms. The TIP3P water molecule is a classical water model that I use in the solvation study. The energy is set at an arbitrary maximum value of 1000 kcal/mol. The distance between the water molecule and the atom is gradually decreased by 0.01Å and the corresponding energy calculated. This is done until the minimum interaction energy is located. I follow the procedure for every periphery atom, with a single water molecule, one at a time. This data is then compared to the target data which I obtain from the QM calculation.

As mentioned above, the CHARMM General Force Field is an additive force field. One of the inherent disadvantages to an additive force field is the way electrostatic interactions are calculated. Coulomb's law is used to calculate the electrostatic interactions and uses fixed point charges on atoms to do so. Using fixed point charges is a major simplification. It is well known that the electron density or distribution around an atom and in this case atomic charges, changes as a function of the electrostatic field around the molecule. So, as the atoms move during simulation, their atomic charge should change or respond to its environment. This is ignored in an additive force field. To compensate for the lack of polarizability the atomic charges are 'tweaked' or 'enhanced', especially for molecules that are polar and in condensed phase. CPD is a polar molecule;

hence, the target data acquired from QM calculations are 'enhanced'. One of the major developers of the CHARMM force fields, MacKerell et al., has outlined the modifications that need to be made to the target data. I follow these recommendations and make the following changes. The target interaction energies for the QM optimized structure are scaled by a factor of 1.16. The target optimal hydrogen bond distances are increased by about 0.2 Å. The target dipole moment is scaled by 0.8-0.9.

The only parameters I can change at this step are the partial atomic charges on CPD. CPD is a symmetric molecule so we expect atoms like C1 and C4 to have the same charge. Likewise, C2 and C3 would have the same charge and also O1 and O4. The aliphatic hydrogen atoms need to have an atomic charge of +0.9, as described earlier. Therefore, there are only four atomic charges I am able to manipulate, including C1 (or C4), C2 (or C3), O1 (or O4) and C5. Minor changes were made to these parameters and then the empirical data was compared to the target data. Several such iterations were done to approach the target data. The approach used to achieve this goal was to have the least possible deviation from the different types of target data like the dipole moment, water interaction energies and their respective minimum energy distance for all the atoms in CPD.

Ideally, the recommended level of agreement, for water interaction energies should be within 0.2 kcal/mol of the target values. The dipole moment should be overestimated by 20 to 50% in comparison to the QM calculated dipole moment. The best partial charges I was able to obtain are listed in Table 2.

Atom Name	MK	Initial CHARMM	Final CHARMM
C1	0.56	0.55	0.48
C2	-0.18	-0.18	-0.19
C3	-0.16	-0.18	-0.19
C4	0.55	0.55	0.48
C5	-0.45	-0.36	-0.2
O1	-0.46	-0.46	-0.46
O4	-0.46	-0.46	-0.46
H21	0.07	0.09	0.09
H22	0.08	0.09	0.09
H31	0.07	0.09	0.09
H32	0.07	0.09	0.09
H51	0.15	0.09	0.09
H52	0.15	0.09	0.09

Table 2: CPD Partial atomic charge

The interaction energies with water molecules, the respective hydrogen bond distances with the water molecule and the dipole moments are listed in Table 3. Note: the QM energies are scaled by 1.16.

	QM (kcal/mol)	CHARMM (kcal/mol)	Δ (CHARMM – QM) (kcal/mol)
O1 - HOH	-4.1479	-4.62955	-0.48165
O4 - HOH	-4.1479	-4.6454	-0.4975
H21 - OHH	-3.0084	-2.15067	0.85773
H22 - OHH	-2.5196	-1.68288	0.83672
H31 - OHH	-2.5196	-1.68288	0.83672
H32 - OHH	-3.0084	-2.15067	0.85773
H51 - OHH	-3.29	-1.84931	1.44069
H52 - OHH	-3.29	-1.84931	1.44069
		AVG Difference	0.661391
		RMS Difference	0.707352

Table 3: CPD – H₂O Interaction Energies

	QM (Å)	CHARMM (Å)	Δ (CHARMM – QM) (Å)
O1 - HOH	2.11	1.83	-0.28
O4 - HOH	2.11	1.83	-0.28
H21 - OHH	2.47	2.64	0.17
H22 - OHH	2.47	2.66	0.19
H31 - OHH	2.47	2.66	0.19
H32 - OHH	2.47	2.64	0.17
H51 - OHH	2.34	2.63	0.29
H52 - OHH	2.34	2.63	0.29

Table 4: CPD – H$_2$O Interaction Distances

QM (D)	CHARMM (D)
1.8625	1.8152

Table 5: Dipole moments

3.1.3. Bond and angle terms

The next step is the optimization of intra-molecular bond and angle parameters. The functional forms for these terms are shown below.

$$\sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2$$

Bond and angle parameters are optimized at the same time. There are two values needed for bond and angle parameters, the equilibrium values and the force constants. The equilibrium values can easily be obtained from a corresponding QM optimization of the molecule. The MP2/6-31G* level of theory is recommended for this purpose. Ideally, deviations of up to 0.03 Å for equilibrium bond distances and up to 3° for equilibrium angles in comparison to QM calculations is acceptable. The optimization of the force constants, however, is quite involved and was possibly the longest and most tedious part of the optimization process for CPD. CHARMM uses a module called MOLVIB, which is a general-purpose vibrational analysis program used for small molecules that contain fewer than 50 atoms.

The MOLVIB module is very structured and the syntax has to be closely followed. The MOLVIB module is also used to get the target data, but the target force matrix for the vibrational analysis is obtained from a QM calculation of a frequency calculation. The force matrix from the QM (Gaussian98) output is then copied into the MOLVIB with the target frequencies scaled by a factor of 0.89.

The force constants are adjusted so that they best match the target data. The goal is to reproduce the frequencies and the contributions of the internal degrees of freedom. Lower frequencies represent motions involving heavy atoms and, therefore, motion that deviates the most from its optimal geometry. Hence, it is a priority to reproduce the low frequencies more accurately. The high frequency motion that represent stretching motion involving hydrogen atom are trivial and in most cases not important. It is preferable to produce vibrational frequencies that are slightly lower than the target values. The reasoning is that it is better to have a more flexible molecule instead of a more rigid molecule.

As a rough guideline the empirical frequencies calculated should be, on average, within 5% of the corresponding QM calculated frequencies. "Apart from these basic guidelines, it is difficult to objectively quantify the quality of the parameterization." "It is often difficult to unambiguously correlate the QM and MM frequencies due to multiple local internal valence coordinate contributions to each frequency. These contributions will often differ between the QM and MM spectra." [72].

```
Frequencies (cm⁻¹) and % contribution (from QM calculated force constants)


     Freq              %               %               %
1    33.8    t-5RNGsy  84.
2    95.9    t-5RNGas  83.
3    370.0   d-C1=O1   34.   d-C4=O4   34.
4    387.3   w-C4=O4   30.   w-C1=O1   30.   r-C5H2    25.
5    444.8   w-C1=O1   32.   w-C4=O4   32.
6    499.7   d-C4=O4   40.   d-C1=O1   40.
7    552.8   d-5RNGas  48.
8    572.6   d-5RNGsy  49.   r-C5H2    16.
9    683.1   r-C3H2    19.   r-C2H2    19.
10   775.3   sC3-C4    29.   sC1-C2    29.
11   790.4   sC4-C5    27.   sC1-C5    27.
12   870.2   r-C5H2    30.
13   884.7   r-C2H2    28.   r-C3H2    28.   t-C5H2    27.
14   977.1   sC2-C3    65.
15   1104.6
16   1122.3  t-C5H2    57.
17   1128.5  t-C3H2    34.   t-C2H2    34.
18   1157.3
19   1204.8  w-C5H2    76.
20   1224.7  t-C3H2    21.   t-C2H2    21.
21   1255.1  w-C3H2    29.   w-C2H2    29.
22   1271.1  w-C2H2    39.   w-C3H2    39.
23   1387.2  sc-C5H2   100.
24   1432.7  sc-C2H2   50.   sc-C3H2   50.
25   1448.3  sc-C3H2   49.   sc-C2H2   49.
26   1695.2  sC4=O4    45.   sC1=O1    45.
27   1725.5  sC1=O1    44.   sC4=O4    44.
28   2943.2  sC3-H3sy  40.   sC2-H2as  29.   sC2-H2sy  20.
29   2945.3  sC3-H3sy  45.   sC2-H2sy  29.   sC2-H2as  21.
30   2951.1  sC5-H5sy  99.
31   3000.9  sC5-H5as 100.
32   3015.4  sC3-H3as  40.   sC2-H2sy  29.   sC2-H2as  21.
33   3022.5  sC3-H3as  45.   sC2-H2as  29.   sC2-H2sy  21.
```

Table 6: QM calculated Frequencies (cm⁻¹) and % contribution

```
     Frequencies (cm-1) and % contribution (from empirical calculated force
                                 constants)


     Freq               %                  %                  %
1     41.6   t-5RNGsy  90.
2    112.9   t-5RNGas  84.
3    371.6   d-C4=O4   33.    d-C1=O1   33.
4    404.6   w-C4=O4   40.    w-C1=O1   40.
5    435.0   w-C4=O4   41.    w-C1=O1   41.
6    507.9   d-5RNGas  47.
7    514.6   d-C4=O4   26.    d-C1=O1   26.
8    571.5   d-5RNGsy  62.
9    747.5   sC3-C4    25.    sC1-C2    25.    d-5RNGas  17.
10   755.8   sC4-C5    15.    sC1-C5    15.
11   788.0   r-C2H2    32.    r-C3H2    32.    r-C5H2    17.
12   895.6   r-C5H2    65.
13   925.5   r-C3H2    39.    r-C2H2    39.
14   984.2   sC2-C3    68.
15  1094.7   t-C5H2    93.
16  1138.1   t-C3H2    43.    t-C2H2    43.
17  1155.8   t-C3H2    42.    t-C2H2    42.
18  1159.4   w-C5H2    30.
19  1240.4   w-C3H2    24.    w-C2H2    24.
20  1272.1   w-C5H2    38.
21  1307.6
22  1385.9   w-C2H2    20.    w-C3H2    20.
23  1507.0   sc-C3H2   47.    sc-C2H2   47.
24  1518.7   sc-C5H2   42.    sc-C3H2   26.    sc-C2H2   26.
25  1523.8   sc-C5H2   45.    sc-C3H2   22.    sc-C2H2   22.
26  1729.1   sC4=O4    36.    sC1=O1    36.
27  1747.8   sC1=O1    32.    sC4=O4    32.
28  2797.5   sC5-H5sy  99.
29  2827.6   sC5-H5as 100.
30  2849.1   sC3-H3sy  49.    sC2-H2sy  45.
31  2855.6   sC2-H2sy  51.    sC3-H3sy  50.
32  2873.5   sC3-H3as  50.    sC2-H2as  45.
33  2884.9   sC2-H2as  51.    sC3-H3as  50.
```

Table 7: MM calculated Frequencies (cm^{-1}) and % contribution

3.1.4. Dihedral and improper terms

The last two sections of the parameter optimization are the dihedral and improper angles and force constants. The torsions are one of the degrees of freedom that are optimized while optimizing the bond and angle force constants. The dihedrals that have no hydrogen are fine-tuned next. The optimization of dihedrals is designed to best reproduce the QM potential energy surface (PES). The QM calculations use the MP2 level of theory and the 6-31G* basis set for the target data.

All the dihedrals are held fixed, while the one under consideration is restrained to a certain value and the energy calculated after a short minimization. The dihedral under consideration is then restrained to a new value, while still keeping the other dihedrals fixed, and the energy is again calculated after a short minimization. This is done for a set of several restrained values for the dihedral under consideration till a PES is mapped out. The short minimization carried out consists of 50 conjugate gradient algorithm steps and 100 Newton-Raphson algorithm steps. For example, the improper angle is first restrained to -45.0 degrees, the system minimized and energy calculated. The improper is angle is then restrained to -44.0 degrees, system minimized and energy calculated. This is done until a PES is obtained for the improper angle 'scan' in 1.0 degree increments from -45.0 degrees to +45.0 degrees. Similarly, I get the PES for all the dihedrals I intend to optimize, one at a time. The regular dihedral angles were restrained from -90.0 degrees to +90.0 degrees in 3.0 degree increments. Emphasis is placed on reproducing low energy regions and low barriers, since high barriers will not be crossed during typical room temperature MD simulations. I get a similar PES from QM calculations for the target

data. Just like the other parameter optimization sections I try to reproduce the target data by adjusting the dihedral parameters.
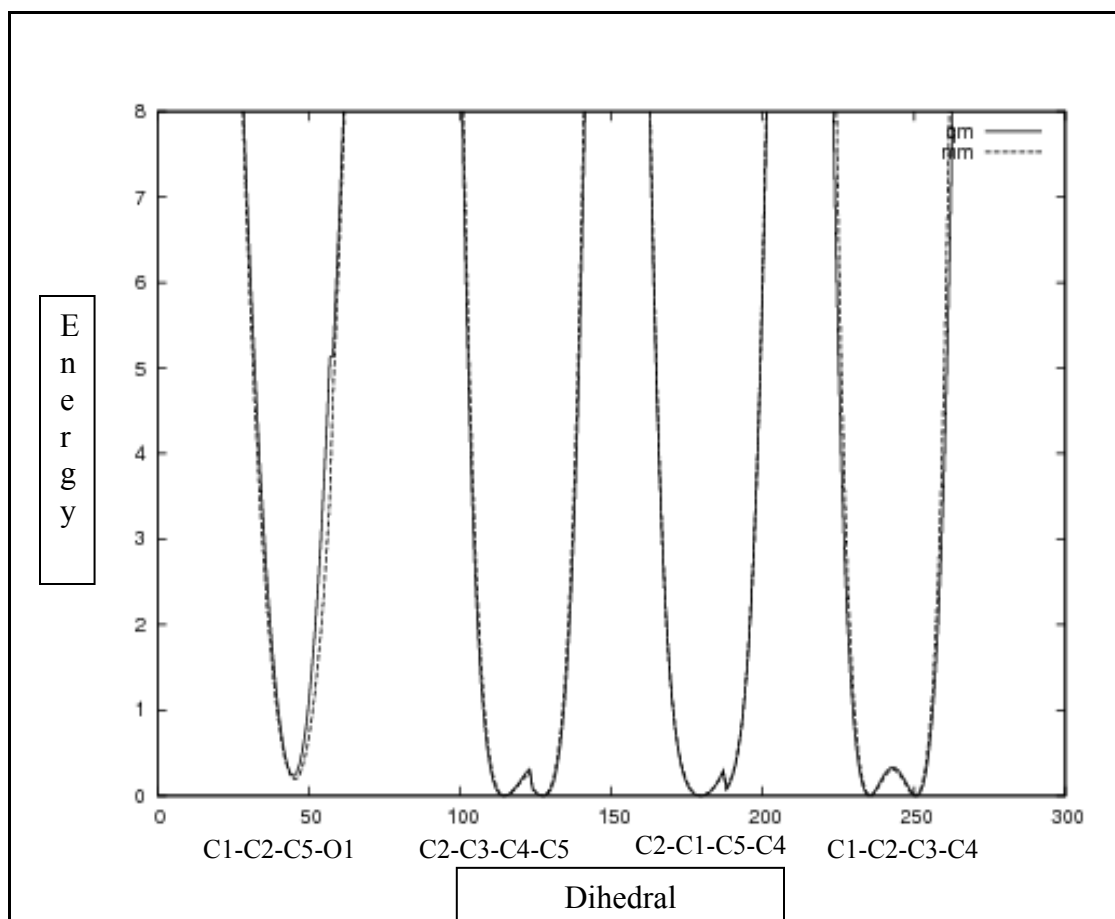

Figure 25: Dihedral potential energy

```
Topology and parameter files for CPD
* Stream file with compound to be parameterized.
*
read rtf card append
* Topology of compound to be parameterized.
*
36 1


RESI CPD            0.00


GROUP
ATOM C1   CG2R53   0.48
ATOM O1   OG2D1   -0.46
ATOM C2   CG3C52  -0.19
ATOM H21  HGA2     0.09
ATOM H22  HGA2     0.09
ATOM C3   CG3C52  -0.19
ATOM H31  HGA2     0.09
ATOM H32  HGA2     0.09
ATOM C4   CG2R53   0.48
ATOM O4   OG2D1   -0.46
ATOM C5   CG3C52  -0.20
ATOM H51  HGA2     0.09
ATOM H52  HGA2     0.09


BOND C1   C2       C2   C3        C3   C4        C4   C5        C5   C1
BOND C2   H21      C2   H22
BOND C3   H31      C3   H32
BOND C5   H51      C5   H52
DOUBLE C1   O1     C4   O4
IMPR C1 C2 C5 O1
IMPR C4 C5 C3 O4
ACCEPTOR O1
ACCEPTOR O4


IC   C1   C2   C3    C4    0.0000   0.00   0.0    0.00   0.0000
IC   C2   C3   C4    C5    0.0000   0.00   0.0    0.00   0.0000
IC   C3   C4   C5    C1    0.0000   0.00   0.0    0.00   0.0000
IC   C5   C2   *C1   O1    0.0000   0.00  120.0   0.00   0.0000
IC   C1   C3   *C2   H21   0.0000   0.00  120.0   0.00   0.0000
IC   C1   C3   *C2   H22   0.0000   0.00  240.0   0.00   0.0000
IC   C2   C4   *C3   H31   0.0000   0.00  120.0   0.00   0.0000
IC   C2   C4   *C3   H32   0.0000   0.00  240.0   0.00   0.0000
IC   C3   C5   *C4   O4    0.0000   0.00  120.0   0.00   0.0000
IC   C4   C1   *C5   H51   0.0000   0.00  120.0   0.00   0.0000
IC   C4   C1   *C5   H52   0.0000   0.00  240.0   0.00   0.0000
PATC  FIRS NONE LAST NONE


END
```

Table 8: CHARMM Topology File for CPD

```
read param card flex append
* Parameters needed for CPD
*

BONDS
CG3C52 HGA2    307.00    1.1000
CG2R53 CG3C52  285.00    1.5300
CG3C52 CG3C52  187.00    1.5300
CG2R53 OG2D1   691.00    1.2350


ANGLES
HGA2    CG3C52 HGA2     39.50    106.00    5.40    1.802
CG2R53 CG3C52 HGA2      41.50    111.00
CG3C52 CG3C52 HGA2      36.00    111.40   22.53    2.179
CG2R53 CG3C52 CG3C52    50.00    106.50
CG3C52 CG2R53 CG3C52    93.00    109.50
CG2R53 CG3C52 CG2R53    75.00    106.50
CG3C52 CG2R53 OG2D1     70.00    122.00



DIHEDRALS
CG2R53 CG3C52 CG3C52 CG2R53    0.0000  3    0.000
CG2R53 CG3C52 CG2R53 CG3C52    0.1000  3    0.000
CG3C52 CG3C52 CG2R53 CG3C52    0.000   3    0.000
OG2D1  CG2R53 CG3C52 CG2R53    0.0078  3  172.550
OG2D1  CG2R53 CG3C52 CG2R53    0.9000  1    0.000
OG2D1  CG2R53 CG3C52 CG3C52    0.0078  3  160.564
OG2D1  CG2R53 CG3C52 CG3C52    0.9000  1    0.000
OG2D1  CG2R53 CG3C52 HGA2      0.0000  3    0.000
OG2D1  CG2R53 CG3C52 HGA2      0.0000  1    0.000
CG3C52 CG2R53 CG3C52 HGA2      0.1000  3    0.000
CG2R53 CG3C52 CG3C52 HGA2      0.1000  3    0.000


IMPROPERS
CG2R53 CG3C52  CG3C52   OG2D1  90.0000  0    0.000


END
```

Table 9: CHARMM Parameter File for CPD

3.2. HOCl
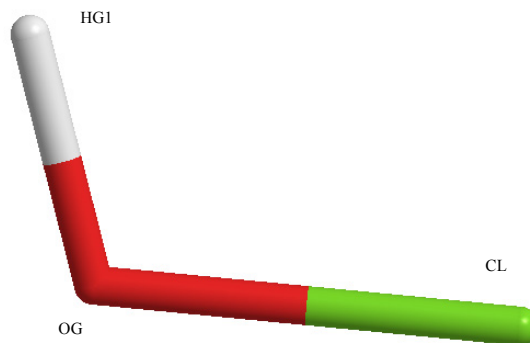
Figure 26: Hypochlorous acid (HOCl). Atom names are shown

Parameters for HOCl were also developed from the ground up, using the procedure described in detail for CPD. Developing parameters for HOCl was simpler than for CPD since HOCl has only three atoms and three atom types. Obviously, no dihedral optimization is needed and only two bond parameters and one angle parameter in addition to the partial atomic charges are needed. The van der Waals parameters are taken directly from similar atom types in the existing parameter file and not modified, as for CPD and as recommended. The finalized parameters are given in Appendix. Table 11 and Table 12 below compares QM and MM observables.

Atom Name	Atom Type	Description
CL	CLGA1	Chlorine e.g. Mono and dichloro ethane

OG	OG311	hydroxyl oxygen
HG1	HGP1	polar hydrogen

Table 10: HOCl Atom Names and Atom Types

	QM (kcal/mol)	CHARMM (kcal/mol)	Δ (CHARMM – QM) (kcal/mol)
H - OHH	-9.6396	-6.61081	3.02879
O - HOH	-2.76436	-3.6361	-0.87174
Cl - OHH	-2.6294	-0.588162	2.041238
		AVG Difference	1.39943
		RMS Difference	1.65579

Table 11: HOCl – H₂O Interaction Energies

	QM (kcal/mol)	CHARMM (kcal/mol)	Δ (CHARMM – QM) (kcal/mol)
H - OHH	1.86	1.86	0
O - HOH	1.97	1.97	0
Cl - OHH	2.92	3.53	0.61

Table 12: HOCl – H₂O Interaction Distances

QM (D)	CHARMM (D)
1.8253	1.96711

Table 13: Dipole moments

```
*Topology file for HOCl
* Stream file with compound to be parameterized.
*


RESI HOCL           0.00
!
GROUP
ATOM HG1   HGP1     0.38
ATOM OG    OG311   -0.49
ATOM CL    CLGA1    0.11

BOND CL   OG   OG HG1
DONO HG1 OG
DONO CL OG
ACCE OG

PATC   FIRS NONE LAST NONE

END
```

Table 14: CHARMM Topology File for HOCl

```
* read param card flex append
* Parameters needed for HOCl
*


BONDS
CLGA1  OG311   428.00      1.71739

ANGLES
CLGA1  OG311  HGP1    50.00    102.637

END
```

Table 15: CHARMM Parameter File for HOCl

3.3. CPO sugars

Sugar parameters were developed using the analogy method, for two reasons. First, the CHARMM force field contains parameters for some related sugars, and these parameters are expected to be largely transferable to the other sugars present in Chloroperoxidase. Second, the sugars themselves are located at the periphery of CPO, far from the active site where reaction takes place. Thus calculations of the reaction barrier for epoxidation at the active site or chlorination, if it occurs at the active site, are not expected to be very sensitive to the sugar parameters, and approximate parameters would be sufficient under such circumstances.

Application of the analogy method involves minor modification of atoms or atom types that would not significantly affect the energetic. This requires some chemical intuition. The process by which sugar parameters for addition to the default CHARMM topology file is described below.

3.3.1. Alpha-D-xylopyranose (AXYS)

Alpha-D-xylopyranose is similar to alpha-D-glucose (AGLC), which is already present in the CHARMM sugar topology file. All that is required to construct AXYS is to remove a small group of atoms from AGLC and replace them with hydrogen atoms. There are a few important details, so I describe the procedure carefully below.

I make a new entry for AXYS in the topology file. The AGLC section in the topology file is copied for AXYS, keeping only the atoms that are common to both molecules. Then the two extra hydrogen atoms are added to complete the AXYS

molecule. The following atoms from AGLC are deleted: C6, H61, H62, O6, HO6. The atoms H51 and H52 are added. Partial atomic charges must be adjusted slightly so that the individual charges remain consistent and the overall charge is zero. In the process of adjusting charges, the aliphatic hydrogen atoms are given charges of +0.09 as recommended by MacKerell. Carbon atom C5 has its charge adjusted to +0.250 to accommodate overall charge neutrality.
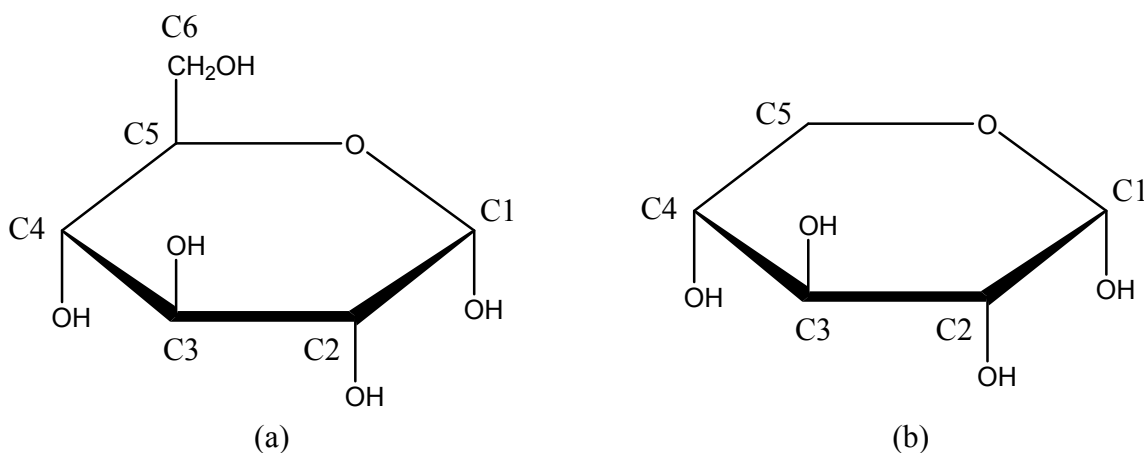
Figure 27: Sugar: (a) alpha-D-glucose (b) alpha-D-xylopyranose

Adjustments are made in the 'BOND' section of the topology file to provide information about which atoms are bonded. The donor ('DONO'), acceptor ('ACCE') and internal coordinate (IC) table sections are also modified. This is not absolutely necessary, though, because the information they provide can be included in a CHARMM input script if required.

3.3.2. Beta-L-xylopyranose (BXYS)

The molecule BXYS and AXYS are similar. Therefore all parameters for BXYS are the same as for AXYS; only the coordinates or stereochemistry are different. An α or β sugar are defined by the stereochemistry of the first carbon (C1) with respect to the last chiral carbon atom (C4). If they have the opposite stereochemistry the sugar is define as α and if they both have the same stereochemistry the sugar is defined as β.

We do not need to consider the orientation or the cartesian coordinates in the topology file (they are gotten from the Protein Data Bank (PDB) or a coordinate (CRD) file). Consequently, all that is needed is to create an entry for BXYS and copy all the parameters directly from AXYS.

3.3.3. Beta-L-arabinose(BARB)

Like BXYS, BARB is similar to AXYS. Essentially, only the coordinates differ between BARB and AXYS; the parameters are again identical. A new entry BARB is created with the AXYS parameters copied in.

3.3.4. N-Acetyl-Glucosamine (BNAG)

Like AXYS, BNAG is similar to a sugar already in the parameter set, alpha-D-glucose (AGLC).   Thus parameterization of BNAG requires a small modification of AGLC parameters. However, in this case the hydroxyl group attached to C2 is replaced by –NHCOCH3 group. The new group is split in to three sub-groups NH, CO and CH3, each with a net partial charge of zero.

3.3.5. Pyroglutamic acid (PCA)

Pyroglutamic acid is an uncommon amino acid and is a terminal amino acid in CPO. It is similar to a more common amino acid, glutamic acid. The terminal hydroxyl is missing in PCA and the delta-carbon in glutamic acid is connected to the backbone nitrogen to form a ring structure in PCA.

3.4. Patches

Patches are tools used in CHARMM when individual residues or molecules are required to be connected to others. These connections can be made explicitly in the CHARMM input script, but if this needs to be done regularly and often, it is worth creating a patch to do so.

A patch usually contains two residues that are to be connected. Often some atoms need to be deleted; bond, angle and perhaps dihedral terms need to be added; and in some cases partial charges of the atoms need to be adjusted to maintain the overall charge deletion of atoms.

In the macromolecule CPO there are several sugar molecules that are connected to each other and also to the enzyme. To accomplish this in the model, several patches were created in the topology file. They are to be invoked in the input script when the entire enzyme is built. We also need a patch to connect the heme to the proximal ligand cysteine (CYS). An example of one patch is given below, and the others are given in the Appendix.

```
PRES SSO4       0.000   ! patch between MAN-MAN( C1-O4 )


DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO4
BOND 1C1 2O4


GROU
ATOM 1C1   CC3162    0.240
ATOM 1H1   HCA1      0.090
ATOM 1C5   CC3163    0.080
ATOM 1H5   HCA1      0.090
ATOM 1O5   OC3C61   -0.500
GROU
ATOM 2C4   CC3161    0.360
ATOM 2H4   HCA1      0.090
ATOM 2O4   OC311    -0.450
```

Table 16: Patch to connect two mannose sugar molecules

3.5. Conclusions

In summary, I developed a set of molecular mechanics parameters for two molecules used in the study of the CPO halogenation mechanism and a set of sugars used in the study of the CPO epoxidation mechanism. My modeling work uses the CHARMM parameter set and therefore, to be consistent, I followed the protocol outlined by MacKerell et al. to develop parameters. I developed parameters for CPD and HOCl from the ground up because precise parameters were needed. The parameters developed were tested and met the requirements outlined in the protocol by MacKerell et al [72].

I used the less time-consuming analogy method to develop parameters for sugars. The sugar structures are very similar to other sugars already in the default CHARMM parameter set, and these sugars are not close to the enzyme active site and do not require a high level of accuracy. For these reasons I expect the analogy method to be satisfactory in this case. Patches were developed to create heme–cysteine, sugar–sugar and sugar–amino acid covalent bonds.

4. STUDY OF CPD DIKETO → KETO-ENOL TAUTOMERIZATION WITH LIMITED HYDRATION

4.1. Gas phase / Intramolecular proton transfer

I begin for reference with the intra-molecular hydrogen transfer process, in which a hydrogen is transferred from the alpha-carbon to one of the keto oxygens of CPD. As expected the adiabatic reaction barrier is quite high, 63.9 kcal/mol. Reactant, transition state and product structures are shown in Figure 28. Single-point calculations using the PCM method to model solvation implicitly do not alter the barrier heights significantly. The PCM barrier is 64.6 kcal/mol.
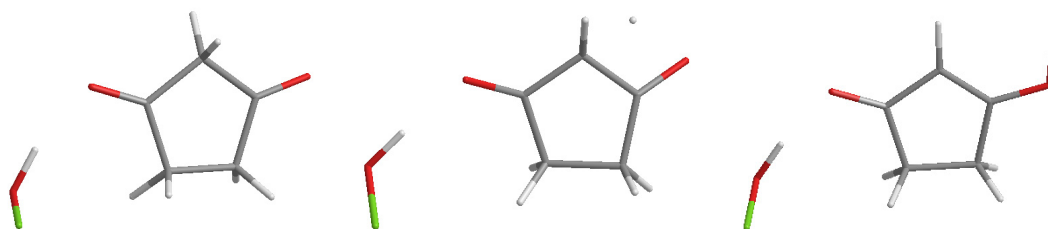
Figure 28: Intramolecular gas phase reaction. Stationary points. HOCl present to play stabilizing role representative of acidic conditions

4.2. Intermolecular proton transfer with one catalytic water molecule

When one catalytic water molecule is included, the reaction barrier decreases to 27.7 kcal/mol. Reactant, transition state and product structures are shown in Figure 29.

The water-catalyzed mechanism is a two-step process, with nearly complete abstraction of the alpha hydrogen by the water preceding transfer of a hydrogen from the water to the keto oxygen. As with intra-molecular hydrogen transfer, the implicit solvation model PCM method single-point calculation did not alter the barrier height significantly. The PCM barrier for the intermolecular hydrogen transfer is 26.8 kcal/mol.

The ketoenol product is slightly more stable that the diketo reactant in both the uncatalyzed and the single-water-catalyzed reactions (Table 17); PCM-modeled solvation increases the degree of product stabilization.
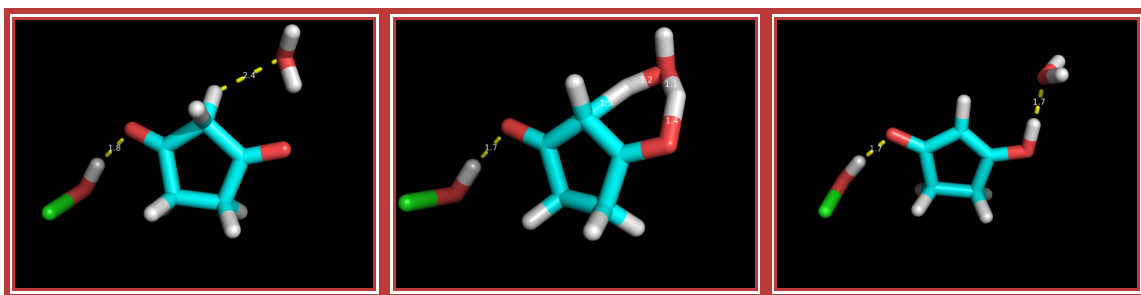
Figure 29: Intermolecular gas phase reaction. Stationary points. HOCl present to play stabilizing role representative of acidic conditions

4.3. Water clusters, fully QM optimized

Previous work on related systems has shown that additional explicit waters can lower the reaction barrier significantly [43-45]. Therefore I systematically evaluated the effect of additional hydration on the reaction barrier. To do so, I constructed clusters of water molecules surrounding the CPD/HOCl system and determined the adiabatic reaction barriers, treating the entire system at the QM level. Systems with multiple

waters possess a large number of degrees of freedom and great flexibility. Therefore the potential energy surface (PES) has a high dimensionality and many local minima. Meaningful, interpretable results require careful placement of waters. My approach was to begin with the stationary points of the single-water system and add water molecules incrementally. The water molecules were placed in highly favorable positions for forming hydrogen bonds. For reactants, the waters were added incrementally until the hydrogen bonding positions around the keto oxygens, the HOCl and the catalytic water were saturated, that is, until these hydrophilic atoms had one full solvation shell of explicit waters. I attempted to maintain similar degrees of hydration of these three moieties as I increased the level of hydration. To increment the degree of hydration, one or two new waters were added to a previously minimized reactant structure, and this structure was fully minimized. The same procedure was applied to the corresponding product and transition state structures. With this procedure, trends in barrier heights and reaction energies exhibited a systematic relation to level of hydration. When incremental waters were not carefully placed, however, trends were not systematic and were difficult to interpret (data not shown).

The results are shown in Table 17 and Figure 34. The overall reaction energy, ΔE, decreases from -3.1 kcal/mol for a single catalytic water to -17.2 kcal/mol for eight waters. Eight explicit waters are sufficient to model the overall reaction energetics, as ΔE is nearly unchanged beyond eight waters. The reaction barrier exhibits a stronger dependence on level of hydration, and in a way that is critically dependent on the placement of incremental waters. The barrier height is 27.7 and 27.6 kcal/mol with one or three explicit waters, respectively. Waters two and three were placed so as to hydrate

the HOCl, having little effect on the reaction barrier. However, the barrier decreases rapidly as more explicit waters are added. Waters four and five were placed so as to hydrate the catalytic water, and thus had a substantial effect on the barrier. For systems with more than five water molecules, I was unable to fully optimize the transition states. This difficulty probably stems from the large number of highly flexible degrees of freedom and consequently quite flat PES in the vicinity of stationary points.
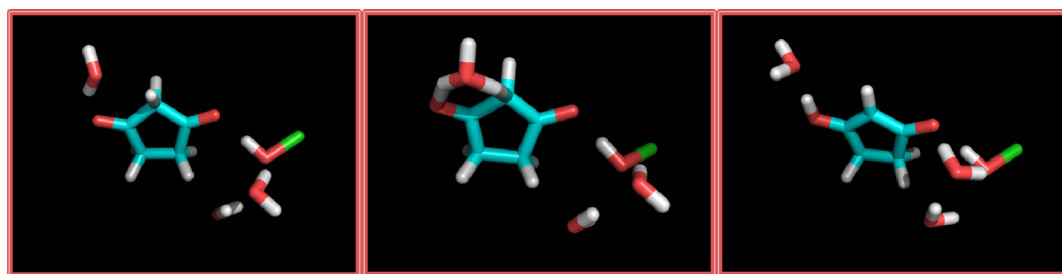
Figure 30: Stationary points of system with 3 explicit water molecules
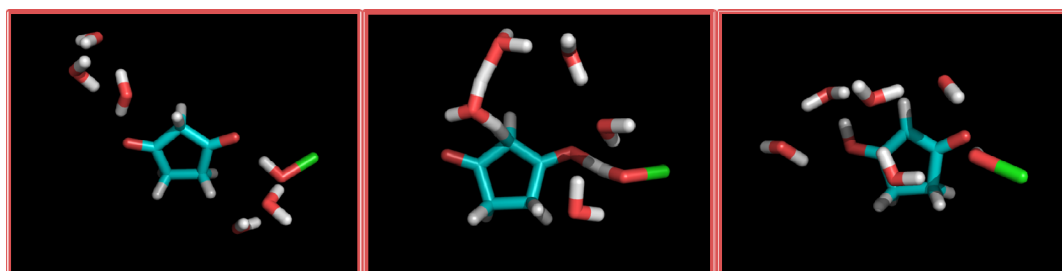
Figure 31: Stationary points of system with 5 explicit water molecules
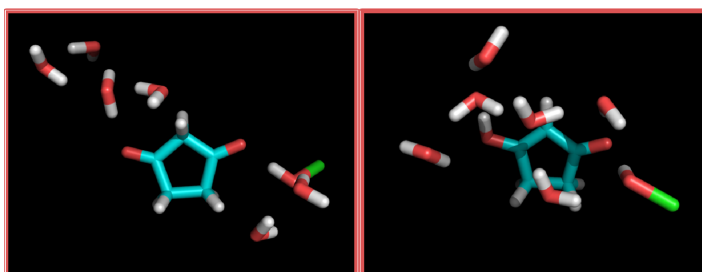
Figure 32: Stationary points of system with 6 explicit water molecules
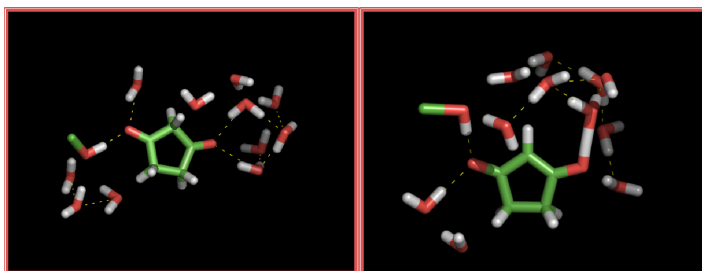
Figure 33: Stationary points of system with 11 explicit water molecules

For each stationary point at each hydration level, I also performed single-point energy calculations using the PCM method. The reaction energies changed by up to 8.3 kcal/mol. However, the barrier heights changed by no more than 1.6 kcal/mol. This finding is not consistent with the results described above, which demonstrate that adding explicit waters can significantly alter the barrier heights, at least for waters added to the first hydration shell of a hydrophilic moiety such as a keto oxygen. I conclude that an implicit water environment is not capable of accurately assessing the energetics of this reaction, in which specific directional solvent-solute interactions, and also solvent-solvent interactions in the vicinity of the solute, are important. It represents the barrier for fully optimized structures.

Incremental explicit water molecules placed in the manner described significantly stabilize all structures (reactants, transition states and products) with respect to infinitely separated species. Stabilization is generally greatest for the transition state, with the result that the reaction barrier decreases with increasing level of hydration. With five explicit water molecules, the forward reaction (diketone to ketoenol) is significantly lower. To make a physical connection, such structures might be thought of fully optimized and stable conformation that one might expect to see in a protective channel in an enzyme such as the major channel of CPO. This stable network of water molecules is present in

the crystal structure reported by Kuehnel et. al. [38]. While this method does allow us to model a highly stable environment like the enzyme major channel of CPO, it is not representative of kinetics in solution at 300 K.

| # H_2O's | in vacuo | | PCM (ε=78.4) | |
	ΔE (kcal/mol)	$\Delta E^{TS,f}$ (kcal/mol)	ΔE (kcal/mol)	$\Delta E^{TS,f}$ (kcal/mol)
0	-1.6	63.9	-2	64.6
1	-3.1	27.7	-6.3	26.8
3	-5.1	27.6	-5.6	
5	-12.7	15.5(22)	-4.4	17.1
6	-11.6			
8	-17.2		-11.8	
10	-16.7			
11	-17.0			

Table 17: Reaction energies and barrier heights (kcal/mol) for reaction in water clusters
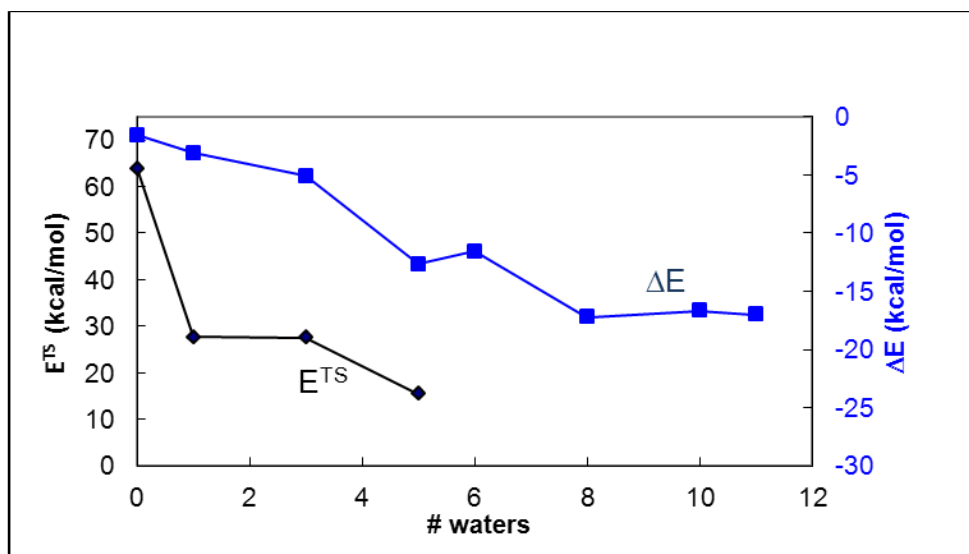
Figure 34: Reaction energies and barrier heights for reaction in water clusters

It is interesting to note that the water molecules significantly affect the barrier height only if they participate directly in a proton transfer, acting as a catalyst or if they help stabilize the catalytic water. For example, in the cluster with three waters only one water molecule is involved in the proton transfer and acts as a catalyst. The other two waters are only involved in hydration. The reaction barrier, 27.6 kcal/mol, is almost the same as for the single-water example (27.7 kcal/mol). When two more water molecules were added in close vicinity to the catalytic water the barrier height dropped significantly. The reaction barrier dropped from 27.7 kcal/mol to 15.5 kcal/mol.

I also used the IRC method [58, 59] to obtain the respective reactant and product structures from the transition states. The same reactant and product structures were obtained as before, except for the system with five water molecules. The IRC barrier is given in parentheses in Table 17 for the system with five water molecules.

4.4. Conclusions

The calculations demonstrate that the barrier height is quite sensitive to the particular pattern of an elaborate hydrogen bonded structure. Thus for reaction at the active site of CPO, it is likely that the placement of water molecules, and possibly also charged side chains, is critical for achieving the lowest barrier to reaction. The barrier height is observed to be most sensitive to the placement of water molecules near the first catalytic water and less sensitive to waters located elsewhere. No proton transfer is observed from HOCl to CPD as might have been expected to initiate or accompany abstraction of the alpha hydrogen as H^+.

The energies of reaction leveled off as the number of waters increased beyond five, but the barrier heights did not. Thus previously reported theoretical calculations on keto-enol tautomerization, which involved 1-4 waters, are unlikely to have achieved reaction barriers representative of solution conditions, which was in most cases the stated goal. More waters, and probably a different approach altogether, are needed to realistically model the barriers in full solvation conditions.

The location of transition states with many water molecules present is very difficult due to the high dimensional, flat potential energy surface. Therefore extending this approach to larger clusters was not feasible. It is emphasized that these structures are what one might expect to see in a protective channel in an enzyme such as the major channel of CPO and unlikely to be representative of liquid water at room temperature. In order to determine the barrier height representative of full solvation conditions, I employed a different approach, which is described in the next chapter.

5. STUDY OF CPD DIKETO → KETO-ENOL TAUTOMERIZATION UNDER FULL SOLVATION CONDITIONS

To represent the solvated environment with explicit solvent molecules, I used Molecular Dynamics (MD) run at 300 K to obtain snapshots of reaction precursor structures. A QM/MM approach was then used to obtain an adiabatic reaction path and barrier height for each precursor structure. The replica path and Climbing Image Nudged Elastic Band (NEB) methods were used to determine the reaction paths.

5.1. MD simulation and selection of frames

For a reaction to occur in solution, the reactive molecules need to be favorably positioned and oriented with respect to each other. Such positioning, sometimes called a 'rare event' occurrence, is not likely to occur frequently on the nanosecond timescale. I therefore chose frames from the ten ns MD simulation that represent such 'rare event' occurrences. The MD provides the randomization of the water molecules present in a solvated environment.

When the HOCl and CPD approach each other in solution, they remain close for a duration on the order of a few ps, trapped by a solvent cage. I selected only one frame for each unique encounter, choosing frames at least ten ps apart. Ten ps represents the residence time for a hydrogen bond. Favorable conformations were selected such that the HOCl was near enough to CPD to form a hydrogen bond with one of the keto oxygens,

and a water molecule was positioned ready to accept a proton from the alpha carbon of CPD and to donate a proton to the other keto oxygen. The selection process thus used the following criteria: The distance between one of the keto oxygens of CPD and the hydrogen of HOCl is less than or equal to 2.4 Å (1). There exists a water molecule whose oxygen is within 2.4 Å of the alpha carbon of CPD (2) and one of whose hydrogens is within 3.2 Å of the keto oxygen not used in the first criterion (3). Statistics regarding the trajectory frames satisfying these criteria are shown in Figure 36. A representative selected structure is shown in Figure 35. From each unique encounter, the best frame was selected for further consideration. The best frame was taken as the one with the smallest value for criterion (1); in the case of a tie, the smallest value for criterion (2) was determining; and if a tie still remained, the smallest value for criterion (3) was decisive.
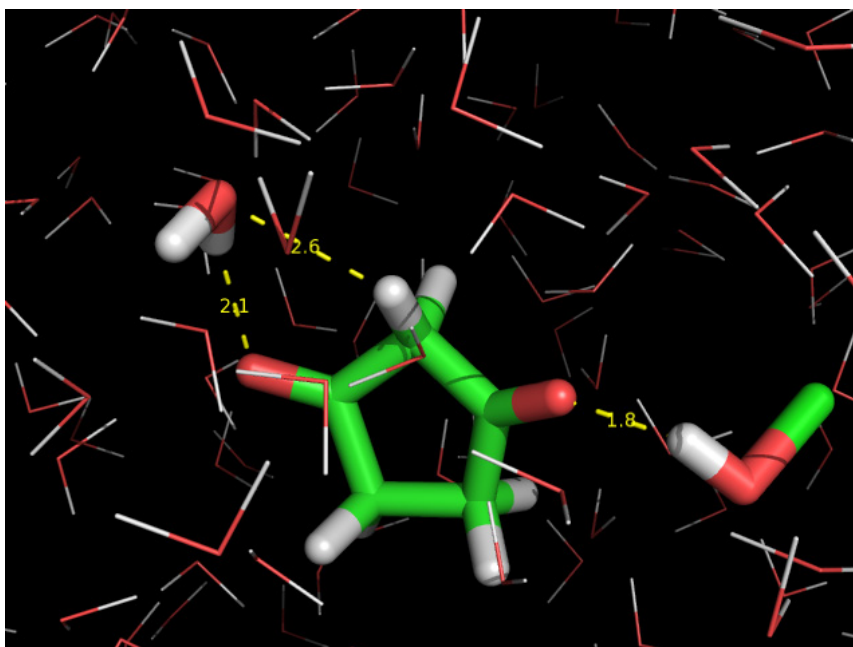
Figure 35: Representative rare event frame favorable for reaction, chosen from the MD simulation according to the selection criteria. Atoms treated at the QM level in subsequent QM/MM calculations are represented with thick sticks, and MM atoms are represented with thin lines
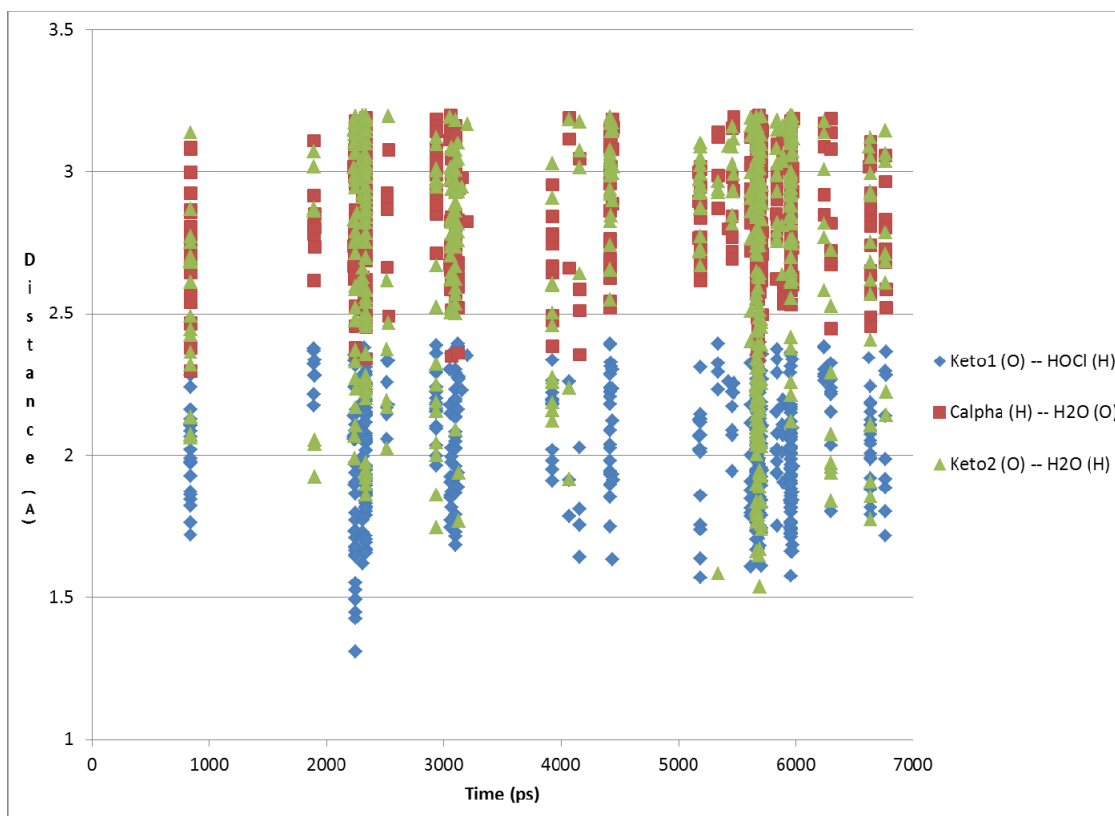
Figure 36: Trajectory frames satisfying the selection criteria. Each vertical group of data points represents one encounter, i.e., all frames lie within 10 ps of the best frame selected. Frames were saved every 20 steps (every 0.02 ps). 617 frames satisfied the selection criteria out of a total of 350,000 saved trajectory frames (0.18%)

5.2. *Determination of minimum energy path*

I determined the reaction barriers for the selected frames using a QM/MM model. I treated the CPD, HOCl and a single water molecule, which participates in the reaction catalytically, at the QM level and the rest of the water molecules at the MM level. All the MM waters were kept fixed while the QM atoms were flexible. Thus the barriers calculated are adiabatic with respect to the flexible degrees of freedom, in a fixed solvent environment representative of a liquid solution at 300 K. Physically, this represents a

reaction that takes place much faster than solvent response. This is clearly not what happens in reality but provides a well-defined limiting behavior for analysis. Each frame represents a different solvent environment consistent with a temperature of 300 K. Boltzmann weighted averaging of barriers across a sufficient number of frames can provide an effective activation barrier for the limiting physical process considered. The MEP was determined with chain-of-replicas methods as described earlier.

5.2.1. *Determining MEP using Replica Path Method*

I used the Replica Path Method as the method of choice to map out the minimum energy path. In this method, a chain of conformations, or replicas, representing stages of reaction between reactants and products is created. The distance between pairs of conformations is defined using an RMSD metric. Simultaneous optimization of all replicas with a constraint to maintain spacing between the replicas yields an MEP. Optimization with the RPATh facility in CHARMM required 500 steps to achieve the convergence criterion. The convergence criterion was an average root mean square gradient (GRMS) value of less than 0.06 kcal/mol/ Å between successive 100-cycle iterations. A typical optimization is shown in Figure 38. For frame 1 Table 18 shows the energy level convergences of the replicas for every 100 optimization cycle, Table 19 shows the gradient convergence criteria for every 100 optimization cycle and Table 20 shows that the RMSD deviations are within limit (i.e. less than 20% RMSD deviation from neighboring replicas or images) to validate the parameters used in RPATH. For the

optimizations, the end points were fixed. Consequently the minima, which correspond to reactants and products, occur at intermediate replicas. The reaction path obtained with RPATh results in approximately equal spacing between replicas on the PES. Thus the transition state is not located with precision. To obtain an accurate transition state, I refined the reaction path using the nudged elastic band method with the climbing image option (CI-NEB).
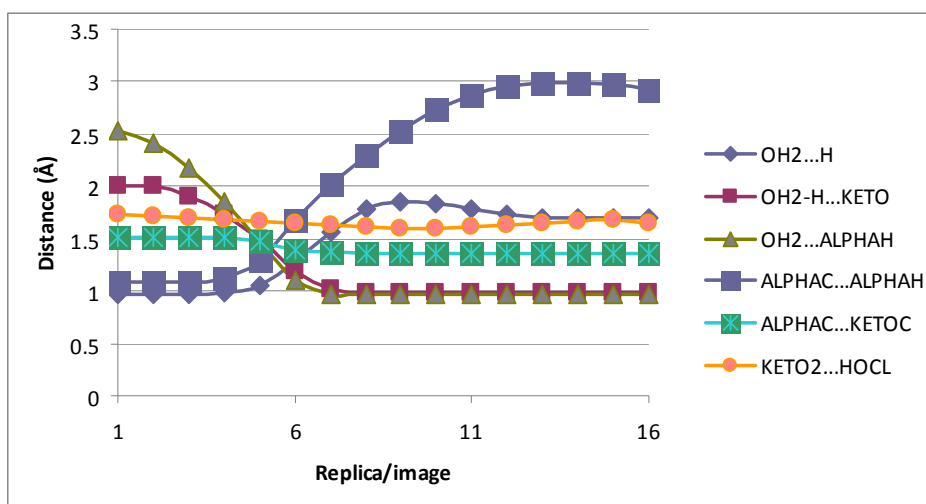
Figure 37: Frame 1: Distance change showing the hydrogen abstraction and enolization from reactant to product as the reaction proceeds in the vicinity of the transition state

	Optimization steps (SD + ABNR)				
Replicas/Images	50+100	50+200	50+300	50+400	50+500
1	21	21	21	21	21
2	2	2	2	1	1

3	3	2	2	2	2
4	4	4	6	9	7
5	9	10	19	==28==	==23==
6	20	26	==30==	18	22
7	==35==	==27==	12	5	6
8	21	12	5	2	2
9	11	6	2	1	1
10	7	4	2	1	1
11	5	3	1	1	1
12	3	3	1	1	1
13	2	2	1	1	1
14	1	1	0	0	0
15	0	0	0	0	0
16	15	15	15	15	15

Table 18: Frame 1: Normalized energy level for each replica for every 100 ABNR optimization cycles. The maximum energy for each optimization cycle is highlighted

Optimization Steps (SD + ABNR)	Avg. RMS Gradient
50+100	2.094
50+200	0.388
50+300	0.165
50+400	0.121
50+500	0.065

Table 19: Frame 1: Average RMS Gradient for every 100 ABNR optimization steps

Replicas/Images	RMSD	%RMSD deviation
1	0.123	2.70%
2	0.122	3.70%
3	0.124	1.80%
4	0.13	3.00%
5	0.147	16.60%
6	0.13	3.30%
7	0.126	0.00%

8	0.125	1.00%
9	0.125	1.00%
10	0.125	0.80%
11	0.124	1.50%
12	0.123	2.40%
13	0.123	2.30%
14	0.123	2.80%
15	0.123	2.80%

Table 20: Frame 1: RMSD and % RMSD deviation from neighboring replicas
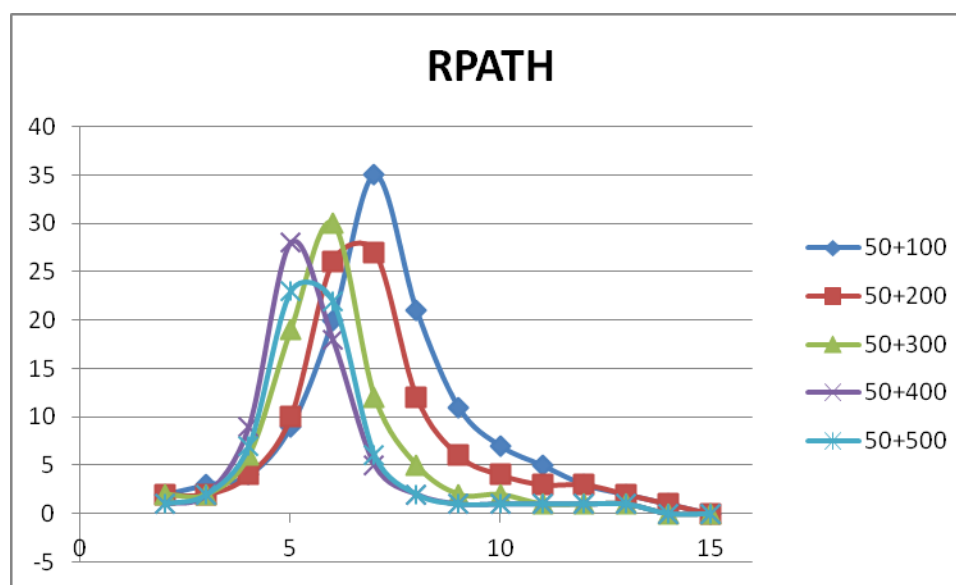
Figure 38: Optimization of Replica Path Method (SD+ABNR)

5.2.2. *Locating transition state using Climbing Image Nudged Elastic Band*

I use CI-NEB to locate the transition state once I have the MEP mapped out by the Replica Path Method. This option causes a path point near the transition state to climb in energy toward the transition state, while all other path points are energy minimized in directions orthogonal to the MEP. A typical optimization is shown in Figure 39. Further details are given in the supplemental information.
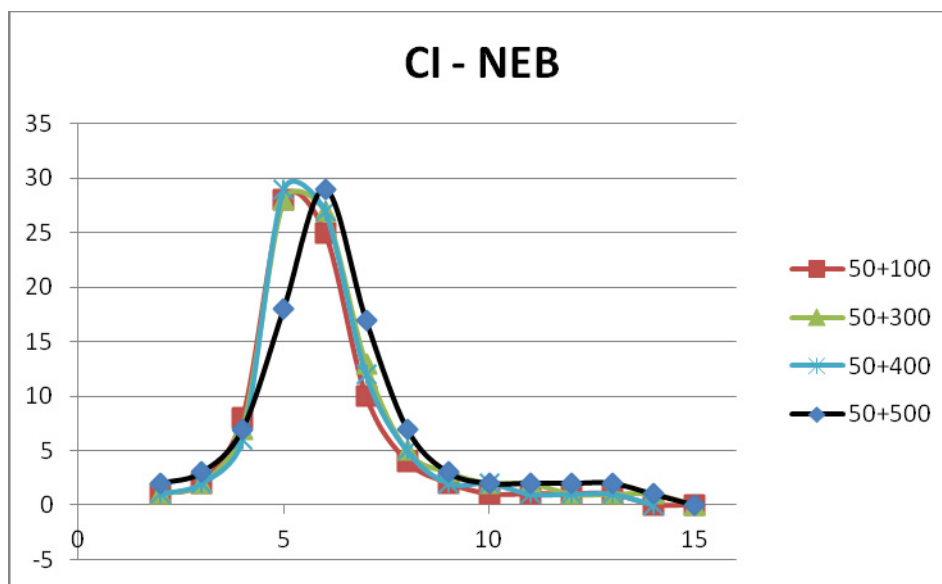
Figure 39: Optimization of CI-NEB to locate Transition state (SD+ABNR)

Table 21 gives the barrier height for each of the selected frames. There is a substantial variability in barrier height of 25-44 kcal/mol. Applying Boltzmann weighting to the barrier heights for the ten frames in Table 21, I obtain an effective activation energy (E_a) of 25 kcal/mol. Some important distances are tabulated in x, y, and table for reactant transition state and product respectively. To investigate the cause of the variation

in barrier heights, I visually analyzed each frame, focusing on hydrogen-bonding interactions likely to be important for reactants and transition states. These are summarized in Table 25.

Frame	ΔE
1	27
2	30
3	36
4	41
5	44
6	42
7	35
8	38
9	25
10	30

Table 21: Barrier heights (kcal/mol) for reaction for selected frames from solvated MD simulation

Frame	OH2 ... H	OH2-H ... KETO	OH2 ... ALPHAH	ALPHAC ... ALPHAH	ALPHAC ... KETOC	KETO2 ... HOCL
1	0.972	1.995	2.379	1.093	1.512	1.721
2	0.968	2.823	3.201	1.100	1.522	1.792
3	0.974	2.107	2.490	1.099	1.521	1.675
4	0.974	2.104	2.675	1.095	1.522	1.942
5	0.979	2.385	3.165	1.098	1.501	2.321
6	0.972	2.651	2.915	1.096	1.523	1.786
7	0.973	2.201	2.517	1.092	1.511	2.024
8	0.97	3.42	2.22	1.10	1.52	1.72
9	0.98	3.15	2.20	1.10	1.52	1.88
10	0.972	1.796	2.247	1.096	1.526	1.714

Table 22: Reactant Distances for all the frames

Frame	OH2 ... H	OH2-H ... KETO	OH2 ... ALPHAH	ALPHAC ... ALPHAH	ALPHAC ... KETOC	KETO2 ... HOCL
1	1.11	1.38	1.22	1.47	1.44	1.64
2	1.13	1.39	1.28	1.53	1.44	1.64

3	1.14	1.34	1.23	1.49	1.44	1.71
4	1.24	1.23	1.27	1.47	1.44	1.95
5	1.15	1.41	1.26	1.66	1.41	3.01
6	1.2	1.26	1.28	1.47	1.44	1.65
7	1.18	1.3	1.22	1.47	1.43	1.92
8	1.11	1.42	1.23	1.54	1.44	1.73
9	1.02	1.67	1.19	1.52	1.44	1.77
10	1.16	1.33	1.27	1.5	1.45	1.66

Table 23: Transition state Distances for all the frames

Frame	OH2 … H	OH2-H … KETO	OH2 … ALPHAH	ALPHAC … ALPHAH	ALPHAC … KETOC	KETO2 … HOCL
1	1.71	0.99	0.97	2.93	1.37	1.66
2	1.69	0.99	0.96	2.48	1.37	1.61
3	2.00	0.98	0.97	3.07	1.36	1.66
4	3.18	0.97	0.97	3.29	1.35	1.82
5	2.26	0.98	0.97	3.16	1.35	2.66

6	1.94	0.98	0.97	3.07	1.36	1.63
7	3.48	0.98	0.97	2.51	1.36	1.87
8	4.04	0.97	0.97	3.00	1.36	1.68
9	1.74	0.99	0.97	2.76	1.36	1.77
10	1.95	0.98	0.97	4.62	1.37	1.59

Table 24: Product Distances for all the frames

	Frame 1		Frame 9		Frame6	
	Reactant	TS	Reactant	TS	Reactant	TS
HBond	1	1	0	0	1	2
RMSD	0.07	0.21	0.38	0.30	0.29	0.52

Table 25: Important hydrogen bonding interactions for reactants, transition states and products calculated for selected frames

To gain further insight into the cause of the variability in barrier heights, I recalculated the MEP, allowing nearby waters to respond to the progress of reaction. The MM water molecules within 2.6 Å of the QM atoms were made flexible, with the remaining MM waters still fixed. Distinguishing the treatment of water molecules in this way is easily achieved in the replica path implementation in CHARMM, as it allows for

the replicated system to be larger than the QM system. When this test was performed on one of the frames with a high barrier, frame 6, the barrier height decreased by 7 kcal/mol. The flexible MM water molecules underwent significant motion to stabilize the transition state structure, but little motion in the reactant structure, which was already solvent stabilized. When the test was performed on low-barrier structures, frames 1 and 9, there was little change in the barrier height.

The hydrogen bond analysis parameters had a cutoff distance of 4.0 Å and an angle of 150.0 degrees between acceptor and donor atoms. Only hydrogen bonds between the QM solvation shell and the CPD, HOCl and the catalytic water were analyzed. The RMSD was calculated for the flexible solvation shell with respect to the fixed shell in the reactant and transition state structures for frames 1, 9 and 6.

Returning to the original reaction barrier calculations without the additional solvent flexibility, I can now trace the variability in barrier heights to the degree to which surrounding solvent waters stabilize the reactants. In no cases were transition states stabilized, but in some reactants were. In reality, it is unlikely that relatively massive solvent oxygens can move on the time scale of the proton transfer processes. Thus solvent stabilization of transition states is unlikely to occur, and the original set of barrier height calculations is more representative of the actual physical process than the calculations in which surrounding waters are allowed to move during the proton transfer processes.

5.3. *Conclusions*

I find that the solvating water environment creates a wide range of barriers for potentially reactive events. In my sampling of ten frames, these range from 25 to 44 kcal/mol. In view of Boltzmann weighting, the lowest of these are dominant and the effective barrier is 26 kcal/mol, which compares with the barrier of 28 kcal/mol for gas-phase reaction with a single catalytic water. Thus, surprisingly, the barrier in a fully solvated system and the barrier with only a single water present are nearly the same. Close analysis reveals that this is not because other waters do not play a significant role in reaction. Rather it is because of a balancing effect. In low-barrier reaction events, which are mainly responsible for the observed rate constant, surrounding waters stabilize the transition state and the reactants to nearly the same extent. This is not true of most potentially reactive encounters; more often, surrounding waters stabilize reactants to a greater extent. Assuming that proton transfer occurs faster than water molecules can reorient, those same solvating waters are usually ill-positioned to stabilize the transition state and thus have a net effect of retarding rather than promoting reaction. In a small percentage of cases, though, the surrounding waters are positioned so as to lower the energies of both reactants and the transition state. Thus I find that transition state stabilization by solvent water is critical for reaction in solution. Transition state stabilization is also thought to be an important mode via which enzymes catalyze reactions. Thus it appears that enzymes and aqueous solution may have similar ways of catalyzing keto-enol tautomerization.

6. OVERALL CONCLUSIONS

The goal of this work was to elucidate the chlorination mechanism of chloroperoxidase using CPD as a model substrate, and to shed light on which of two possible reaction pathways is more likely: reaction with free HOCl in aqueous solution or reaction at the enzyme active site. I show that the rate limiting step for the chlorination of CPD is the diketo-ketoenol tautomerization step and that the solvent environment plays a crucial role in the reaction. I model the hydration environments of the two possible pathways and calculate the reaction barriers for both. I observe that a limited and optimized (active-site-like) hydration environment leads to a lower energetic barrier than the fully solvated model representative of the aqueous environment at room temperature. This result indicates that the stable water network near the active site and the access channel is likely to facilitate the chlorination mechanism. Confirmation of this predication could be obtained from calculations of the reaction barrier with the enzyme active site explicitly modeled, including active site residues and hydrating waters, using a QM/MM approach. This, however, is beyond the scope of the current project.

I determined the nature of the influence of the solvent environment on the reaction barrier. I observe that stabilization of the catalytic water by other solvent molecules is critical for keto-enol tautomerization. When studying a limited-hydration environment, placement of water molecules is more important than the number of water molecules. Careful placement of hydrating water molecules saves a significant amount of computational time in such studies. Previous theoretical limited-hydration studies have been accepted as a tractable method for determining effective reaction barriers in

solution. My work demonstrates that the limited-hydration approach is inadequate for determining effective reaction barriers for keto-enol tautomerization. In particular, I demonstrate that previous theoretical studies of keto-enol tautomerization in a limited-hydration environment, which used five or fewer waters, are not representative of a solution environment. My modeling of the reaction in a fully-solvating environment demonstrates that reaction proceeds when the instantaneous dynamical water environment is close to optimal for stabilizing the transition state.

7. OTHER SYSTEMS STUDIED/CONTRIBUTED TO

This chapter describes my contributions to three other projects carried out in the Chatfield research group. Section 7.1 describes work on dehaloperoxidase, an enzyme bearing similarity to chloroperoxidase, the subject of my main project. Section 7.2 contains the preliminary results of an MD study of the solution dynamics of a small, highly cross-linked peptide, α-conotoxin GI. Section 7.3 contains work done on the CPO-catalyzed epoxidation reaction. My major contribution to the studies described in sections 7.1 and 7.3 was the development of parameters for molecular mechanics modeling, including MD simulation. This work was a natural application of the expertise I gained in parameter development working on parameters for molecules involved in the CPO-catalyzed halogenation reaction, as explained in chapter 3. My contribution to the project described in section 7.1 also involved carrying out MD simulations using techniques and scripts I developed for my main project. The entire project described in section 7.2 was carried out primarily by me. This work in section 7.2 was a preliminary project carried out before beginning the CPO-catalyzed halogenation work that is focus of this dissertation. In carrying out the work described in section 7.2, I became familiar with molecular modeling techniques that I was able to apply to the other system studied. My contribution to each project is detailed in the subsequent sections.

7.1. DHP

I made contributions to an MD study of the enzyme dehaloperoxidase (DHP), which is found in the marine worm Amphitrite ornata. The enzyme has a globin fold and can bind dioxygen as can most globins, but its primary function is to catalyze the dehalogenation of p-halophenols and related compounds. Thus DHP can function either as a dehalogenase or as a globin. The mechanism via which the enzyme switches between the two modalities is of interest. An active-site residue, His55, which can adopt multiple conformations, has been thought to play a major role in the switching between modalities. Molecular dynamics simulations of wildtype DHP and two mutants, H55V, H55R, with and without a p-iodophenol ligand present, were performed.

My contribution to this project was development of parameters for the p-iodophenol ligand, much of the simulation setup, and the running of several of the native and mutated enzyme simulations. Project design and analysis of the results were carried out by other members of the research group. A very brief summary of the simulation method and analysis to date is described in the next paragraph.

A 10-ns simulation was performed and analyzed for each system. Hydrogen bonds between the ligand, Tyr38 and His55 were found to highly stabilize the ligand at the active site. The ligand was found to block access to the heme iron, for most of the 10-ns simulation. When His55 was mutated to Val or Arg, the p-iodophenol ligand was no longer maintained in a position where it blocks access to the heme iron. The simulations also provided insight into the external binding site of the substrate.

7.2. α-Conotoxin GI

I began work on this project at the outset of my Ph.D. The project was framed as a joint MD/NMR study of motional motifs, with an ultimate goal of contributing to the development of MD-based methods for interpreting the results of dynamical NMR. For practical reasons relating to difficulties in obtaining the labeled peptide for the NMR portion of the study, the project was not completed.

Figure 40: Cone snail

The peptide α-conotoxin GI belongs to a family of small multiply cross-linked peptide toxins produced by marine mollusks known as cone snails. The slow-moving snails use the toxins to hunt their prey, which are usually fish. Its venom needs to act fast

in order to bring down its prey given the snails' limited mobility in comparison to its prey, which is one of the reasons its venom has evolved to be highly potent. Depending on the species, the conotoxins range in length from 13-30 residues, making them among the smallest of peptide toxins (snake venom toxins, by comparison, tend to be in the 70-residue range). Conotoxins act on the nervous system, serving to paralyze the snails' prey. For example, the α-conotoxins bind to the nicotinic acetylcholine receptor (nAChR), to Ca-gated voltage channels.

Conotoxins are notable for their specificity of binding. For example, the α-conotoxins of one species of cone snail are often inactive against the nAChR of the prey of another species of cone snail. The specificity of conotoxins are achieved through conservation of a scaffold based on patterns of cross-linking combined with a remarkable variability of loop composition. For this reason, conotoxins are of interest as possible lead compounds for drug development. The α-conotoxins are of particular interest because their binding properties may help to map out the nAChR, whose three-dimensional structure is thus far unknown.

The α-conotoxin GI was one of the first-isolated conotoxins and has been widely studied. Its sequence is ECCNPACGRHYSC, and there are disulfide bonds between residues 2 and 7, and 3 and 13. The three-dimensional structure has been characterized by X-ray crystallography and solution-state NMR. Two regions of secondary structure have been identified. In all structures, residues 4-8 have been described as 3-10 helical. Residues 9-13 have been described as a β-turn. In the case of the X-ray structure, this was on the basis of both backbone dihedral angles and hydrogen bonds. For the NMR structures, secondary structure assignment was based mainly on dihedral angles.
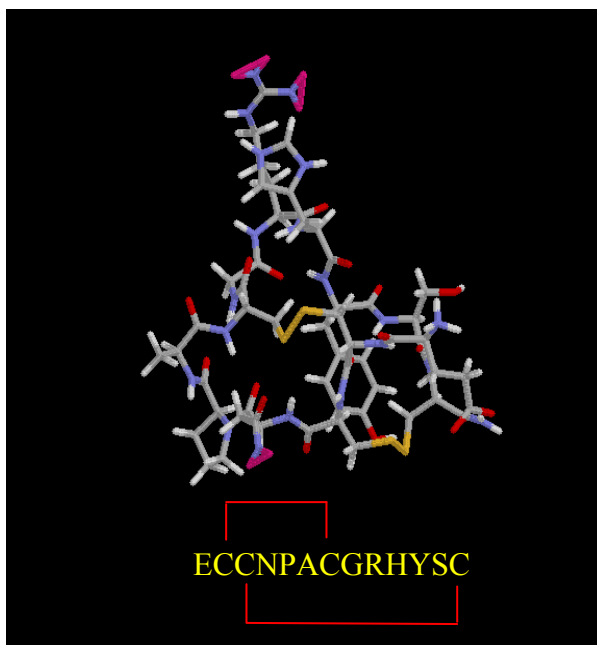
Figure 41: α-conotoxin GI

Notably, a recent NMR study identified two distinct conformers present in a 70:30 ratio (referred to hereafter as the major and minor conformers) on the basis of simulated annealing with the NMR structural restraints. It may be important to understand whether the conformers interconvert and whether one or both is active in binding to nAChR.

Multiple molecular dynamics simulations were performed on a 13-residue, doubly cross-linked peptide toxin, α-conotoxin GI, in each of two solution conformations identified by NMR. The two conformers remained distinct on the 100-ns simulation time scale, supporting the hypothesis that only one of the conformers is biologically active. The secondary structure as judged by backbone dihedral angles was preserved during simulation, but backbone hydrogen bonds were almost entirely absent. This observation suggests that binding at the active site probably involves hydrogen bonds to the peptide

backbone. The backbone exhibited unusual, concerted motions of dihedral angles in nonconsecutive residues, which may have implications for the binding free energy.

Below is a set of figures that bring out the salient features of the simulation. Figure 42 shows overlays of the several structures obtained from NMR structural data on peptide in solution using simulated annealing. The backbone and most sided chains have well defined conformations, but several side chains are clearly mobile. From this structure, two primary backbone conformations were determined. These are shown in Figure 43 and are called the major and minor structures. Figure 44 depicts the root mean square deviation (RMSD) in the backbone coordinates for ten 10-ns simulations of each structure (major and minor). The RMSD for both structures are shown together on each panel. It can be seen that the structures remain distinct on the 10-100 ns time scale.

Figure 45 depicts correlation functions, $C_i(t)$, for motion of backbone N-H bond vectors for all residues of each of the two structures. The correlation functions are defined by

$$C_i(t) = \langle P_2(\hat{u}_i(0) \cdot (\hat{u}_i(t)) \rangle$$

where $P_2 = \frac{3}{2}x^2 - \frac{1}{2}$ is a second Legendre polynomial and $\hat{u}_i$ is a unit vector in the direction of the N-H bond vector. The asymptotic value of $C_i(t)$ as time gets large is the order parameter, S^2. Small values of S^2 correspond to relatively unrestricted motion, and values of S^2 close to one correspond to highly restricted mobility. The two sets of data (major and minor) have very different asymptotic values, but this is because the data were treated differently. The trajectory frames for the major structures were reoriented

before calculating the correlation functions, while those for the minor structures were not. When the trajectory frames for the minor structures are reoriented to remove overall translation and rotation (data not shown), the asymptotic behavior of the correlation functions is very similar to what is seen for the major structures in Figure 45. Values of S^2 by residue are depicted with a color code for both structures and also for the crystal structure (obtained from X-ray B factors) in Figure 46. The crystal structure shows generally more motional restriction (larger average value of S^2) as a consequence of crystal packing; the simulation environment is aqueous solution.

Maps showing the pattern of N-H backbone bond vector motion are shown in Figure 47. These are a major finding of the work. The N-H bond vectors display differing modes of motion, with some exhibiting motion within a cone, others elongated mobility, and yet others bimodal sampling. This indicates that interpretation of NMR-derived order parameters in terms of motional models must be made with care.

Finally, Figure 48 shows that some backbone dihedrals angles exhibit correlated motions. When interpreting order parameters in terms of conformational entropy, this correlated motion will need to be taken into account.
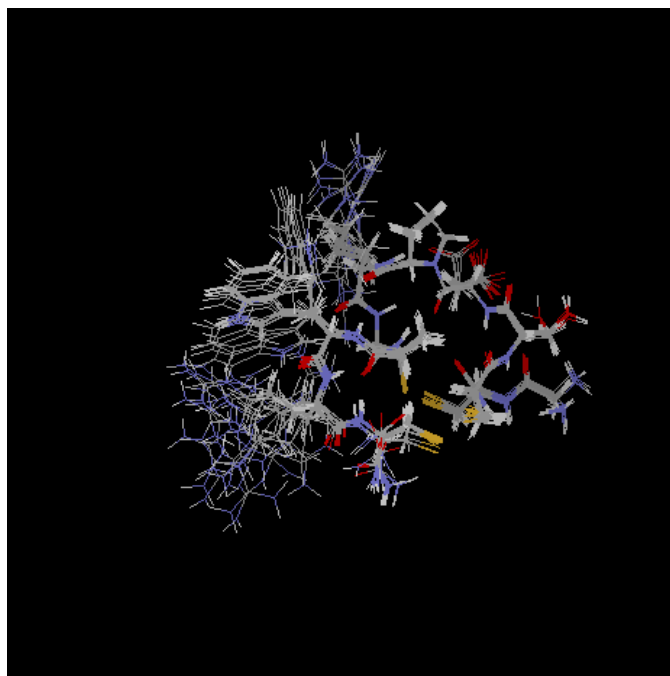
Figure 42: NMR structure showing stable backbone
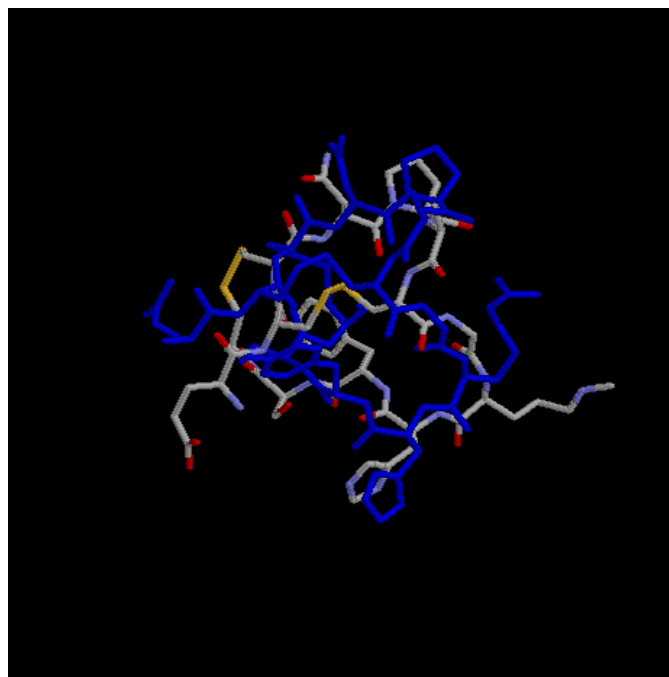
Figure 43: distinct structures in solution – Major and Minor
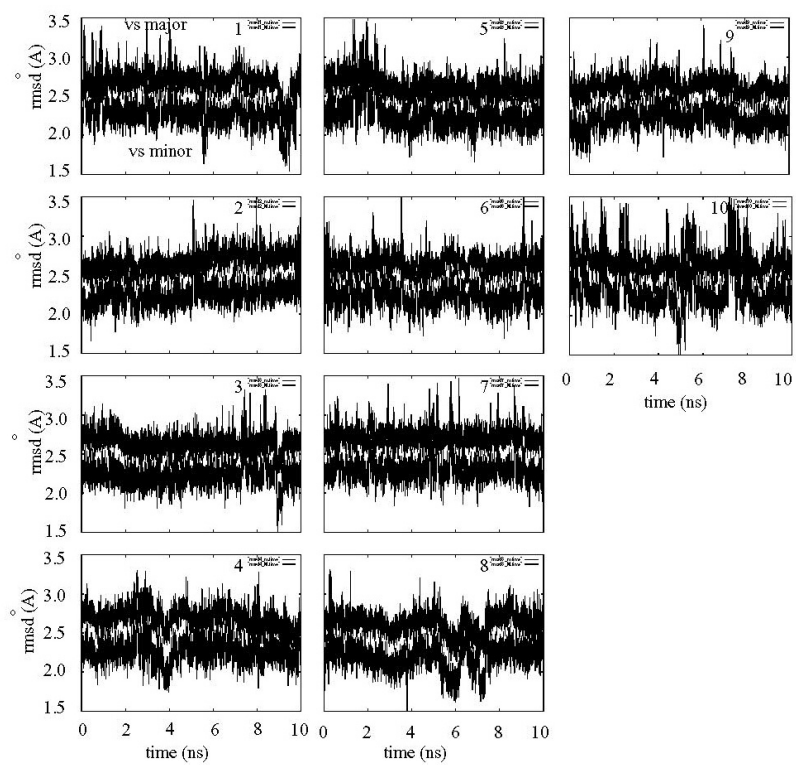
Figure 44: Simulation of major and minor conformation for 10x10 ns simulation demonstrating that the two structures are stable and distinct
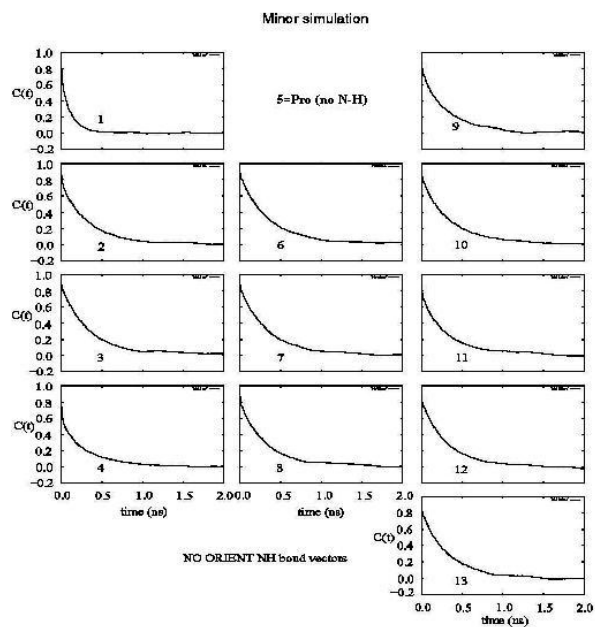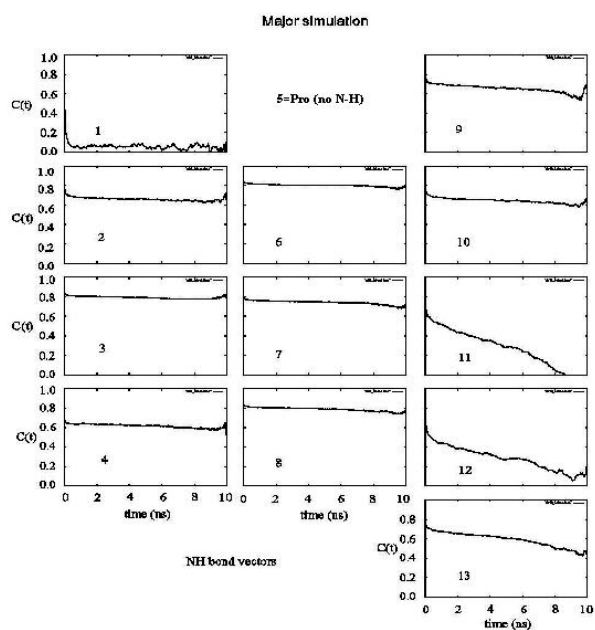
Figure 45: N-H correlation functions. Trajectory frames were reoriented for Major conformation but not for Minor conformation (S^2 values shown elsewhere are based on reoriented data.)

S^2	
0.8+	(purple)
0.7+	(blue)
0.6+	(green)
0.5+	(yellow)
0.4+	(orange)
0.0+	(red)

Crystal $\overline{S^2}$=0.84   Major $\overline{S^2}$=0.64   Minor $\overline{S^2}$=0.66
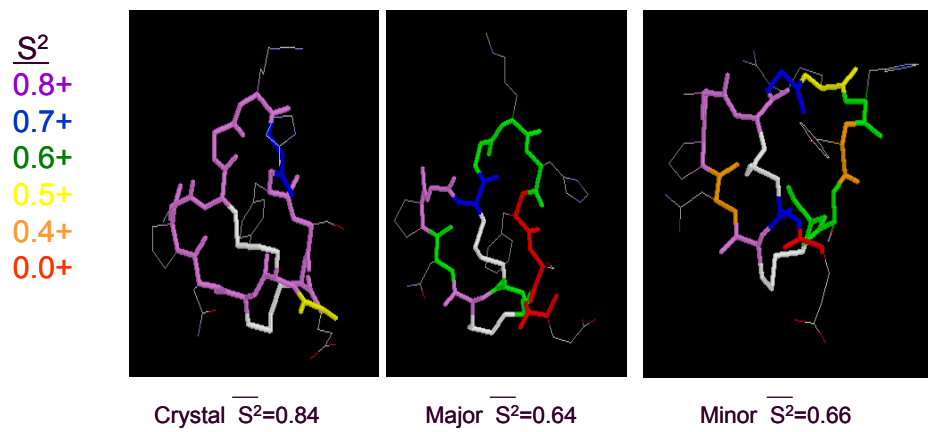
Figure 46: Backbone NH order parameter

Crystal   Solution (Major)
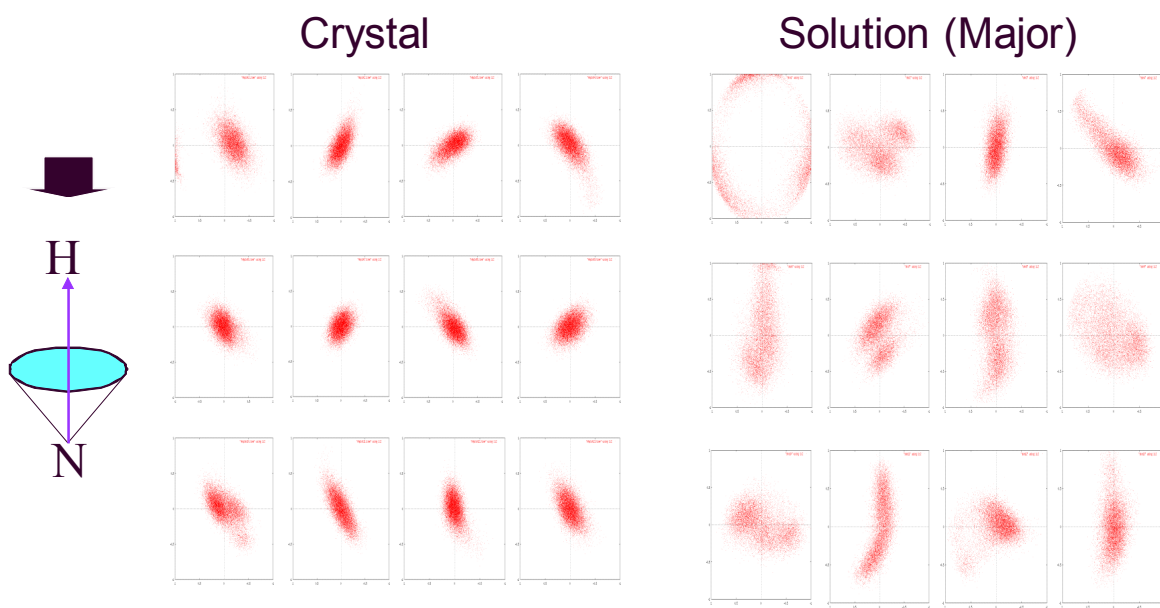
Figure 47: NH bond vector orientation and sampling
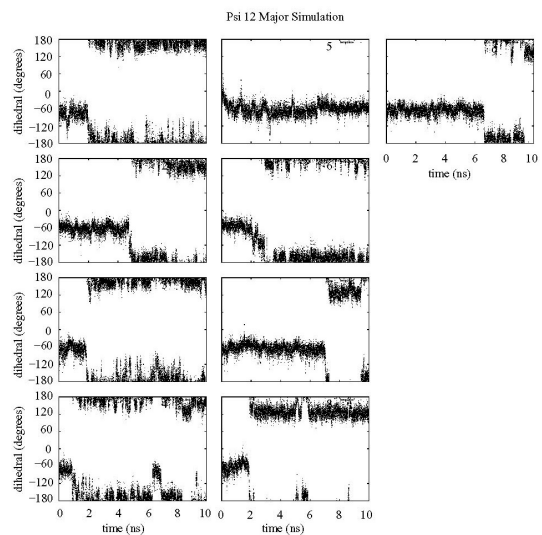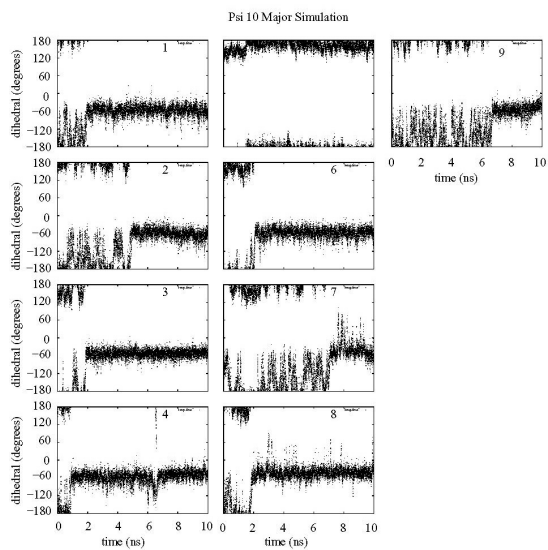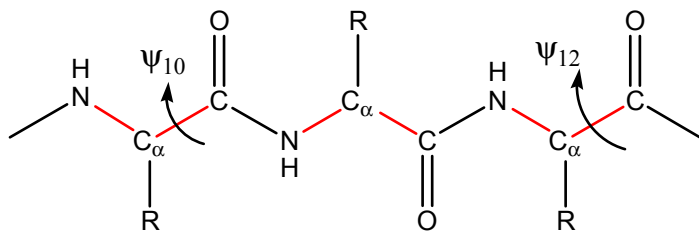
Figure 48: Dihedral (Ψ) of residue 10 and 12 of the major conformation

7.3. *Chloroperoxidase epoxidation*

I made contributions to an MD study of CPO aimed at elucidating the mechanism

of the epoxidation reaction, which takes place at the active site. My contribution to the

project was development of parameters for the sugar molecules and help with

development of parameters for the substrate, cis-β-methylstyrene. In addition, I created

patches to connect the heme to the proximal ligand, cysteine, and to connect sugar

molecules to each other and to amino acids with covalent bonds. Project design and

analysis of the results were carried out by other members of the research group, primarily

Alexander Morozov. Below is the abstract from a paper published on this work [74]:

Molecular dynamics simulations of an explicitly solvated cis-β-methylstyrene /chloroperoxidase-Compound I complex are performed to determine the cause of the high enantiospecificity of epoxidation. From the simulations, a two-dimensional free energy potential is calculated to distinguish binding potential wells from which reaction to 1S2R and 1R2S epoxide products may occur. Convergence of the free energy potential is accelerated with an adaptive biasing potential. Analysis of binding is followed by analysis of 1S2R and 1R2S reaction precursor structures in which the substrate, having left the binding wells, places its reactive double bond in steric proximity to the oxyferryl heme center. Structural analysis of binding and reaction precursor conformations is presented. We find that i) a distortion of Glu183 is important for CPO-catalyzed epoxidation as was postulated previously based on experimental results; ii) the free energy of binding does not provide significant differentiation between structures leading to the respective epoxide enantiomers; iii) CPO's enantiospecificity toward cis-β-methylstyrene is likely to be caused by a specific group of residues which form a hydrophobic core surrounding the oxyferryl heme center.

REFERENCES

1.   Shaw, P.D. and L.P. Hager, *An enzymic chlorination reaction.* J. Am. Chem. Soc., 1959. **81**: p. 1011-12.

2.   Shaw, P.D. and L.P. Hager, *Biological chlorination. VI. Chloroperoxidase: a component of the β-ketoadipate chlorinase system.* J. Biol. Chem., 1961. **236**: p. 1626-30.

3.   Sundaramoorthy, M., J. Terner, and T.L. Poulos, *The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid.* Structure (London), 1995. **3**: p. 1367-77.

4.   Neidleman, S.L. and J. Geigert, *Biological halogenation: roles in nature, potential in industry.* Endeavour, 1987. **11**: p. 5-15.

5.   Wang, X., *Special Topics in Biochemistry Lecture Notes, Lecture 8, Slide 13*, 2010.

6.   Casella, L., et al., *The Chloroperoxidase-Catalyzed Oxidation of Phenols. Mechanism, Selectivity, and Characterization of Enzyme-Substrate Complexes.* Biochemistry, 1994. **33**: p. 6377-86.

7.   Thomas, J.A., D.R. Morris, and L.P. Hager, *Chloroperoxidase. VIII. Formation of peroxide and halide complexes and their relation to the mechanism of the halogenation reaction.* J. Biol. Chem., 1970. **245**: p. 3135-42.

8.   Bello-Ramirez, A.M., B.Y. Carreon-Garabito, and A.A. Nava-Ocampo, *A theoretical approach to the mechanism of biological oxidation of organophosphorus pesticides.* Toxicology, 2000. **149**: p. 63-68.

9.   Carlsen, L. and P. Lassen, *Enzymically mediated formation of chlorinated humic acids.* Org. Geochem., 1992. **18**: p. 477-80.

10.   Casey, W.H., *Geochemistry: The fate of chlorine in soils.* Science (Washington, DC, U. S.), 2002. **295**: p. 985-986.

11.   Niedan, V., I. Pavasars, and G. Oberg, *Chloroperoxidase-mediated chlorination of aromatic groups in fulvic acid.* Chemosphere, 2000. **41**: p. 779-785.

12.   Ortiz-Bermudez, P., E. Srebotnik, and K.E. Hammel, *Chlorination and cleavage of lignin structures by fungal chloroperoxidases.* Appl. Environ. Microbiol., 2003. **69**: p. 5015-5018.

13. Reina, R.G., A.C. Leri, and S.C.B. Myneni, *Cl K-edge X-ray Spectroscopic Investigation of Enzymatic Formation of Organochlorines in Weathering Plant Material.* Environ. Sci. Technol., 2004. **38**: p. 783-789.

14. Walter, B. and K. Ballschmiter, *Biohalogenation as a source of halogenated ansioles in air.* Chemosphere, 1991. **22**: p. 557-67.

15. Franssen, M.C.R. and d.P.H.C. Van, *The chlorination of barbituric acid and some of its derivatives by chloroperoxidase.* Bioorg. Chem., 1987. **15**: p. 59-70.

16. Geigert, J., et al., *Haloperoxidases: enzymic synthesis of α,β-halohydrins from gaseous alkenes.* Appl. Environ. Microbiol., 1983. **45**: p. 366-74.

17. Griffin, B.W., *Mechanism of halide-stimulated activity of chloroperoxidase. Evidence for enzymatic formation of free hypohalous acid.* Biochem. Biophys. Res. Commun., 1983. **116**(3): p. 873-9.

18. Griffin, B.W. and P.L. Ashley, *Evidence for a radical mechanism of halogenation of monochlorodimedone catalyzed by chloroperoxidase.* Arch. Biochem. Biophys., 1984. **233**: p. 188-96.

19. Griffin, B.W. and R. Haddox, *Chlorination of NADH: similarities of the hydrogen hypochloride-supported and chloroperoxidase-catalyzed reactions.* Arch. Biochem. Biophys., 1985. **239**: p. 305-9.

20. Lee, T.D., et al., *Neighboring group migration in enzyme-mediated halohydrin formation.* Biochem. Biophys. Res. Commun., 1983. **110**: p. 880-3.

21. Manoj, K.M., *Chlorinations catalyzed by chloroperoxidase occur via diffusible intermediate(s) and the reaction components play multiple roles in the overall process.* Biochim. Biophys. Acta, Proteins Proteomics, 2006. **1764**: p. 1325-1339.

22. Dunford, H.B., et al., *On the mechanism of chlorination by chloroperoxidase.* Arch. Biochem. Biophys., 1987. **252**: p. 292-302.

23. Libby, R.D., T.M. Beachy, and A.K. Phipps, *Quantitating direct chlorine transfer from enzyme to substrate in chloroperoxidase-catalyzed reactions.* J. Biol. Chem., 1996. **271**: p. 21820-21827.

24. Libby, R.D., et al., *Chloroperoxidase halogenation reactions. Chemical versus enzymic halogenating intermediates.* J. Biol. Chem., 1982. **257**(9): p. 5030-7.

25. Murphy, C.D., *New frontiers in biological halogenation.* J. Appl. Microbiol., 2003. **94**: p. 539-548.

26. Ramakrishnan, K., et al., *Stereoselectivity of chloroperoxidase-dependent halogenation.* Biochemistry, 1983. **22**(13): p. 3271-7.

27.   Wagenknecht, H.-A., C. Claude, and W.-D. Woggon, *New enzyme models of chloroperoxidase. Improved stability and catalytic efficiency of iron porphyrinates containing a thiolato ligand.* Helv. Chim. Acta, 1998. **81**: p. 1506-1520.

28.   Wagenknecht, H.-A. and W.-D. Woggon, *Identification of intermediates in the catalytic cycle of chloroperoxidase.* Chem. Biol., 1997. **4**: p. 367-372.

29.   Woggon, W.-D., *Metalloporphyrines as Active Site Analogues - Lessons from Enzymes and Enzyme Models.* Acc. Chem. Res., 2005. **38**: p. 127-136.

30.   Woggon, W.D., H.A. Wagenknecht, and C. Claude, *Synthetic active site analogues of heme-thiolate proteins. Characterization and identification of intermediates of the catalytic cycles of cytochrome P450cam and chloroperoxidase.* J. Inorg. Biochem., 2001. **83**: p. 289-300.

31.   Wannstedt, C., D. Rotella, and J.F. Siuda, *Chloroperoxidase mediated halogenation of phenols.* Bull. Environ. Contam. Toxicol., 1990. **44**(2): p. 282-7.

32.   Fu, H., et al., *Chloroperoxidase-catalyzed asymmetric synthesis: enantioselective reactions of chiral hydroperoxides with sulfides and bromohydroxylation of glycals.* J. Org. Chem., 1992. **57**: p. 7265-70.

33.   Levine, S.D., S.L. Neidleman, and M. Oberc, *An enzymic route to α-bromo steroidal ketones.* Tetrahedron, 1968. **24**: p. 2979-84.

34.   Neidleman, S.L., et al., *The enzymatic halogenation of steroids.* Tetrahedron Lett, 1966. **44**: p. 5337-42.

35.   Liu, K.K.C. and C.H. Wong, *Enzymic halohydration of glycals.* J. Org. Chem., 1992. **57**: p. 3748-50.

36.   Beckwith, J.R. and L.P. Hager, *Biological chlorination. VIII. Late intermediates in the biosynthesis of caldariomycin.* J. Biol. Chem., 1963. **238**: p. 3091-4.

37.   Hager, L.P., et al., *Chloroperoxidase. II. Utilization of halogen anions.* J. Biol. Chem., 1966. **241**: p. 1769-77.

38.   Kuehnel, K., et al., *Crystal structures of chloroperoxidase with its bound substrates and complexed with formate, acetate, and nitrate.* J. Biol. Chem., 2006. **281**: p. 23990-23998.

39.   Karelson, M., U. Maran, and A.R. Katritzky, *Theoretical study of the keto-enol tautomerism in aqueous solutions.* Tetrahedron, 1996. **52**(34): p. 11325-11328.

40. Nagy, P.I., G. Alagona, and C. Ghio, *Theoretical investigation of tautomeric equilibria for isonicotinic acid, 4-pyridone, and acetylacetone in vacuo and in solution.* J. Chem. Theory Comput., 2007. **3**(4): p. 1249-1266.

41. Belova, N.V., et al., *Tautomeric and conformational properties of beta-diketones.* J. Mol. Struct., 2010. **978**(1-3): p. 282-293.

42. Lee, D., et al., *A theoretical study on keto-enol tautomerization involving simple carbonyl derivatives.* J. Comput. Chem., 1997. **18**(1): p. 56-69.

43. Pei, K.M., Y.F. Ma, and X.M. Zheng, *Theoretical study on the intramolecular proton transfer reactions of 3-methyl-5-hydroxyisoxazole and its water complexes.* J. Mol. Struct.-Theochem, 2007. **820**(1-3): p. 107-111.

44. Freitag, M.A., et al., *On the keto-enol tautomerization of malonaldehyde: An effective fragment potential study.* J. Phys. Chem. A, 2007. **111**(9): p. 1659-1666.

45. Cucinotta, C.S., et al., *Ab initio molecular dynamics study of the keto-enol tautomerism of acetone in solution.* ChemPhysChem, 2006. **7**(6): p. 1229-1234.

46. Cramer, C.J., *Essentials of Computational Chemistry: Theories and Models.* 2002: John Wiley & Sons Ltd. 562 pp.

47. Hinchliffe, A., *Molecular Modelling for Beginners.* 2003: Wiley. 428 pp.

48. Andrews, D.H., *The relation between the Raman spectra an the structure of organic molecules.* Phys. Rev., 1930. **36**: p. 544-54.

49. Kettering, C.F., L.W. Shutts, and D.H. Andrews, *A representation of the dynamic properties of molecules by mechanical models.* Phys. Rev., 1930. **36**: p. 531 43.

50. Snyder, R.G. and J.H. Schachtschneider, *A valence force field for saturated hydrocarbons.* Spectrochim. Acta, 1965. **21**: p. 169-95.

51. Brooks, B.R., et al., *CHARMM: The biomolecular simulation program.* J. Comput. Chem., 2009. **30**(10): p. 1545-1614.

52. BioModeling and bioInformatics, E.U.o.T. *Potentials.* 2007; Available from: http://www.pumma.nl/index.php/Theory/Potentials.

53. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B.

Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *GAUSSIAN 98* 1998, Gaussian, Inc.: Pittsburgh PA.

54.   Wiberg, K.B., *A scheme for strain energy minimization. Application to the cycloalkanes.* J. Am. Chem. Soc., 1965. **87**: p. 1070-8.

55.   Fletcher, R. and C.M. Reeves, *Function minimization by conjugate gradients.* The Computer Journal, 1964. **7**(2): p. 149-154.

56.   M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *GAUSSIAN 09*, 2009, Gaussian, Inc.: Wallingford CT.

57.   Y. Shao, L.F.-M., Y. Jung, J. Kussmann, C. Ochsenfeld,   S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko,S. V. Levchenko, D. P. O'Neill, R. A. Distasio Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M., Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Advances in methods and algorithms in a modern quantum chemistry program package.* Phys. Chem. Chem. Phys., 2006. **8**: p. 3172 - 3191.

58.   Fukui, K., *Formulation of the reaction coordinate.* J. Phys. Chem., 1970. **74**(23): p. 4161-4163.

59.   Gonzalez, C. and H.B. Schlegel, *Reaction path following in mass-weighted internal coordinates.* J. Phys. Chem., 1990. **94**(14): p. 5523-5527.

60.   Sheppard, D., R. Terrell, and G. Henkelman, *Optimization methods for finding minimum energy paths.* J. Chem. Phys., 2008. **128**: p. 134106/1-134106/10.

61.   Maragakis, P., et al., *Adaptive nudged elastic band approach for transition state calculation.* J. Chem. Phys., 2002. **117**: p. 4651-4658.

62.   Liu, H., et al., *Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions.* J. Chem. Phys., 2004. **121**: p. 697-706.

63.   Lee, Y.-S., V.W. Pike, and M. Hodoscek, *Identification of the Transition States in the Inversion of 1,4-Benzodiazepines with the Ab Initio Replica Path Method.* J. Phys. Chem. A, 2008. **112**: p. 1604-1611.

64.   Henkelman, G., B.P. Uberuaga, and H. Jonsson, *A climbing image nudged elastic band method for finding saddle points and minimum energy paths.* J. Chem. Phys., 2000. **113**: p. 9901-9904.

65.   Henkelman, G. and H. Jonsson, *Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points.* J. Chem. Phys., 2000. **113**: p. 9978-9985.

66.   Henkelman, G., G. Johannesson, and H. Jonsson, *Methods for finding saddle points and minimum energy paths.* Prog. Theor. Chem. Phys., 2000. **5**: p. 269-302.

67.   Elber, R. and M. Karplus, *A method for determining reaction paths in large molecules: application to myoglobin.* Chem. Phys. Lett., 1987. **139**: p. 375-80.

68.   Czerminski, R. and R. Elber, *Reaction path study of conformational transitions in flexible systems: applications to peptides.* J. Chem. Phys., 1990. **92**: p. 5580-601.

69.   Czerminski, R. and R. Elber, *Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems.* Int. J. Quantum Chem., Quantum Chem. Symp., 1990. **24**: p. 167-86.

70.   Cisneros, G.A., et al., *Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods.* J. Chem. Phys., 2005. **122**: p. 114502/1-114502/7.

71.   Ayala, P.Y. and H.B. Schlegel, *A combined method for determining reaction paths, minima, and transition state geometries.* J. Chem. Phys., 1997. **107**: p. 375-384.

72.   Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* J. Comput. Chem., 2010. **31**: p. 671-690.

73.   Singh, U.C. and P.A. Kollman, *An approach to computing electrostatic charges for molecules.* J. Comput. Chem., 1984. **5**: p. 129-45.

74.   Morozov, A.N., et al., *Enantiospecificity of Chloroperoxidase-Catalyzed Epoxidation: Biased Molecular Dynamics Study of a Cis-β-Methylstyrene/Chloroperoxidase-Compound I Complex.* Biophys. J., 2011. **100**: p. 1066-1075.

APPENDICES

Appendix A: Computer scripts

A1   Scripts used in the study of limited-hydration CPD keto - enol tautomerization


```
# B3LYP/6-31G* opt

Optimization of reactant structure (CPD+HOCl+1H20)

0  1
 c
 c   1 cc2
 c   2 cc3        1 ccc3
 c   3 cc4        2 ccc4        1 dih4
 c   4 cc5        3 ccc5        2 dih5
 o   5 oc6        4 occ6        3 dih6
 o   4 oc7        3 occ7        2 dih7
 h   4 hc8        3 hcc8        2 dih8
 h   4 hc9        3 hcc9        2 dih9
 h   1 hc10       2 hcc10       3 dih10
 h   1 hc11       2 hcc11       3 dih11
 h   2 hc12       3 hcc12       4 dih12
 h   2 hc13       3 hcc13       4 dih13
 h   7 ho14       4 hoc14       3 dih14
 h   7 ho15       4 hoc15       3 dih15
 o   3 oc16       2 occ16       1 dih16
 h  16 ho17       3 hoc17       2 dih17
 o  17 oh18      16 oho18       3 dih18
cl  18 clo19      17 cloh19    16 dih19

cc2        1.545410
cc3        1.502805
ccc3        105.470
cc4        1.458564
ccc4        112.500
dih4          9.623
cc5        1.553092
ccc5        103.369
dih5        -20.776
oc6        1.197888
occ6        124.984
dih6       -158.289
oc7        3.012537
occ7        106.729
dih7         62.621
```


141

```
hc8          1.117716
hcc8         107.667
dih8          85.649
hc9          1.091939
hcc9         116.089
dih9        -151.414
hc10         1.092588
hcc10        113.544
dih10        127.397
hc11         1.095078
hcc11        111.498
dih11       -110.412
hc12         1.096916
hcc12        109.711
dih12        133.326
hc13         1.099154
hcc13        108.637
dih13       -111.324
ho14         0.970819
hoc14        109.561
dih14       -160.888
ho15         0.970299
hoc15        136.088
dih15         60.908
oc16         1.220000
occ16        120.000
dih16        180.000
ho17         1.500000
hoc17        179.000
dih17        180.000
oh18         0.950000
oho18        179.000
dih18        180.000
clo19        1.700000
cloh19       109.471
dih19        180.000
```

Table 26: Optimization script for reactant (or product) structure of CPD + HOCl + explicit water molecule

```
# B3LYP/6-31G* opt=ModRedundant

Scan to locate structure close to transition state

0  1
 c
 c   1 cc2
 c   2 cc3        1 ccc3
 c   3 cc4        2 ccc4        1 dih4
 c   4 cc5        3 ccc5        2 dih5
 o   3 oc6        2 occ6        1 dih6
 o   5 oc7        4 occ7        3 dih7
```

```
 o    4 oc8          3 occ8          2 dih8
 h    4 hc9          3 hcc9          2 dih9
 h    4 hc10         3 hcc10         2 dih10
 h    1 hc11         2 hcc11         3 dih11
 h    1 hc12         2 hcc12         3 dih12
 h    2 hc13         3 hcc13         4 dih13
 h    2 hc14         3 hcc14         4 dih14
 h    6 ho15         3 hoc15         2 dih15
 h    8 ho16         4 hoc16         3 dih16
 h    8 ho17         4 hoc17         3 dih17
 o   15 oh18         6 oho18         3 dih18
cl   18 clo19       15 cloh19        6 dih19


cc2          1.544633
cc3          1.508992
ccc3         104.626
cc4          1.423838
ccc4         112.514
dih4          15.495
cc5          1.514990
ccc5         106.402
dih5         -10.465
oc6          1.280959
occ6         124.332
dih6        -164.197
oc7          1.211892
occ7         124.205
dih7         174.516
oc8          2.559776
occ8         123.858
dih8          87.988
hc9          1.282261
hcc9         115.146
dih9          94.779
hc10         1.093171
hcc10        117.758
dih10       -147.959
hc11         1.096993
hcc11        111.438
dih11        103.568
hc12         1.092699
hcc12        114.102
dih12       -135.021
hc13         1.094783
hcc13        110.746
dih13        139.816
hc14         1.099688
hcc14        108.489
dih14       -103.953
ho15         1.700000
hoc15        114.173
dih15          1.528
ho16         0.979979
hoc16         92.849
dih16       -143.860
```

```
ho17          0.974744
hoc17         119.979
dih17         104.314
oh18          0.950000
oho18         179.000
dih18         180.000
clo19         1.700000
cloh19        109.471
dih19         180.000


8 9 1.32 S 5 -0.1
6 15 1.72 S 5 -0.1


```

Table 27: Script for scan of potential energy surface to locate structure close to the transition state

```
# B3LYP/6-31G* opt=(TS,CalcFC) IOp(1/8=10)

Transition state optimization

0  1
 c
 c    1 cc2
 c    2 cc3        1 ccc3
 c    3 cc4        2 ccc4       1 dih4
 c    4 cc5        3 ccc5       2 dih5
 o    4 oc6        3 occ6       2 dih6
 o    1 oc7        2 occ7       3 dih7
 o    6 oo8        4 ooc8       3 dih8
cl    8 clo9       6 cloo9      4 dih9
 o    5 oc10       4 occ10      3 dih10
 h   10 ho11       5 hoc11      4 dih11
 h    5 hc12       4 hcc12      3 dih12
 h    2 hc13       3 hcc13      4 dih13
 h    2 hc14       3 hcc14      4 dih14
 h    3 hc15       2 hcc15      1 dih15
 h    3 hc16       2 hcc16      1 dih16
 h    6 ho17       4 hoc17      3 dih17
 h   10 ho18       5 hoc18      4 dih18
 h   10 ho19       5 hoc19      4 dih19

cc2          1.524179
cc3          1.542551
ccc3         104.650
cc4          1.527352
ccc4         104.750
dih4         -18.017
cc5          1.453226
ccc5         110.078
dih5          17.393
oc6          1.242614
occ6         124.139
```

```
dih6         -162.496
oc7            1.233933
occ7         125.202
dih7         -158.458
oo8            2.330560
ooc8         110.574
dih8             3.363
clo9           1.725564
cloo9        104.618
dih9         103.668
oc10           2.501081
occ10        131.864
dih10          89.754
ho11           1.220000
hoc11          14.063
dih11          30.876
hc12           1.093236
hcc12        116.890
dih12        -145.272
hc13           1.098553
hcc13        110.979
dih13          98.138
hc14           1.093351
hcc14        115.038
dih14        -139.992
hc15           1.091907
hcc15        115.884
dih15        -139.659
hc16           1.098098
hcc16        111.551
dih16          98.877
ho17           1.220000
hoc17        113.730
dih17           3.751
ho18           1.002126
hoc18          80.741
dih18        -136.840
ho19           0.973841
hoc19        112.608
dih19        117.714
```

Table 28: Transition state optimization script

```
# B3LYP/6-31G* SCRF=PCM

Optimization of reactant/product

0  1
 c
 c    1 cc2
 c    2 cc3          1 ccc3
 c    3 cc4          2 ccc4          1 dih4
```

```
 c    4 cc5          3 ccc5         2 dih5
 o    2 oc6          3 occ6         4 dih6
 o    5 oc7          4 occ7         3 dih7
 o    7 oo8          5 ooc8         4 dih8
cl    8 clo9         7 cloo9        5 dih9
 o    6 oo10         2 ooc10        3 dih10
 h    1 hc11         2 hcc11        3 dih11
 h    1 hc12         2 hcc12        3 dih12
 h    3 hc13         2 hcc13        1 dih13
 h    3 hc14         2 hcc14        1 dih14
 h    4 hc15         3 hcc15        2 dih15
 h    4 hc16         3 hcc16        2 dih16
 h   10 ho17         6 hoo17        2 dih17
 h   10 ho18         6 hoo18        2 dih18
 h    8 ho19         7 hoo19        5 dih19


cc2          1.522321
cc3          1.532052
ccc3         110.030
cc4          1.538389
ccc4         106.565
dih4          -1.736
cc5          1.528754
ccc5         106.324
dih5           9.117
oc6          1.215396
occ6         124.391
dih6         179.531
oc7          1.219052
occ7         125.165
dih7         167.039
oo8          2.763081
ooc8         108.786
dih8          -5.875
clo9         1.725303
cloo9        103.543
dih9          86.440
oo10         2.895279
ooc10         99.285
dih10        153.329
hc11         1.096825
hcc11        111.853
dih11       -131.969
hc12         1.100739
hcc12        109.515
dih12        111.013
hc13         1.095062
hcc13        109.525
dih13       -125.427
hc14         1.096736
hcc14        107.722
dih14        119.290
hc15         1.098302
hcc15        112.200
dih15       -107.804
```

```
hc16        1.093686
hcc16       115.025
dih16       131.078
ho17        0.975757
hoo17        18.144
dih17       173.629
ho18        0.968553
hoo18       114.146
dih18       113.727
ho19        0.991861
hoo19         9.666
dih19       170.874
```

Table 29: Single point energy calculation script using implicit solvent environment (PCM), of gas-phase optimized structure of reactant (or product) of CPD + HOCl + explicit water molecule

```
#HF/6-31G* NoSymm opt POP=(MK,DIPOLE)

Calc for initial partial atomic charge guess and target dipole

0 1
 c
 c    1 cc2
 c    2 cc3         1 ccc3
 c    3 cc4         2 ccc4          1 dih4
 c    4 cc5         3 ccc5          2 dih5
 o    2 oc6         3 occ6          4 dih6
 o    5 oc7         4 occ7          3 dih7
 h    3 hc8         2 hcc8          6 dih8
 h    3 hc9         2 hcc9          6 dih9
 h    4 hc10        3 hcc10         2 dih10
 h    4 hc11        3 hcc11         2 dih11
 h    1 hc12        2 hcc12         6 dih12
 h    1 hc13        2 hcc13         6 dih13

cc2          1.521729
cc3          1.528426
ccc3         109.193
cc4          1.530854
ccc4         105.445
dih4          19.512
cc5          1.528425
ccc5         105.445
dih5         -23.418
oc6          1.221537
occ6         125.081
dih6        -160.564
oc7          1.221538
occ7         125.082
dih7        -160.561
hc8          1.097780
hcc8         107.519
dih8          80.770
hc9          1.092772
hcc9         110.168
dih9         -35.822
hc10         1.092771
hcc10        115.075
dih10       -145.032
hc11         1.097780
hcc11        111.148
dih11         92.796
hc12         1.096378
hcc12        111.170
dih12        -66.825
hc13         1.096379
```

```
hcc13          111.260
dih13           51.984
```

Table 30: Parameterization script (QM): Initial partial atomic guess and target dipole calculation

```
#HF/6-31G* Opt

Calc for water - CPD(H12) interaction distance and energy

0 1
 c
 c    1 cc2
 c    2 cc3           1 ccc3
 c    3 cc4           2 ccc4           1 dih4
 c    4 cc5           3 ccc5           2 dih5
 o    2 oc6           3 occ6           4 dih6
 o    5 oc7           4 occ7           3 dih7
 h    3 hc8           2 hcc8           6 dih8
 h    3 hc9           2 hcc9           6 dih9
 h    4 hc10          3 hcc10          2 dih10
 h    4 hc11          3 hcc11          2 dih11
 h    1 hc12          2 hcc12          6 dih12
 h    1 hc13          2 hcc13          6 dih13
 o   12 oh14          1 ohc14          2 dih14
 h   14 ho15         12 hoh15          1 dih15
 h   14 ho16         12 hoh16          1 dih16


cc2          1.521729
cc3          1.528426
ccc3         109.193
cc4          1.530854
ccc4         105.445
dih4          19.512
cc5          1.528425
ccc5         105.445
dih5         -23.418
oc6          1.221537
occ6         125.081
dih6        -160.564
oc7          1.221538
occ7         125.082
dih7        -160.561
hc8          1.097780
hcc8         107.519
dih8          80.770
hc9          1.092772
hcc9         110.168
dih9         -35.822
hc10         1.092771
hcc10        115.075
dih10       -145.032
hc11         1.097780
```

```
hcc11          111.148
dih11           92.796
hc12           1.096378
hcc12          111.170
dih12          -66.825
hc13           1.096379
hcc13          111.260
dih13           51.984
oh14           2.100000
ohc14          179.000
dih14          180.000
ho15           0.950000
hoh15          109.471
dih15          180.000
ho16           0.950000
hoh16          109.471
dih16          300.000
```

Table 31: Parameterization script (QM): Target water-CPD (H12) interaction energy and distance calculation

```
* From tutorial on A. MacKerell's website
* Modified by Cassian
* Water interactions
*

ioformat extended

set pdbdir .

!### Set Variables ###
set residue cpd
set zeed 1 C1  1 C2   1 C3 ! seed to create molecule
set frist none ! first capping residue
set lsat none  ! last capping residue
set bombover 0 !needs to be set to -1 for 3-membered rings
set ffsuffix _cpd

set stage 5 !4=do everything with mp2 geometry for h2o inter
            !5=do everything with charmm geometry for h2o inter


!### Things that usually should not be changed. ####
set dontmini 0 ! Do not minimize main conformation. Should normally
not be used.

system "./do_remas"
open unit 10 read form name top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name par_all22_prot.prm
read param card flex unit 10
```

```
close unit 10
open unit 10 read card name top_all36_cgenff_remas@ffsuffix.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name par_all36_cgenff_remas@ffsuffix.prm
read para card flex unit 10 append
close unit 10
if @?str eq 1 stream @str
stream toppar_dum_nobel_gases.str

read sequence card
* @residue
*
1
@residue

bomlev @bombover
generate @residue first @frist last @lsat setup warn
bomlev 0

! For testing and debugging purposes
   open unit 30 write form name test.psf
   write psf card unit 30
   close unit 30

ic param
   ic seed @zeed
   ic build
   ic purge
   ic print

! Write out initial structure for testing and debugging purposes
   open unit 30 write form name @residue_init.pdb
   write coor pdb unit 30
   close unit 30

!minimize structure and calculate dipole moment
if @dontmini eq 1 goto skip_min
   MINI CONJ nstep 200 nprint 20 inbfrq 1000 cutnb 999.
   MINI NRAP nstep 50 tolgrd 0.00001
   prnlev 3
   coor dipole oxyz select segid @residue end

!save minimized geometry
   coor print
   open unit 30 write form name @residue_min.crd
   write coor card unit 30

   open unit 30 write form name @pdbdir/@residue_min.pdb
   write coor pdb unit 30
   close unit 30
label skip_min
if @stage lt 4 stop

!save coordinates in comparison set
```

```
coor copy comp
ic save
ic fill
ic print

!water interactions
if @stage eq 4 then
  open unit 50 write form name @residue_water_constr.ene
 else
  open unit 50 write form name @residue_water.ene
 endif

!variables for averages etc.
set count 0
set asum 0.0
set sum 0.0
set sum2 0.0

!read QM optimized coordinates into the comparison set
!and align charmm optimized structure to allow for the
!components of the dipole moment to be checked.
open unit 20 read form name @residue_mp2.crd
if @stage eq 4 then
  read coor card unit 20
 else
  read coor card unit 20 comp
  coor orient rms
  coor orient mass noro
  coor print
  open unit 30 write form name @residue_min_orient.pdb
  write coor pdb unit 30
 endif
coor dipole oxyz select segid @residue end

! For testing and debugging purposes
open unit 30 write form name test.pdb
write coor pdb unit 30
* @residue
*

! Dipole values below from QM calculation for comparison
echu 50
echo @residue water interaction
if @stage eq 4 echo ######## USING MP2 GEOMETRY!!! ########
echo QM dipole (Debye)
echo mp2/6-31g*: X=      1.4581    Y=    -0.4891    Z=      0.0853
Tot=     1.5404
echo hf/6-31g*: X=      1.5098    Y=     0.1997    Z=      1.1617
Tot=     1.9153
echo empirical dipole: ?xdip ?ydip ?zdip ?rdip
echo

ic delete sele all end

!1) C3H32...OHH  90.0 degrees
```

```
set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H31   1 C3    1 H32    3 dum   0.0
dihe 1 C3    1 H32    3 dum   4 dum   0.0
dihe 4 dum   3 dum   1 H32    2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum   1 H32    2 oh2   2 h1    90.0
dihe 3 dum   1 H32    2 oh2   2 h2   -90.0
bond 1 H32    3 dum   1.
bond 3 dum   4 dum   1.
bond 1 H32    2 oh2   @d !varied
bond 2 oh2   2 h1    0.9572
bond 2 oh2   2 h2    0.9572
angl 1 C3    1 H32    3 dum   90.0
angl 1 H32    3 dum   4 dum   90.0
angl 3 dum   1 H32    2 oh2   90.0
angl 1 H32    2 oh2   2 h1   127.74
angl 1 H32    2 oh2   2 h2   127.74
end


ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy
```

```
label loop_1

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_1
set u @m

ic edit
bond 2 oh2    1 H32    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_1
label done_1

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat1.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -3.0084  !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.47
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
```

```
echo 1) C3H32..OHH  90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!2) C3H31...OHH  90.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H32   1 C3   1 H31   3 dum   0.0
dihe 1 C3   1 H31   3 dum  4 dum   0.0
dihe 4 dum  3 dum  1 H31   2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum  1 H31   2 oh2  2 h1   90.0
dihe 3 dum  1 H31   2 oh2  2 h2  -90.0
bond 1 H31   3 dum  1.
bond 3 dum  4 dum  1.
bond 1 H31   2 oh2  @d !varied
bond 2 oh2  2 h1   0.9572
bond 2 oh2  2 h2   0.9572
angl 1 C3   1 H31   3 dum  90.0
angl 1 H31   3 dum  4 dum  90.0
angl 3 dum  1 H31   2 oh2  90.0
angl 1 H31   2 oh2  2 h1  127.74
angl 1 H31   2 oh2  2 h2  127.74
end
```

```
ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_2

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_2
set u @m

ic edit
bond 2 oh2   1 H31    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_2
label done_2

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat2.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -2.5196   ! From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.47
```

```
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 2) C3H31..OHH  90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!3) C2H22...OHH  90.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H21   1 C2    1 H22   3 dum   0.0
dihe 1 C2    1 H22   3 dum   4 dum   0.0
dihe 4 dum   3 dum   1 H22   2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum   1 H22   2 oh2   2 h1    90.0
dihe 3 dum   1 H22   2 oh2   2 h2   -90.0
bond 1 H22   3 dum   1.
bond 3 dum   4 dum   1.
bond 1 H22   2 oh2   @d !varied
```

```
bond 2 oh2   2 h1     0.9572
bond 2 oh2   2 h2     0.9572
angl 1 C2     1 H22    3 dum   90.0
angl 1 H22    3 dum   4 dum   90.0
angl 3 dum   1 H22    2 oh2   90.0
angl 1 H22    2 oh2   2 h1   127.74
angl 1 H22    2 oh2   2 h2   127.74
end


ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_3

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_3
set u @m

ic edit
bond 2 oh2    1 H22    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_3
label done_3

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat3.pdb
```

```
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -2.5196 !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.47
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 3) C2H22..OHH  90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!4) C2H21...OHH  90.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H22   1 C2   1 H21   3 dum   0.0
dihe 1 C2    1 H21   3 dum  4 dum   0.0
```

```
dihe 4 dum   3 dum   1 H21    2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum   1 H21    2 oh2   2 h1    90.0
dihe 3 dum   1 H21    2 oh2   2 h2   -90.0
bond 1 H21    3 dum   1.
bond 3 dum   4 dum   1.
bond 1 H21    2 oh2   @d !varied
bond 2 oh2   2 h1    0.9572
bond 2 oh2   2 h2    0.9572
angl 1 C2    1 H21    3 dum   90.0
angl 1 H21    3 dum   4 dum   90.0
angl 3 dum   1 H21    2 oh2   90.0
angl 1 H21    2 oh2   2 h1   127.74
angl 1 H21    2 oh2   2 h2   127.74
end


ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_4

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_4
set u @m

ic edit
bond 2 oh2    1 H21    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

prnlev -5

incr d by 0.01
```

```
if @d lt 4.5 goto loop_4
label done_4

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat4.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -3.0084    !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.47
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 4) C2H21..OHH  90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!5) C5H52...OHH   0.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum
```

```
generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H51   1 C5    1 H52    3 dum   0.0
dihe 1 C5    1 H52    3 dum   4 dum   0.0
dihe 4 dum   3 dum   1 H52    2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum   1 H52    2 oh2   2 h1    90.0
dihe 3 dum   1 H52    2 oh2   2 h2   -90.0
bond 1 H52    3 dum   1.
bond 3 dum   4 dum   1.
bond 1 H52    2 oh2   @d !varied
bond 2 oh2   2 h1   0.9572
bond 2 oh2   2 h2   0.9572
angl 1 C5    1 H52    3 dum   90.0
angl 1 H52    3 dum   4 dum   90.0
angl 3 dum   1 H52    2 oh2   90.0
angl 1 H52    2 oh2   2 h1   127.74
angl 1 H52    2 oh2   2 h2   127.74
end


ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_5

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_5
set u @m

ic edit
bond 2 oh2   1 H52    @d
end
```

```
!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

!prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_5
label done_5

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat5.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -3.2900  !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.34
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 5) C5H52..OHH  90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!6) C5H51...OHH   0.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral
```

```
join @residue tip3 renumber

read sequence card
* dummy
*
2
dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 H52   1 C5    1 H51    3 dum    0.0
dihe 1 C5    1 H51    3 dum   4 dum    0.0
dihe 4 dum   3 dum   1 H51    2 oh2 180.0
end
ic fill preserve
ic edit
dihe 3 dum   1 H51    2 oh2   2 h1    90.0
dihe 3 dum   1 H51    2 oh2   2 h2   -90.0
bond 1 H51    3 dum   1.
bond 3 dum   4 dum   1.
bond 1 H51    2 oh2   @d !varied
bond 2 oh2   2 h1    0.9572
bond 2 oh2   2 h2    0.9572
angl 1 C5    1 H51    3 dum   90.0
angl 1 H51    3 dum   4 dum   90.0
angl 3 dum   1 H51    2 oh2   90.0
angl 1 H51    2 oh2   2 h1   127.74
angl 1 H51    2 oh2   2 h2   127.74
end


ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_6

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
```

```
set x ?vdw
set m ?ener

if @u lt @m goto done_6
set u @m

ic edit
bond 2 oh2    1 H51    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

!prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_6
label done_6

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat6.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -3.2900  !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.34
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)

calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 6) C5H51..OHH   90.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!7) O1...OHH    0.0 degrees
```

```
set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
3
dum dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!obtain the inner angle.
!quick 3 1 12
!calculate the outer angle
!calc oangle = ( 360.0 - ?thet ) / 2.

ic edit
dihe 1 C2    1 C5    1 C1    3 dum 180.0
dihe 3 dum  1 C1    1 O1    4 dum   0.0
dihe 1 C1    1 O1    4 dum  5 dum   0.0
dihe 5 dum  4 dum  1 O1    2 h1 180.0
end
ic fill preserve
ic edit
dihe 4 dum  1 O1    2 h1   2 oh2    0.0
dihe 1 O1    2 h1   2 oh2   2 h2 180.0
!dihe 5 dum  2 h1   2 oh2   2 h2    0.0
dihe 5 dum  2 h1   2 oh2   2 h2  -90.0
bond 1 C1    3 dum  1.
bond 1 O1    4 dum  1.
bond 4 dum  5 dum  1.
bond 1 O1    2 h1   @d !varied
bond 2 oh2  2 h1    0.9572
bond 2 oh2  2 h2    0.9572
angl 1 C5    1 C1    3 dum 127.9075
angl 1 C1    1 O1    4 dum  90.0
angl 1 O1    4 dum  5 dum  90.0
angl 4 dum  1 O1    2 h1   90.0
angl 1 O1    2 h1   2 oh2   180.0
angl 2 h1    2 oh2  2 h2   109.47
end

ic print
ic build
```

```
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_7

inter -
   sele resn @residue end -
   sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_7
set u @m

ic edit
bond 2 h1   1 O1   @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

!prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_7
label done_7

prnlev 5

delete atom sele type dum end

open unit 30 write form name @residue_wat7.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -4.1479 !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.11
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
```

```
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)


calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 7) O1..HOH    0.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!8) O4...OHH    0.0 degrees

set d 1.60 ! initial interaction distance

read sequence card
* water
*
1
tip3

generate tip3 first none last none setup warn noangle nodihedral

join @residue tip3 renumber

read sequence card
* dummy
*
3
dum dum dum

generate dum first none last none setup warn noangle nodihedral

join @residue dum renumber

!preparation of IC table for surface of interest

ic edit
dihe 1 C5    1 C3    1 C4    3 dum 180.0
dihe 3 dum  1 C4    1 O4    4 dum   0.0
dihe 1 C4    1 O4    4 dum  5 dum   0.0
dihe 5 dum  4 dum   1 O4    2 h1  180.0
end
ic fill preserve
ic edit
dihe 4 dum  1 O4    2 h1    2 oh2   0.0
dihe 1 O4    2 h1    2 oh2   2 h2 180.0
!dihe 5 dum  2 h1    2 oh2   2 h2   0.0
dihe 5 dum  2 h1    2 oh2   2 h2   90.0
bond 1 C4    3 dum  1.
bond 1 O4    4 dum  1.
bond 4 dum  5 dum  1.
```

```
bond 1 O4   2 h1   @d !varied
bond 2 oh2  2 h1   0.9572
bond 2 oh2  2 h2   0.9572
angl 1 C3   1 C4   3 dum 127.9075
angl 1 C4   1 O4   4 dum  90.0
angl 1 O4   4 dum  5 dum  90.0
angl 4 dum  1 O4   2 h1   90.0
angl 1 O4   2 h1   2 oh2  180.0
angl 2 h1   2 oh2  2 h2   109.47
end

ic print
ic build
coor print

update cutnb 999.0 ctofnb 998.0 ctonnb 997.0

set u 1000. ! minimum energy


label loop_8

inter -
    sele resn @residue end -
    sele resn tip3 end

!check for lowest energy and store data
set n @d
set o ?elec
set x ?vdw
set m ?ener

if @u lt @m goto done_8
set u @m

ic edit
bond 2 h1   1 O4    @d
end

!initialize water coordinates
coor init sele resn tip3 .or. resn dum end

ic build

!prnlev -5

incr d by 0.01

if @d lt 4.5 goto loop_8
label done_8

prnlev 5

delete atom sele type dum end
```

```
open unit 30 write form name @residue_wat8.pdb
write coor pdb unit 30
* @residue
*

!accumulate statistics
calc count = @count + 1
set ai -4.1479 !  From QM calc and scaled by 1.16 for neutral
compound
set aidist 2.11
calc diff = @m - @ai
calc sum = @sum + @diff
calc diff2 = abs(@diff*@diff)
calc sum2 = @sum2 + @diff2
calc asum = @asum + abs(@diff)


calc distdiff = @n - @aidist

! write ab initio and empirical information to unit 50
echo 8) O4..HOH   0.0 degrees
echo a.i. @ai @aidist
echo emp. @m @o @x @n
echo ene diff: @diff dist diff: @distdiff
echo

delet atom sele resn tip3 .or. resn dum end

!calculate statistics from all water calculations
!average difference
calc avediff = @sum/@count
calc avediff2 = @sum2/@count
calc rmsdif = sqrt(@avediff2 -(@avediff*@avediff))
!AAE; average absolute error
calc aae = @asum/@count

echo avediff, rmsdiff, average absolute error
echo @avediff @rmsdif @aae

stop
```

Table 32: Parameterization script (MM): Water – CPD dipole and optimal interaction energies and distances

```
%CHK=CPD_OPT_FREQ
#MP2/6-31G* NOSYMM OPT=TIGHT POP=(MK,DIPOLE) DENSITY=CURRENT

Initial opt

0 1
 C    -0.01119     0.01187    -0.18971
 O    -0.74465    -0.97376    -0.20661
```

```
  C      1.51541     -0.02896     -0.12172
  H      1.91154     -0.52385     -1.02065
  H      1.84573     -0.63344      0.73519
  C      1.98602      1.42895     -0.02109
  H      2.80533      1.65720     -0.71746
  H      2.35965      1.66407      0.98651
  C      0.76424      2.29273     -0.32771
  O      0.77831      3.51602     -0.44209
  C     -0.48850      1.45196     -0.17369
  H     -1.19878      1.64493     -0.98696
  H     -0.99637      1.68132      0.77091


--LINK1--
%CHK=CPD_OPT_FREQ_MP2-2
#MP2/6-31G* NOSYMM GEOM=ALLCHECK FREQ IOP(7/33=1)

Freq calc for force constants to get target normal mode freq and
contribution

0 1



```

Table 33: Frequency calculation script for force constants to get target normal mode frequency and percent contribution

```
* MOLVIB analysis of Gaussian second derivatives
* NOTE: The MOLVIB module is very sensitive to the format
*

ioformat extended

MOLVIB NDI1 33 NDI2 33 NDI3 44 NATOM  13  NOTO
GFX
PRNT    1
DIM    33   33   44
CART
    0.001835    0.015210   -0.040459    12.000
   -0.705793   -0.955884    0.179520    16.000
    1.522114   -0.022283   -0.193538    12.000
    1.743134   -0.349330   -1.217897     1.000
    1.950783   -0.765002    0.483782     1.000
    1.990997    1.416118    0.040270    12.000
    2.848032    1.727522   -0.561949     1.000
    2.255480    1.580291    1.092989     1.000
    0.780126    2.294838   -0.272340    12.000
    0.826424    3.487778   -0.531000    16.000
   -0.482586    1.448084   -0.207391    12.000
   -1.083215    1.559459   -1.117822     1.000
   -1.120590    1.752238    0.630754     1.000
```

```
IC
 1  1   2   0   0 !1 R1     R(1,2)       C1 = O1
 1  1   3   0   0 !2 R2     R(1,3)       C1 - C2
 1  1  11   0   0 !3 R3     R(1,11)      C1 - C5
 1  3   4   0   0 !4 R4     R(3,4)       C2 - H21
 1  3   5   0   0 !5 R5     R(3,5)       C2 - H22
 1  3   6   0   0 !6 R6     R(3,6)       C2 - C3
 1  6   7   0   0 !7 R7     R(6,7)       C3 - H31
 1  6   8   0   0 !8 R8 R(6,8)   C3 - H32
 1  6   9   0   0 !9 R9     R(6,9)       C3 - C4
 1  9  10   0   0 !10 R10 R(9,10)        C4 = O4
 1  9  11   0   0 !11 R11 R(9,11)        C4 - C5
 1 11  12   0   0 !12 R12 R(11,12)       C5 - H51
 1 11  13   0   0 !13 R13 R(11,13)       C5 - H52
 2  2   1   3   0 !14 A1     A(2,1,3)     O1 - C1 - C2    1b2
 2  2   1  11   0 !15 A2     A(2,1,11)    O1 - C1 - C5    1b1
 2  3   1  11   0 !16 A3     A(3,1,11)    C2 - C1 - C5    r1a4
 2  1   3   4   0 !17 A4     A(1,3,4)     C1 - C2 - H21   2b3
 2  1   3   5   0 !18 A5     A(1,3,5)     C1 - C2 - H22   2b4
 2  1   3   6   0 !19 A6     A(1,3,6)     C1 - C2 - C3    r1a5 gamma
 2  4   3   5   0 !20 A7     A(4,3,5)     H21 - C2 - H22 2 alpha
 2  4   3   6   0 !21 A8     A(4,3,6)     H21 - C2 - C3    2b1
 2  5   3   6   0 !22 A9     A(5,3,6)     H22 - C2 - C3    2b2
 2  3   6   7   0 !23 A10    A(3,6,7)     C2 - C3 - H31    3b3
 2  3   6   8   0 !24 A11    A(3,6,8)     C2 - C3 - H32    3b4
 2  3   6   9   0 !25 A12    A(3,6,9)     C2 - C3 - C4    r1a1 gamma
 2  7   6   8   0 !26 A13    A(7,6,8)     H31 - C3 - H32 3 alpha
 2  7   6   9   0 !27 A14    A(7,6,9)     H31 - C3 - C4    3b1
 2  8   6   9   0 !28 A15    A(8,6,9)     H32 - C3 - C4    3b2
 2  6   9  10   0 !29 A16    A(6,9,10)    C3 - C4 - O4     4b1
 2  6   9  11   0 !30 A17    A(6,9,11)    C3 - C4 - C5    r1a2
 2 10   9  11   0 !31 A18    A(10,9,11)   O4 - C4 - C5     4b2
 2  1  11   9   0 !32 A19    A(1,11,9)    C1 - C5 - C4    r1a3 gamma
 2  1  11  12   0 !33 A20    A(1,11,12)   C1 - C5 - H51    5b1
 2  1  11  13   0 !34 A21    A(1,11,13)   C1 - C5 - H52    5b2
 2  9  11  12   0 !35 A22    A(9,11,12)   C4 - C5 - H51    5b3
 2  9  11  13   0 !36 A23    A(9,11,13)   C4 - C5 - H52    5b4
 2 12  11  13   0 !37 A24    A(12,11,13) H51 - C5 - H52 5 alpha
 3  1   3  11   2 !38 I1     I(1,3,11,2)  C1-C2-C5-O1      C1
 3  9  11   6  10 !39 I1     I(9,11,6,10) C5-C1-C2-O4      C4
 4 11   1   3   6 !40 D1     D(11,1,3,6)  C5-C1-C2-C3     r1t4
 4  1   3   6   9 !41 D2     D(1,3,6,9)   C1-C2-C3-C4     r1t5
 4  3   6   9  11 !42 D3     D(3,6,9,11)  C2-C3-C4-C5     r1t1
 4  6   9  11   1 !43 D4     D(6,9,11,1)  C3-C4-C5-C1     r1t2
 4  9  11   1   3 !44 D5     D(9,11,1,3)  C4-C5-C1-C2     r1t3
UMAT     0    1    0               ! row normalization
  1   1  1.            2   2  1.           3   3  1.
  4   4  1.            4   5  1.
  5   4  1.            5   4  -1.
  6   6  1.
  7   7  1.            7   8  1.
  8   7  1.            8   8  -1.
  9   9  1.           10  10  1.
 11  11  1.
 12  12  1.           12  13  1.
```

```
 13 12   1.             13 13  -1.
 14 19   1.             14 20   5.
 15 17   1.             15 18  -1.             15 21   1.             15 22  -1.
 16 17  -1.             16 18  -1.             16 21   1.             16 22   1.
 17 17  -1.             17 18   1.             17 21   1.             17 22  -1.
 18 25   1.             18 26   5.
 19 23   1.             19 24  -1.             19 27   1.             19 28  -1.
 20 23  -1.             20 24  -1.             20 27   1.             20 28   1.
 21 23  -1.             21 24   1.             21 27   1.             21 28  -1.
 22 32   1.             22 37   5.
 23 33   1.             23 34  -1.             23 35   1.             23 36  -1.
 24 33   1.             24 34   1.             24 35  -1.             24 36  -1.
 25 33   1.             25 34  -1.             25 35  -1.             25 36   1.
 26 38   1.
 27 15   1.             27 14  -1.
 28 39   1.
 29 29   1.             29 31  -1.
 30 25   1.             30 30  -0.81            30 32   0.31           30 16   0.31
 30 19  -0.81
 31 30  -1.12           31 32   1.81           31 16  -1.81           31 19   1.12
 32 42   0.31           32 43  -0.81           32 44   1.            32 40  -0.81
 32 41   0.31
 33 42  -1.81           33 43   1.12           33 40  -1.12           33 41   1.81
-1
PED        0    15
  1   sC1=O1
  2   sC1-C2
  3   sC1-C5
  4   sC2-H2sy
  5   sC2-H2as
  6   sC2-C3
  7   sC3-H3sy
  8   sC3-H3as
  9   sC3-C4
 10   sC4=O4
 11   sC4-C5
 12   sC5-H5sy
 13   sC5-H5as
 14   sc-C2H2
 15   r-C2H2
 16   w-C2H2
 17   t-C2H2
 18   sc-C3H2
 19   r-C3H2
 20   w-C3H2
 21   t-C3H2
 22   sc-C5H2
 23   r-C5H2
 24   w-C5H2
 25   t-C5H2
 26   w-C1=O1
 27   d-C1=O1
 28   w-C4=O4
 29   d-C4=O4
 30   d-5RNGsy
```

```
 31  d-5RNGas
 32  t-5RNGsy
 33  t-5RNGas
-1
FMAT    1    1    0            ! Gaussian FX matrix is in input
orientation
 Force constants in Cartesian coordinates:
              1                2                3                4
5
   1  0.643611D+00
   2  0.236491D+00  0.805451D+00
   3 -0.770802D-01 -0.108522D+00  0.263264D+00
   4 -0.318400D+00 -0.293306D+00  0.705900D-01  0.332572D+00
   5 -0.296109D+00 -0.506491D+00  0.983692D-01  0.344365D+00
0.553658D+00
   6  0.692472D-01  0.967589D-01 -0.871856D-01 -0.832153D-01 -
0.115810D+00
   7 -0.169439D+00  0.318536D-01  0.368375D-02 -0.296714D-01 -
0.215520D-01
   8  0.173909D-01 -0.881021D-01 -0.136055D-02 -0.421041D-01
0.352002D-02
   9  0.348225D-02  0.172581D-02 -0.893633D-01  0.108227D-01
0.510280D-02
  10 -0.733210D-02  0.801079D-02  0.366843D-01 -0.464062D-03 -
0.372434D-03
  11  0.634848D-03 -0.134744D-02  0.116934D-02 -0.106853D-02 -
0.118202D-03
  12  0.140842D-02 -0.321319D-03 -0.252530D-02 -0.487130D-04 -
0.259966D-03
  13 -0.134787D-01  0.252618D-01 -0.242071D-01  0.493770D-02 -
0.129191D-02
  14  0.203779D-02 -0.184045D-02 -0.148578D-02 -0.123142D-02
0.667746D-03
  15  0.554168D-02 -0.403406D-02  0.287438D-02 -0.873699D-03 -
0.698227D-03
  16 -0.166763D-01 -0.325877D-01 -0.642492D-02 -0.776981D-02 -
0.149683D-03
  17 -0.184391D-01 -0.620067D-02  0.154042D-02  0.268947D-02
0.124631D-02
  18  0.460571D-03  0.600984D-02  0.266769D-03  0.368527D-02
0.520298D-04
  19 -0.122688D-02 -0.256346D-03  0.676854D-04 -0.841820D-04 -
0.325315D-03
  20  0.286359D-02 -0.684261D-02 -0.673492D-03 -0.348106D-02
0.497678D-04
  21 -0.141565D-02 -0.179596D-02  0.108578D-03 -0.811366D-03 -
0.251657D-03
  22 -0.350794D-03 -0.209015D-03 -0.858540D-03 -0.188155D-03 -
0.230343D-03
  23  0.164796D-02 -0.142444D-02 -0.853379D-03 -0.795777D-03 -
0.281608D-05
  24  0.415373D-03 -0.469615D-03 -0.338904D-03 -0.286168D-03
0.536667D-04
  25  0.265346D-01 -0.248586D-01  0.217737D-02 -0.553549D-02
0.167721D-01
```

```
   26 -0.252890D-01 -0.388526D-01  0.467323D-02  0.335264D-02 -
0.236797D-02
   27  0.489794D-02  0.378506D-02 -0.114741D-02 -0.113154D-02
0.149518D-02
   28 -0.143017D-01  0.560882D-02 -0.819573D-04  0.263160D-02 -
0.722220D-02
   29 -0.788183D-02  0.613259D-02 -0.467272D-03  0.104146D-02 -
0.538102D-02
   30  0.683863D-03 -0.618144D-03  0.202353D-02  0.149466D-03
0.642444D-03
   31 -0.116618D+00  0.389450D-01 -0.452435D-02  0.211636D-01 -
0.305527D-01
   32  0.512841D-01 -0.151697D+00  0.522296D-02 -0.688964D-02 -
0.485713D-01
   33 -0.346672D-02  0.468933D-02 -0.921110D-01  0.109782D-02
0.117843D-01
   34 -0.554912D-02  0.173462D-02 -0.572823D-02  0.242513D-03 -
0.154103D-02
   35  0.162405D-01 -0.133538D-02  0.300695D-01 -0.107804D-02
0.747642D-03
   36  0.315346D-03 -0.355828D-04 -0.854985D-03  0.874142D-04 -
0.112599D-03
   37 -0.677248D-02  0.331252D-02  0.570218D-02  0.565655D-03 -
0.179078D-02
   38  0.191282D-01 -0.745061D-02 -0.276817D-01 -0.149373D-02
0.304232D-02
   39 -0.449006D-02  0.282816D-02  0.498922D-02 -0.659496D-04 -
0.367584D-03
                 6               7               8               9
10
    6  0.493274D-01
    7  0.863396D-02  0.458667D+00
    8  0.944539D-02 -0.671308D-01  0.571449D+00
    9  0.170653D-01  0.172014D-01 -0.344504D-01  0.578994D+00
   10 -0.479862D-02 -0.580467D-01  0.203512D-01  0.603948D-01
0.615450D-01
   11 -0.115974D-02  0.174177D-01 -0.685688D-01 -0.722820D-01 -
0.160573D-01
   12  0.126200D-02  0.499845D-01 -0.715961D-01 -0.286280D+00 -
0.577158D-01
   13  0.290698D-02 -0.912681D-01  0.729754D-01 -0.734666D-01
0.399346D-02
   14 -0.128320D-03  0.731191D-01 -0.176164D+00  0.120416D+00 -
0.449950D-02
   15  0.263988D-04 -0.732942D-01  0.121059D+00 -0.161957D+00 -
0.142046D-01
   16  0.258439D-02 -0.927224D-01 -0.335632D-01 -0.103106D-01
0.241314D-02
   17 -0.173542D-02 -0.328830D-01 -0.184450D+00 -0.185659D-01 -
0.535128D-02
   18  0.701632D-03 -0.258670D-02 -0.211970D-01 -0.857305D-01 -
0.140985D-01
   19  0.570686D-04 -0.272243D-02  0.749229D-03  0.153965D-02
0.627639D-03
   20  0.568902D-03 -0.242386D-01 -0.561428D-02  0.157636D-01
```

```
0.801433D-03
  21  0.196016D-03 -0.742235D-02 -0.183888D-02  0.372903D-02 -
0.303484D-03
  22  0.325049D-03 -0.152622D-02 -0.728631D-03 -0.639794D-02
0.214924D-03
  23  0.624941D-03 -0.810447D-02 -0.641852D-02 -0.327574D-01 -
0.371775D-02
  24 -0.231438D-04 -0.997858D-04  0.329583D-03 -0.165535D-02 -
0.517449D-03
  25 -0.230640D-02  0.106298D-01  0.523737D-02  0.417508D-02 -
0.137768D-02
  26  0.425099D-03  0.185893D-01 -0.343851D-01  0.828822D-02
0.231544D-03
  27  0.175799D-02  0.648766D-02 -0.873153D-03  0.114493D-02
0.504702D-04
  28  0.877512D-03 -0.500167D-02  0.242211D-02 -0.359245D-02
0.329929D-03
  29  0.880118D-03  0.501807D-02 -0.709010D-03  0.284934D-02 -
0.359325D-04
  30 -0.723913D-03 -0.388804D-02  0.474336D-03 -0.111159D-03
0.157395D-03
  31  0.567157D-02 -0.165118D-01  0.235584D-01 -0.297146D-02 -
0.159781D-02
  32  0.108660D-01  0.501342D-02 -0.462486D-02  0.359910D-02 -
0.113561D-03
  33  0.164037D-01  0.176829D-02 -0.503289D-03  0.264746D-01 -
0.602929D-02
  34  0.313376D-03 -0.123056D-02  0.390565D-03  0.594366D-03 -
0.505103D-03
  35 -0.418335D-02  0.119983D-02 -0.213784D-02 -0.479347D-02
0.154357D-02
  36  0.105926D-02  0.292567D-03 -0.154095D-03 -0.894066D-03
0.237080D-03
  37 -0.296826D-03 -0.115595D-02  0.451618D-03 -0.147123D-02
0.199414D-03
  38  0.344697D-02  0.169789D-02 -0.379491D-02  0.510468D-02 -
0.790725D-03
  39  0.133040D-03 -0.761011D-03  0.665258D-03 -0.141656D-02
0.143717D-03
             11              12              13              14
15
  11  0.767915D-01
  12  0.801262D-01  0.307310D+00
  13 -0.590871D-02  0.573665D-02  0.964577D-01
  14  0.722121D-02 -0.709795D-02 -0.809059D-01  0.190079D+00
  15  0.217836D-01 -0.189961D-01  0.766496D-01 -0.127959D+00
0.171711D+00
  16  0.566443D-02  0.507231D-03  0.612635D-02  0.129413D-01
0.459892D-02
  17 -0.992997D-02 -0.668624D-04 -0.973633D-02 -0.155654D-01 -
0.631078D-02
  18 -0.292925D-01 -0.208300D-02  0.890501D-02  0.156378D-01
0.483121D-02
  19  0.544045D-04 -0.155694D-03 -0.219925D-03  0.153276D-03 -
0.242028D-03
```

```
 20   0.221667D-02   0.312738D-03  -0.104758D-02  -0.140019D-02
0.613354D-04
 21   0.110237D-04   0.106507D-02   0.229087D-03  -0.265743D-04
0.140257D-03
 22  -0.384838D-03   0.134050D-04   0.740748D-03   0.266241D-03
0.205000D-03
 23  -0.534917D-02  -0.891875D-03   0.104549D-02   0.205104D-02  -
0.994566D-04
 24  -0.519136D-03   0.115096D-02   0.242193D-03   0.150236D-03
0.111765D-02
 25  -0.152929D-02  -0.670501D-03  -0.444032D-02  -0.455045D-02  -
0.276386D-03
 26  -0.390719D-03  -0.996987D-04  -0.127996D-02  -0.365047D-02  -
0.230085D-02
 27  -0.136409D-02  -0.345696D-03  -0.895183D-03  -0.846489D-03
0.129802D-03
 28   0.444757D-03   0.282017D-03   0.169769D-02   0.211328D-02
0.577973D-03
 29  -0.542521D-03  -0.115433D-03  -0.906777D-03  -0.183235D-02  -
0.716781D-03
 30   0.763825D-03  -0.152931D-05   0.488896D-03   0.785801D-03
0.296360D-03
 31   0.447406D-03   0.407948D-03  -0.410375D-02   0.278751D-03
0.100180D-02
 32   0.113373D-03   0.942788D-04   0.414532D-03   0.401626D-03  -
0.502780D-03
 33   0.872590D-03  -0.643198D-03   0.394115D-02   0.464986D-03  -
0.225226D-03
 34   0.219860D-03   0.868758D-04   0.156310D-04   0.229723D-03
0.129800D-03
 35  -0.189482D-03   0.131292D-03  -0.326911D-04   0.000000D+00  -
0.790170D-04
 36  -0.914851D-04   0.453243D-04  -0.118127D-03  -0.256257D-04
0.869752D-05
 37   0.652878D-04   0.163669D-03  -0.458526D-03   0.478122D-04
0.186132D-03
 38   0.935180D-04  -0.215301D-03   0.141261D-02   0.312884D-04  -
0.203120D-03
 39  -0.175969D-04   0.416793D-04  -0.412586D-03   0.115106D-03
0.428694D-04
               16              17              18              19
20
 16   0.571894D+00
 17   0.724673D-01   0.465043D+00
 18  -0.378625D-01  -0.560860D-02   0.572176D+00
 19  -0.219683D+00  -0.611247D-01   0.127480D+00   0.237014D+00
 20  -0.611944D-01  -0.716679D-01   0.426428D-01   0.676141D-01
0.746130D-01
 21   0.127735D+00   0.432709D-01  -0.138040D+00  -0.134012D+00  -
0.458592D-01
 22  -0.641312D-01  -0.111544D-01  -0.680704D-01   0.823900D-02
0.343068D-02
 23  -0.853247D-02  -0.486087D-01  -0.308201D-01   0.250816D-02
0.125947D-02
 24  -0.594504D-01  -0.368429D-01  -0.300155D+00  -0.608619D-02  -
```

```
0.258534D-02
  25 -0.162045D+00  0.390823D-01 -0.107669D-01 -0.232956D-01
0.147770D-01
  26  0.243923D-01 -0.937078D-01  0.621511D-02 -0.908322D-02
0.651253D-02
  27 -0.122758D-01  0.540509D-02 -0.911532D-01  0.159606D-01 -
0.144089D-01
  28  0.138550D-01  0.330693D-01 -0.704746D-02  0.476299D-02 -
0.188235D-02
  29  0.130206D-01 -0.381940D-01  0.139718D-01 -0.162979D-02
0.101916D-02
  30 -0.196653D-02  0.182470D-01  0.152532D-01 -0.183814D-02
0.142190D-02
  31 -0.256477D-01 -0.744970D-02 -0.678705D-03 -0.235078D-02
0.179655D-02
  32  0.107190D-01  0.489158D-02  0.298485D-02  0.186597D-02 -
0.791767D-03
  33 -0.758232D-02  0.643724D-03  0.260934D-01 -0.314995D-02
0.296684D-02
  34 -0.368487D-02 -0.796603D-03  0.404955D-02 -0.984783D-03
0.720524D-03
  35 -0.247510D-02 -0.169196D-02  0.295920D-02 -0.623685D-03
0.548014D-03
  36  0.857873D-03  0.000000D+00 -0.990941D-03  0.312354D-03 -
0.991142D-04
  37 -0.192712D-02 -0.373228D-03 -0.346943D-02 -0.759583D-04 -
0.159897D-03
  38 -0.702308D-03 -0.116480D-02 -0.355511D-02  0.979142D-04
0.980941D-04
  39 -0.410187D-03  0.234187D-04 -0.117001D-02  0.663159D-04 -
0.112025D-03
              21              22              23              24
25
  21  0.146622D+00
  22  0.272468D-01  0.653862D-01
  23  0.845276D-02  0.109985D-01  0.582061D-01
  24 -0.172798D-01  0.686030D-01  0.394050D-01  0.322054D+00
  25 -0.101360D-01 -0.612777D-02  0.289630D-02 -0.179875D-02
0.473549D+00
  26 -0.973156D-03 -0.303973D-02  0.146542D-02  0.132960D-03
0.287739D-01
  27  0.433836D-02 -0.288046D-01  0.232948D-01 -0.654212D-02
0.169147D-01
  28  0.839421D-03  0.533248D-04  0.109572D-02 -0.284565D-03 -
0.103811D+00
  29 -0.129614D-02  0.143053D-03 -0.107556D-02 -0.140643D-03 -
0.224325D-01
  30 -0.150154D-03  0.299086D-02 -0.363637D-02  0.170140D-02
0.238061D-02
  31 -0.138999D-02 -0.165446D-02  0.790787D-03 -0.402583D-03 -
0.173293D+00
  32  0.418765D-03  0.120388D-02 -0.437436D-03  0.441979D-03 -
0.352390D-01
  33 -0.784960D-03  0.518451D-02 -0.287046D-02 -0.353566D-04 -
0.170334D-02
```

```
 34 -0.404923D-03  0.544985D-03 -0.393999D-03 -0.193777D-03 -
0.159136D-01
 35 -0.134071D-03  0.476723D-03 -0.216469D-03 -0.145765D-04 -
0.913023D-02
 36  0.525107D-04 -0.889925D-04  0.404284D-04  0.606303D-05
0.341239D-02
 37 -0.154557D-03 -0.120062D-02  0.561504D-03 -0.140894D-03 -
0.148733D-01
 38  0.221867D-04 -0.772191D-03  0.551067D-03  0.588108D-04 -
0.979896D-02
 39  0.319495D-05 -0.348196D-03  0.111184D-03  0.000000D+00 -
0.140187D-02
              26             27             28             29
30
 26  0.963939D+00
 27 -0.155551D+00  0.274839D+00
 28 -0.250020D-01  0.237532D-02  0.821285D-01
 29 -0.712441D+00  0.141470D+00  0.286371D-01  0.793657D+00
 30  0.139093D+00 -0.958204D-01 -0.128779D-02 -0.166675D+00
0.597674D-01
 31 -0.229564D-01 -0.971065D-03  0.128733D-01 -0.176033D-01
0.230323D-02
 32 -0.955278D-01 -0.213081D-02 -0.411860D-01 -0.404755D-01
0.983449D-02
 33 -0.625441D-03 -0.916082D-01  0.733890D-02  0.102881D-01
0.165987D-01
 34  0.375541D-02 -0.236227D-01  0.279708D-02  0.189871D-02
0.253504D-02
 35  0.291052D-02 -0.196720D-01  0.121268D-02  0.645698D-03
0.269836D-02
 36  0.241069D-04  0.377010D-02  0.677427D-04 -0.153527D-03
0.298439D-03
 37  0.755519D-02  0.210143D-01  0.198456D-02  0.731084D-03 -
0.270887D-02
 38  0.649626D-02  0.193964D-01  0.688827D-03 -0.803797D-03 -
0.303236D-02
 39  0.698107D-03  0.637152D-03 -0.646684D-04  0.105105D-03
0.868158D-03
              31             32             33             34
35
 31  0.566030D+00
 32 -0.433064D-01  0.452808D+00
 33  0.428433D-02  0.632470D-02  0.543893D+00
 34 -0.122989D+00  0.123682D-01 -0.133908D+00  0.135731D+00
 35  0.147921D-01 -0.490156D-01  0.269710D-01 -0.169360D-01
0.476775D-01
 36 -0.128140D+00  0.231541D-01 -0.237214D+00  0.138408D+00 -
0.254654D-01
 37 -0.135300D+00  0.438656D-01  0.132225D+00  0.115263D-01 -
0.518961D-02
 38  0.412596D-01 -0.670738D-01 -0.610064D-01 -0.164998D-02
0.205644D-02
 39  0.125409D+00 -0.603077D-01 -0.206842D+00  0.177404D-01 -
0.848745D-02
              36             37             38             39
```

```
  36  0.255941D+00
  37 -0.156444D-01  0.147488D+00
  38  0.291891D-02 -0.490771D-01  0.679189D-01
  39 -0.212271D-01 -0.135405D+00  0.647460D-01  0.223940D+00
SCALE
 0.89
END
```

Table 34: Parameterization script (QM): Bond, angle and torsion based on QM calculated force constants for CPD

```
* From tutorial on A. MacKerell's website
* Modified by Cassian
* Quick measurements and molvib
*

ioformat extended

set pdbdir .

!### Set Variables ###
set residue cpd
set zeed 1 C1  1 C2  1 C3 ! seed
set frist none ! first capping residue
set lsat none  ! last capping residue
set bombover 0 !needs to be set to -1 for 3-membered rings

!### Things that usually should not be changed. ####
set dontmini 0 ! Do not minimize main conformation. Should normally
not be used.

open unit 10 read card name top_all36_cgenff_@residue.rtf
read rtf card unit 10
close unit 10
open unit 10 read card name par_all36_cgenff_@residue.prm
read para card flex unit 10
close unit 10

! Read in CPD parameter file in command line
if @?str eq 1 stream @str

read sequence card
* @residue
*
1
@residue

bomlev @bombover
generate @residue first @frist last @lsat setup warn
bomlev 0
```

```
   ic param
   ic seed @zeed
   ic build
   ic purge
   ic print

! Write out initial structure for testing and debugging purposes
   open unit 30 write form name @residue_init.pdb
   write coor pdb unit 30
   close unit 30

!minimize structure and calculate dipole moment
if @dontmini eq 1 goto skip_min
   MINI CONJ nstep 200 nprint 20 inbfrq 1000 cutnb 999.
   MINI NRAP nstep 50 tolgrd 0.00001
   prnlev 3
   coor dipole oxyz select segid @residue end

!save minimized geometry
   coor print
   open unit 30 write form name @residue_min.crd
   write coor card unit 30

   open unit 30 write form name @pdbdir/@residue_min.pdb
   write coor pdb unit 30
   close unit 30
label skip_min

!mp2/6-31g* data
!
!C1 -O1  1.2215
quick  1  2
!C4 -O4  1.2215
quick  9 10
!C1 -C2  1.5284
quick  1  3
!C3 -C4  1.5284
quick  6  9
!C1 -C5  1.5217
quick  1 11
!C4 -C5  1.5217
quick  9 11
!C2 -C3  1.5309
quick  3  6
!C2 -H21 1.0978
quick  3  4
!C3 -H32 1.0978
quick  6  8
!C2 -H22 1.0928
quick  3  5
!C3 -H31 1.0928
quick  6  7
!C5 -H51 1.0964
quick 11 12
!C5 -H52 1.0964
```

```
quick 11 13


!A1   A(2,1,3)    O1 - C1 - C2  125.0815
quick 2 1 3
!A2   A(2,1,11)   O1 - C1 - C5  125.7255
quick 2 1 11
!A3   A(3,1,11)   C2 - C1 - C5 109.193
quick 3 1 11
!A4   A(1,3,4)    C1 - C2 - H21 107.5191
quick 1 3 4
!A5   A(1,3,5)    C1 - C2 - H22 110.1684
quick 1 3 5
!A6   A(1,3,6)    C1 - C2 - C3 105.4448
quick 1 3 6
!A7   A(4,3,5)    H21 - C2 - H22 107.2716
quick 4 3 5
!A8   A(4,3,6)    H21 - C2 - C3 111.1477
quick 4 3 6
!A9   A(5,3,6)    H22 - C2 - C3 115.0749
quick 5 3 6
!A10  A(3,6,7)    C2 - C3 - H31 115.075
quick 3 6 7
!A11  A(3,6,8)    C2 - C3 - H32 111.1476
quick 3 6 8
!A12  A(3,6,9)    C2 - C3 - C4 105.4446
quick 3 6 9
!A13  A(7,6,8)    H31 - C3 - H32 107.2717
quick 7 6 8
!A14  A(7,6,9)    H31 - C3 - C4 110.1685
quick 7 6 9
!A15  A(8,6,9)    H32 - C3 - C4 107.5191
quick 8 6 9
!A16  A(6,9,10)   C3 - C4 - O4 125.0816
quick 6 9 10
!A17  A(6,9,11)   C3 - C4 - C5 109.1928
quick 6 9 11
!A18  A(10,9,11)  O4 - C4 - C5 125.7256
quick 10 9 11
!A19  A(1,11,9)   C1 - C5 - C4 105.3362
quick 1 11 9
!A20  A(1,11,12)  C1 - C5 - H51 111.1703
quick 1 11 12
!A21  A(1,11,13)  C1 - C5 - H52 111.2602
quick 1 11 13
!A22  A(9,11,12)  C4 - C5 - H51 111.2609
quick 9 11 12
!A23  A(9,11,13)  C4 - C5 - H52 111.1697
quick 9 11 13
!A24  A(12,11,13) H51 - C5 - H52 106.7286
quick 12 11 13


!must be easily greppable (grep EXTRA) in output file
!echu
!D3   D(3,6,9,11) C2-C3-C4-C5 19.5145 gv
quick  3 6 9 11
```

```
echo extra: C2-C3-C4-C5 19.5145 gv
echo extra:  3 6 9 11 ?phi
!D4    D(6,9,11,1) C3-C4-C5-C1 -7.5311 gv
quick 6 9 11 1
echo extra: C3-C4-C5-C1 -7.5311 gv
echo extra: 6 9 11 1 ?phi
!D5    D(3,1,11,9) C4-C5-C1-C2 -7.5267 gv
quick 3 1 11 9
echo extra: C4-C5-C1-C2 -7.5267 gv
echo extra: 3 1 11 9 ?phi
!D1    D(11,1,3,6) C5-C1-C2-C3 19.5117 gv
quick 11 1 3 6
echo extra: C5-C1-C2-C3 19.5117 gv
echo extra: 11 1 3 6 ?phi
!D2    D(1,3,6,9)  C1-C2-C3-C4 -23.4181 gv
quick 1 3 6 9
echo extra: C1-C2-C3-C4 -23.4181 gv
echo extra: 1 3 6 9  ?phi


!D2    D(1,3,11,2) C1-C2-C5-O1 -0.036 gv
quick 1 3 11 2
echo extra: C1-C2-C5-O1 -0.036 gv
echo extra: 1 3 11 2  ?phi


vibran
diag
end


!MOLVIB analysis by CHARMm
!Although the minimum is slightly asymmetric, the parameters are
!perfectly symmetric, so we use a symmetric molvib analysis.
! NOTE: The MOLVIB module is very sensitive to the format
MOLVIB NDI1 33 NDI2 33 NDI3 44 SECO
GFX
PRNT    1
DIM    33    33    44
IC
 1  1  2  0  0 !1 R1    R(1,2)       C1 = O1
 1  1  3  0  0 !2 R2    R(1,3)       C1 - C2
 1  1 11  0  0 !3 R3    R(1,11)      C1 - C5
 1  3  4  0  0 !4 R4    R(3,4)       C2 - H21
 1  3  5  0  0 !5 R5    R(3,5)       C2 - H22
 1  3  6  0  0 !6 R6    R(3,6)       C2 - C3
 1  6  7  0  0 !7 R7    R(6,7)       C3 - H31
 1  6  8  0  0 !8 R8 R(6,8)   C3 - H32
 1  6  9  0  0 !9 R9    R(6,9)       C3 - C4
 1  9 10  0  0 !10 R10 R(9,10)       C4 = O4
 1  9 11  0  0 !11 R11 R(9,11)       C4 - C5
 1 11 12  0  0 !12 R12 R(11,12)      C5 - H51
 1 11 13  0  0 !13 R13 R(11,13)      C5 - H52
 2  2  1  3  0 !14 A1   A(2,1,3)     O1 - C1 - C2   1b2
 2  2  1 11  0 !15 A2   A(2,1,11)    O1 - C1 - C5   1b1
 2  3  1 11  0 !16 A3   A(3,1,11)    C2 - C1 - C5   r1a4
 2  1  3  4  0 !17 A4   A(1,3,4)     C1 - C2 - H21  2b3
 2  1  3  5  0 !18 A5   A(1,3,5)     C1 - C2 - H22  2b4
```

```
 2  1  3  6  0 !19 A6   A(1,3,6)    C1 - C2 - C3   r1a5 gamma
 2  4  3  5  0 !20 A7   A(4,3,5)    H21 - C2 - H22 2 alpha
 2  4  3  6  0 !21 A8   A(4,3,6)    H21 - C2 - C3  2b1
 2  5  3  6  0 !22 A9   A(5,3,6)    H22 - C2 - C3  2b2
 2  3  6  7  0 !23 A10  A(3,6,7)    C2 - C3 - H31  3b3
 2  3  6  8  0 !24 A11  A(3,6,8)    C2 - C3 - H32  3b4
 2  3  6  9  0 !25 A12  A(3,6,9)    C2 - C3 - C4   r1a1 gamma
 2  7  6  8  0 !26 A13  A(7,6,8)    H31 - C3 - H32 3 alpha
 2  7  6  9  0 !27 A14  A(7,6,9)    H31 - C3 - C4  3b1
 2  8  6  9  0 !28 A15  A(8,6,9)    H32 - C3 - C4  3b2
 2  6  9 10  0 !29 A16  A(6,9,10)   C3 - C4 - O4   4b1
 2  6  9 11  0 !30 A17  A(6,9,11)   C3 - C4 - C5   r1a2
 2 10  9 11  0 !31 A18  A(10,9,11)  O4 - C4 - C5   4b2
 2  1 11  9  0 !32 A19  A(1,11,9)   C1 - C5 - C4   r1a3 gamma
 2  1 11 12  0 !33 A20  A(1,11,12)  C1 - C5 - H51  5b1
 2  1 11 13  0 !34 A21  A(1,11,13)  C1 - C5 - H52  5b2
 2  9 11 12  0 !35 A22  A(9,11,12)  C4 - C5 - H51  5b3
 2  9 11 13  0 !36 A23  A(9,11,13)  C4 - C5 - H52  5b4
 2 12 11 13  0 !37 A24  A(12,11,13) H51 - C5 - H52 5 alpha
 3  1  3 11  2 !38 I1   I(1,3,11,2)  C1-C2-C5-O1   C1
 3  9 11  6 10 !39 I1   I(9,11,6,10) C5-C1-C2-O4   C4
 4 11  1  3  6 !40 D1   D(11,1,3,6)  C5-C1-C2-C3   r1t4
 4  1  3  6  9 !41 D2   D(1,3,6,9)   C1-C2-C3-C4   r1t5
 4  3  6  9 11 !42 D3   D(3,6,9,11)  C2-C3-C4-C5   r1t1
 4  6  9 11  1 !43 D4   D(6,9,11,1)  C3-C4-C5-C1   r1t2
 4  9 11  1  3 !44 D5   D(9,11,1,3)  C4-C5-C1-C2   r1t3
UMAT      0    1    0              ! row normalization
  1  1  1.         2  2  1.         3  3  1.
  4  4  1.         4  5  1.
  5  4  1.         5  4  -1.
  6  6  1.
  7  7  1.         7  8  1.
  8  7  1.         8  8  -1.
  9  9  1.        10 10  1.
 11 11  1.
 12 12  1.        12 13  1.
 13 12  1.        13 13  -1.
 14 19  1.        14 20  5.
 15 17  1.        15 18  -1.       15 21  1.        15 22  -1.
 16 17  -1.       16 18  -1.       16 21  1.        16 22  1.
 17 17  -1.       17 18  1.        17 21  1.        17 22  -1.
 18 25  1.        18 26  5.
 19 23  1.        19 24  -1.       19 27  1.        19 28  -1.
 20 23  -1.       20 24  -1.       20 27  1.        20 28  1.
 21 23  -1.       21 24  1.        21 27  1.        21 28  -1.
 22 32  1.        22 37  5.
 23 33  1.        23 34  -1.       23 35  1.        23 36  -1.
 24 33  1.        24 34  1.        24 35  -1.       24 36  -1.
 25 33  1.        25 34  -1.       25 35  -1.       25 36  1.
 26 38  1.
 27 15  1.        27 14  -1.
 28 39  1.
 29 29  1.        29 31  -1.
 30 25  1.        30 30  -0.81      30 32  0.31      30 16  0.31
 30 19 -0.81
```

```
 31 30 -1.12        31 32  1.81        31 16 -1.81        31 19  1.12
 32 42  0.31        32 43 -0.81        32 44  1.         32 40 -0.81
 32 41  0.31
 33 42 -1.81        33 43  1.12        33 40 -1.12        33 41  1.81
-1
PED       0   15
  1  sC1=O1
  2  sC1-C2
  3  sC1-C5
  4  sC2-H2sy
  5  sC2-H2as
  6  sC2-C3
  7  sC3-H3sy
  8  sC3-H3as
  9  sC3-C4
 10  sC4=O4
 11  sC4-C5
 12  sC5-H5sy
 13  sC5-H5as
 14  sc-C2H2
 15  r-C2H2
 16  w-C2H2
 17  t-C2H2
 18  sc-C3H2
 19  r-C3H2
 20  w-C3H2
 21  t-C3H2
 22  sc-C5H2
 23  r-C5H2
 24  w-C5H2
 25  t-C5H2
 26  w-C1=O1
 27  d-C1=O1
 28  w-C4=O4
 29  d-C4=O4
 30  d-5RNGsy
 31  d-5RNGas
 32  t-5RNGsy
 33  t-5RNGas
-1
END
```

Table 35: Parameterization script (MM): Bond, angle and torsion parameter optimization for CPD

```
* From tutorial on A. MacKerell's website
* Modified by Cassian
* Potential energy scan for dihedrals
*

ioformat extended
```

```
set residue cpd
set basnam scan_cpd
set zeed 1 C1  1 C2  1 C3 ! seed
set frist none ! first capping residue
set lsat none  ! last capping residue
set pesbas ..
set TOPPAR  ..
set peses 4
set pes1 gausst1
set pes2 gausst2
set pes3 gausst3
set pes4 gausst4
set cons 7
set con1 1 C1 1 C2 1 C3 1 C4
set con2 1 C2 1 C3 1 C4 1 C5
set con3 1 C3 1 C4 1 C5 1 C1
set con4 1 C2 1 C1 1 C5 1 C4
set con5 1 C5 1 C1 1 C2 1 C3
set con6 1 C1 1 C2 1 C5 1 O1
set con7 1 C4 1 C5 1 C3 1 O4


set meas 11
set mea1 @con1
set mea2 @con2
set mea3 @con3
set mea4 @con4
set mea5 @con5
set mea6 1 C3 1 C2 1 C1 1 O1
set mea7 1 C4 1 C5 1 C1 1 O1
set mea8 1 C2 1 C3 1 C4 1 O4
set mea9 1 C1 1 C5 1 C4 1 O4
set mea10 1 O1 1 C1 1 C2 1 H21
set mea11 1 C5 1 C1 1 C2 1 H21
set offset 31
open unit 31 append form name @basnam.c1ccc4
open unit 32 append form name @basnam.ccc4c
open unit 33 append form name @basnam.cc4cc1
open unit 34 append form name @basnam.c4cc1c
open unit 35 append form name @basnam.cc1cc
open unit 36 append form name @basnam.ccc1o
open unit 37 append form name @basnam.c4cc1o
open unit 38 append form name @basnam.ccc4o
open unit 39 append form name @basnam.c1cc4o
open unit 40 append form name @basnam.oc1ch
open unit 41 append form name @basnam.cc1ch


set eun 29
open unit 29 append form name @basnam.mme


open unit 10 read card name @TOPPAR/top_all36_cgenff_cpd.rtf
read rtf card unit 10
close unit 10
open unit 10 read card name @TOPPAR/par_all36_cgenff_cpd.prm
read para card flex unit 10
close unit 10
```

```
if @?str eq 1 stream @str

open unit 10 write form name ini_geom.pdb
open unit 11 write form name fin_geom.pdb

read sequence card
* @residue
*
1
@residue

generate @residue first @frist last @lsat setup warn

!  patch @residue @residue 1 setup warn
!  autogenerate angles dihedral

set scan 1
label loop_scan
set pesdir @pesbas/@pes@@scan
open unit 25 read form name @pesdir/s_file.txt

label loop_pes
get pdbnam unit 25
if "@PDBNAM" eq "" goto nextscan
open unit 27 read form name @pdbnam
read coor pdb unit 27
close unit 27

write coor pdb unit 10

set cond 1
label loop_cond
cons dihe @con@@cond FORCE 9999. MAIN PERIOD 0
incr cond
if @cond le @cons goto loop_cond

MINI CONJ nstep 50 nprint 20 inbfrq 1000 cutnb 999.
MINI NRAP nstep 100 tolgrd 0.00001

write coor pdb unit 11

echu @eun
cons cldh
energy
echo ?ener

if @meas eq 0 goto loop_pes

set mead 1
set mun @offset
label loop_mead
echu @mun
quick @mea@@mead
echo ?phi
```

```
incr mun
incr mead
if @mead le @meas goto loop_mead

goto loop_pes

label nextscan
close unit 25
incr scan
if @scan le @peses goto loop_scan
```

Table 36: Parameterization script (MM): Potential energy scans of dihedrals and impropers for CPD

A3 Scripts used for setup and run of molecular dynamic simulations

```
* Structure Optimization
*

ioformat extended

set pdbdir .

!### Set variables ###
set residue cpd-hocl
set frist none ! first capping residue
set lsat none  ! last capping residue
set bombover 0 !needs to be set to -1 for 3-membered rings

open unit 10 read form name top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

! May need dummy atom to generate HOCl
stream toppar_dum_nobel_gases.str
if @?str eq 1 stream @str

read sequence card
* CPD
*
1
CPD

bomlev @bombover
generate CPD first @frist last @lsat setup warn

read sequence card
* HOCL
*
1
HOCL

bomlev @bombover
generate HOCL first @frist last @lsat setup warn

open unit 4 card read name cpd-hocl.crd
read coor card unit 4
```

```
   coor print

   open unit 30 write form name @residue_init.pdb
   write coor pdb unit 30
   close unit 30

!minimize structure and calculate dipole moment
   MINI CONJ nstep 200 nprint 20 inbfrq 1000 cutnb 999.
   MINI NRAP nstep 50 tolgrd 0.00001

!save minimized geometry
   coor print
   open unit 30 write form name @residue_min.psf
   write psf card unit 30
   close unit 30

   open unit 30 write form name @residue_min.crd
   write coor card unit 30

   open unit 30 write form name @residue_min.pdb
   write coor pdb unit 30
   close unit 30

stop
```

Table 37: Initial minimization script for CPD and HOCl

```
*NAME: Solvate.inp
*PURPOSE: build energetically-adjusted ion-water environment around
protein
*AUTHOR: Yeo H.C.Modified by Cassian
*

ioformat extended

set pdbdir .

!### Set variables ###
set residue cpd-hocl
set frist none ! first capping residue
set lsat none  ! last capping residue
set bombover 0 !needs to be set to -1 for 3-membered rings

open unit 10 read form name top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name top_all36_cgenff_remas.rtf
```

```
read rtf card unit 10 append
close unit 10
open unit 10 read card name par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

! May need dummy atom to generate HOCl
stream toppar_dum_nobel_gases.str
if @?str eq 1 stream @str

open read formatted unit 3 name cpd-hocl_min.psf
read psf card unit 3
open read formatted unit 4 name cpd-hocl_min.crd
read coor card unit 4

ener

! find out where protein is and put it with center at origin
coor stat sele segi cpd .or. segid hocl end

! orient molecule with moments of inertia along x,y,z-axes
! everything will be translated/rotated
coor orie sele segid cpd .or. segid hocl end
coor stat sele segi cpd .or. segid hocl end

! Add water; here no ions are added, this should perhaps also be
done
! Box size could be autmatically determined here by adding say 8A on
each side to the
! dimensions of the protein, we also make the sizes integer numbers
(not necessary)
calc XSIZ = INT( ?XMAX - ?XMIN + 16)
calc YSIZ = INT( ?YMAX - ?YMIN + 16)
calc ZSIZ = INT( ?ZMAX - ?ZMIN + 16)

! Set size based on previous calculation
set XSIZ 24
set YSIZ 24
set ZSIZ 24

define SOLUTE sele segid cpd .or. segid hocl end

!Read the default water box and modify it
stream mod-solvent-box.str

! Variable A is used in the stream file
set A @XSIZ

! slice box into truncated-octahedral
stream setup-truncated-octahedron.str

open write card unit 1 name temp-oct.pdb
write coor pdb unit 1
* octahedron
*
```

```
coor conv align symm @ax @ax @ax @b @b @b  !do not put this in
dynamics !!


! save the PSF and coordinates for later use
open write formatted unit 27 name cpd-hocl-solv.psf
write psf card unit 27
*  with crystal waters solvated with TIP3 waters
*

open write formatted unit 27 name cpd-hocl-solv.crd
write coor card unit 27
* Energy Minimized Coordinate Set for @FILE
*

! For viewing in VMD
open unit 1 write card name cpd-hocl-solv.pdb
write coor pdb unit 1
* @FILE for viewing in VMD
* in rectangular box: @XSIZ x @YSIZ x @ZSIZ
*

stop
```

Table 38: Script for solvating CPD and HOCl in a truncated octahedron box

```
* CHARMM script for initial minimization
*

! Read in pre-defined variables
stream datadir.def

! Read initial data files
open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

if @?str eq 1 stream @str

open read card unit 1 name @0cpd-hocl-solv.psf
read psf card unit 1
```

```
open read card unit 1 name @0cpd-hocl-solv.crd
read coor card unit 1

calc xsiz 24. * sqrt(0.75)
calc ysiz 24. * sqrt(0.75)
calc zsiz 24. * sqrt(0.75)
set angle 109.4712206344907
crystal define octahedral @XSIZ @YSIZ @ZSIZ @angle @angle @angle
crystal build cutoff 14.0

image byresidue sele resn tip3 end  xcen 0.0 ycen 0.0 zcen 0.0
image bysegment sele segi cpd .or. segid hocl end  xcen 0.0 ycen 0.0
zcen 0.0

! Specify nonbond parameters and evaluate energy once
nbonds cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
  cutim 14.0 wmin 1.5 -
  ewald pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6

ener

mini sd nstep 200 nprint 10
mini abnr nstep 200 nprint 10

open write card unit 1 name @1/init_mini.crd
write coor card unit 1
* init mini
*

stop
```

Table 39: Initial minimization script for solvated CPD and HOCl

```
* CHARMM script for initialization of NVE dynamics
*

! Read in pre-defined variables
stream datadir.def

! Read initial data files
open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
```

```
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

if @?str eq 1 stream @str

open read card unit 1 name @0cpd-hocl-solv.psf
read psf card unit 1
open read card unit 1 name @0cpd-hocl-solv.crd
read coor card unit 1

calc xsiz 24. * sqrt(0.75)
calc ysiz 24. * sqrt(0.75)
calc zsiz 24. * sqrt(0.75)
set angle 109.4712206344907
crystal define octahedral @XSIZ @YSIZ @ZSIZ @angle @angle @angle
crystal build cutoff 14.0

image byresidue sele resn tip3 end  xcen 0.0 ycen 0.0 zcen 0.0
image bysegment sele segi cpd .or. segid hocl end  xcen 0.0 ycen 0.0
zcen 0.0

! Specify nonbond parameters and evaluate energy once
nbonds cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
  cutim 14.0 wmin 1.5 -
  ewald pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6

ener

open read card unit 1 name @1/init_mini.crd
read coor card unit 1

! Open files for writing dynamics data.
open write card unit 31 name dyn.rst
open write file unit 41 name dyn.trj
open write file unit 51 name dyn.vel

! Dynamics command with options described
DYNAMICS -
 - ! algorithm
  LEAP -
  VERLET - !
 - ! start/restart/initialization info
  START -  ! START for initial run; RESTART to pick up from previous
run
  IASORS 1 -  ! specifies scaling (0) or assigning (other value) of
velocities
  IASVEL 1 -  ! species method of assigning velocities during
heating
  ISCVEL 0 -  ! specifies method of scaling velocities (0=same
factor)
 - ! time step info
  NSTEP 50000     - ! number of steps
  TIMESTEP 0.001  - ! timestep in picoseconds
```

```
  IPRFRQ 1000      - ! frequency for calculating averages and
fluctuations
                   - ! (must be at least as large as NTRFRQ unless
NTRFRQ=0)
 - ! heating info
  IHTFRQ 2000  - ! heating frequency
  FIRSTT 98.0  - ! initial temperature
  FINALT 298.0 - ! final temperature
  TEMINC 100.0 - !temperature increment
 - ! equilibration info
  IEQFRQ 5000  - ! frequency for assigning or scaling velocities
(during equil.)
  ICHECW 1     - ! every IEQFREQ steps, always change velocities (0)
               - ! or change only if temperature is outside window
(other value)
  TWINDH  10.0 - ! upper temperature window limit
  TWINDL -10.0 - ! lower temperature window limit
  NTRFRQ 5000  - ! frequency for removing overall translation and
rotation
 -! io units
  IUNREA -1 - ! unit number for reading restart info (-1 means no
restart)
  IUNWRI 31 - ! unit number for writing restart info
  IUNCRD 41 - ! unit number for writing coordinates (coordinate
trajectory)
  IUNVEL 51 - ! unit number for writing velocities (velocity
trajectory)
  IUNIT  -1 - ! unit number to which to write miscellaneous
information
 - ! printing & writing directives
  NPRINT 100 - ! frequency for printing output info
  NSAVC  20 - ! frequency for saving coordinates
  NSAVV  20 - ! frequency for saving velociti4es
 - ! energy
  NBONDS cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
        cutim 14.0 wmin 1.5 -
  EWALD pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6 -
  INBFRQ 25    ! frequency for undating nonbond list

open write card unit 1 name init_nve.crd
write coor card unit 1
* Coordinates for simulation after
* initial NVE heating of 50,000 steps (0.002 ps time step)
* from 98 K to 298 K.  This is preparation
* for a subsequent NPT simulation.

  stop
```

Table 40: Initial NVE dynamic simulation script (slowly heating the system to 298 K)

```
* CHARMM script for initialization of NPT dynamics
*

! Read in pre-defined variables
stream datadir.def

! Read initial data files
open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

if @?str eq 1 stream @str

open read card unit 1 name @0cpd-hocl-solv.psf
read psf card unit 1
open read card unit 1 name @0cpd-hocl-solv.crd
read coor card unit 1

calc xsiz 24. * sqrt(0.75)
calc ysiz 24. * sqrt(0.75)
calc zsiz 24. * sqrt(0.75)
set angle 109.4712206344907
crystal define octahedral @XSIZ @YSIZ @ZSIZ @angle @angle @angle
crystal build cutoff 14.0

image byresidue sele resn tip3 end  xcen 0.0 ycen 0.0 zcen 0.0
image bysegment sele segi cpd .or. segid hocl end  xcen 0.0 ycen 0.0
zcen 0.0

! Specify nonbond parameters and evaluate energy once
nbonds cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
  cutim 14.0 wmin 1.5 -
  ewald pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6

ener

open read card unit 1 name @1/init_mini.crd
read coor card unit 1

open read card unit 1 name init_nve.crd
read coor card unit 1
ener

! Open files for writing dynamics data.
open write card unit 31 name dyn.rst
```

```
open write file unit 41 name dyn.trj
open write file unit 51 name dyn.vel

! Dynamics command with options described
DYNAMICS -
 - ! algorithm
  CPT - !
 - ! start/restart/initialization info
  START -  ! START for initial run; RESTART to pick up from previous
run
  IASORS 1 -  ! specifies scaling (0) or assigning (other value) of
velocities
  IASVEL 1 -  ! species method of assigning velocities during
heating
  ISCVEL 0 -  ! specifies method of scaling velocities (0=same
factor)
 - ! time step info
  NSTEP 500000     - ! number of steps
  TIMESTEP 0.001  - ! timestep in picoseconds
  IPRFRQ 1000     - ! frequency for calculating averages and
fluctuations
                  - ! (must be at least as large as NTRFRQ unless
NTRFRQ=0)
 - ! heating info
  IHTFRQ 2000  - ! heating frequency
  FIRSTT 298.0  - ! initial temperature
  FINALT 298.0 - ! final temperature
  TEMINC 0.0 - !temperature increment
 - ! cpt parameters
  PCONst pmass 400.0 pgamma 20.0 tbath 298.0 -
 - ! equilibration info
  IEQFRQ 5000  - ! frequency for assigning or scaling velocities
(during equil.)
  ICHECW 1      - ! every IEQFREQ steps, always change velocities (0)
                - ! or change only if temperature is outside window
(other value)
  TWINDH  10.0 - ! upper temperature window limit
  TWINDL -10.0 - ! lower temperature window limit
  NTRFRQ 5000  - ! frequency for removing overall translation and
rotation
 -! io units
  IUNREA -1 - ! unit number for reading restart info (-1 means no
restart)
  IUNWRI 31 - ! unit number for writing restart info
  IUNCRD 41 - ! unit number for writing coordinates (coordinate
trajectory)
  IUNVEL 51 - ! unit number for writing velocities (velocity
trajectory)
  IUNIT  -1 - ! unit number to which to write miscellaneous
information
 - ! printing & writing directives
  NPRINT 100 - ! frequency for printing output info
  NSAVC  20 - ! frequency for saving coordinates
  NSAVV  20 - ! frequency for saving velociti4es
 - ! energy
```

```
  NBONDS cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
          cutim 14.0 wmin 1.5 -
  EWALD pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6 -
  INBFRQ 25     ! frequency for undating nonbond list

  stop
```

Table 41: Initial NPT dynamic simulation script

```
* CHARMM script for NPT dynamics
*

! Read in pre-defined variables
stream datadir.def

! Read initial data files
open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

if @?str eq 1 stream @str

open read card unit 1 name @0cpd-hocl-solv.psf
read psf card unit 1
open read card unit 1 name @0cpd-hocl-solv.crd
read coor card unit 1

calc xsiz 24. * sqrt(0.75)
calc ysiz 24. * sqrt(0.75)
calc zsiz 24. * sqrt(0.75)
set angle 109.4712206344907
crystal define octahedral @XSIZ @YSIZ @ZSIZ @angle @angle @angle
crystal build cutoff 14.0

image byresidue sele resn tip3 end  xcen 0.0 ycen 0.0 zcen 0.0
image bysegment sele segi cpd .or. segid hocl end  xcen 0.0 ycen 0.0
zcen 0.0

! Specify nonbond parameters and evaluate energy once
nbonds cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
```

```
7.0 -
  cutim 14.0 wmin 1.5 -
  ewald pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6

ener

open read card unit 1 name @1/init_mini.crd
read coor card unit 1

! Open files for writing dynamics data.
open read card unit 30 name dyn.rea
open write card unit 31 name dyn.rst
open write file unit 41 name dyn.trj
open write file unit 51 name dyn.vel

! Dynamics command with options described
DYNAMICS -
 - ! algorithm
  CPT - !
 - ! start/restart/initialization info
  REST  -  ! START for initial run; RESTART to pick up from previous
run
  IASORS 0 -  ! specifies scaling (0) or assigning (other value) of
velocities
  IASVEL 1 -  ! species method of assigning velocities during
heating
  ISCVEL 0 -  ! specifies method of scaling velocities (0=same
factor)
 - ! time step info
  NSTEP 500000    - ! number of steps
  TIMESTEP 0.001  - ! timestep in picoseconds
  IPRFRQ 1000     - ! frequency for calculating averages and
fluctuations
                  - ! (must be at least as large as NTRFRQ unless
NTRFRQ=0)
 - ! heating info
  IHTFRQ 2000  - ! heating frequency
  FIRSTT 298.0  - ! initial temperature
  FINALT 298.0 - ! final temperature
  TEMINC 0.0 - !temperature increment
 - ! cpt parameters
  PCONst pmass 400.0 pgamma 20.0 tbath 298.0 -
 - ! equilibration info
  IEQFRQ 5000  - ! frequency for assigning or scaling velocities
(during equil.)
  ICHECW 1     - ! every IEQFREQ steps, always change velocities (0)
               - ! or change only if temperature is outside window
(other value)
  TWINDH  10.0 - ! upper temperature window limit
  TWINDL -10.0 - ! lower temperature window limit
  NTRFRQ 5000  - ! frequency for removing overall translation and
rotation
 -! io units
  IUNREA 30 - ! unit number for readiing restart info (-1 means no
restart)
```

```
   IUNWRI 31 - ! unit number for writing restart info
   IUNCRD 41 - ! unit number for writing coordinates (coordinate
trajectory)
   IUNVEL 51 - ! unit number for writing velocities (velocity
trajectory)
   IUNIT  -1 - ! unit number to which to write miscellaneous
information
 - ! printing & writing directives
   NPRINT 100 - ! frequency for printing output info
   NSAVC  20 - ! frequency for saving coordinates
   NSAVV  20 - ! frequency for saving velociti4es
 - ! energy
   NBONDS cdie eps 1.0 atom vdw vswitch cutnb 12.0 ctofnb 10.0 ctonnb
7.0 -
         cutim 14.0 wmin 1.5 -
   EWALD pmewald kappa 0.34 fftx 24 ffty 24 fftz 24 order 6 -
   INBFRQ 25     ! frequency for undating nonbond list

   stop
```

Table 42: Script for NPT dynamics simulation used for analysis

```
#!/bin/bash

#! example of job file to submit parallel MPI applications
#! lines starting with #PBS are options for the qsub command

#! Name of job
#PBS -N CPD-HOCL_MPI

#! Number of nodes (in this case I require 1 nodes with 8 CPU each)
#! The total number of nodes passed to mpirun will be nodes*ppn
#PBS -l nodes=1:ppn=8

#! Specific the shell types
#PBS -S /bin/bash

#! Name of output files for std output and error;
#! if non specified defaults are <job-name>.o<job number> and <job-
name>.e<jobnumber>
#PBS -e submit.err
#PBS -o submit.out

 #!change the working directory (default is home directory)
 cd $PBS_O_WORKDIR
 echo Running on host `hostname`
 echo Time is `date`
 echo Directory is $PBS_O_WORKDIR
 echo This jobs runs on the following processors:
 echo `cat $PBS_NODEFILE`
```

```
    #! Run the parallel MPI executable with 8 processor (nodes*ppn)
/opt/mpich2/osu_ch3_mrail_gen2-gnu41/bin/mpiexec -np 8
/opt/c34b1/exec/gnu/charmmgff str:toppar_cpd-hocl.str -i
init_mini.inp -o init_mini.out

/opt/mpich2/osu_ch3_mrail_gen2-gnu41/bin/mpiexec -np 8
/opt/c34b1/exec/gnu/charmmgff str:toppar_cpd-hocl.str -i
init_nve.inp -o init_nve.out
mv dyn.rst dyn.rst.nve
mv dyn.trj dyn.trj.nve
mv dyn.vel dyn.vel.nve

/opt/mpich2/osu_ch3_mrail_gen2-gnu41/bin/mpiexec -np 8
/opt/c34b1/exec/gnu/charmmgff str:toppar_cpd-hocl.str -i
init_npt.inp -o dyn.out.1
cp dyn.rst dyn.rea
mv dyn.rst dyn.rst.1
mv dyn.trj dyn.trj.1
mv dyn.vel dyn.vel.1

for i in 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
   do
        /opt/mpich2/osu_ch3_mrail_gen2-gnu41/bin/mpiexec -np 8
/opt/c34b1/exec/gnu/charmmgff str:toppar_cpd-hocl.str -i
equil_npt.inp -o dyn.out.$i
  cp -u dyn.rst dyn.rst.$i
  cp -u dyn.rst dyn.rea
  mv dyn.trj dyn.trj.$i
  mv dyn.vel dyn.vel.$i
     done
```

Table 43: Shell script to run various molecular dynamics scripts via queue on computer cluster and appropriately archive the output files

```
* read in traj and write out a grp of frames - cassian
*

! Read in pre-defined variables
stream datadir.def

! Read initial data files
open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
```

```
read para card flex unit 10 append
close unit 10

! Read in CPD + HOCl parameter file in command line
if @?str eq 1 stream @str

open read card unit 1 name @0cpd-hocl-solv.psf
read psf card unit 1
open read card unit 1 name @0cpd-hocl-solv.crd
read coor card unit 1

set a 9999
set b 9999
set c 9999
set a2 9999
set b2 9999
set c2 9999
set d 0
set m 0

! Selection criteria
set cutoff1 2.4
set cutoff2 3.2
set cutoff3 3.2
set trajfirst 1
set trajlast 20
calc diff = @trajlast - @trajfirst + 1
calc termination = 500000 * @diff / 20

set i @trajfirst
set j @trajlast
label trajread
open read unit @i file name @1dyn.trj.@i
incr i
if @i le @j goto trajread

open unit 99 write form name hbond3.tmp
open unit 98 write form name hbond3.pdb

!save the hbond distance
open unit 97 write form name hbond3.dat

echu 97
echo FRAME HOCL...O   OH2...H (RESID)  (O type) O...H20
TRAJECTORY FIRSTU @trajfirst NUNIT @diff
set 1 0
set q 1
label loop
      traj read
      incr 1 by 1


!*****************************************************************
! 24 = size of box of water
calc xsiz 24. * sqrt(0.75)
calc ysiz 24. * sqrt(0.75)
```

```
calc zsiz 24. * sqrt(0.75)
set angle 109.4712206344907
crystal define octahedral @XSIZ @YSIZ @ZSIZ @angle @angle @angle
crystal build cutoff 14.0

! Recenter CPD in box of water
coor stat select type C5 end
set xmin ?XMIN
set ymin ?YMIN
set zmin ?ZMIN

image byresidue sele resn tip3 end  xcen @xmin ycen @ymin zcen @zmin
image bysegment sele segi cpd .or. segid hocl end  xcen @xmin ycen
@ymin zcen @zmin

UPDATE IMGFRQ 1

!****************************************************************

      COOR MINDistance SELECT type HG1 END select type O4 .OR. type
O1 END
      set keto1 ?MINDA2
      COOR  DISTANCE CUT @cutoff1 SELECT type HG1 END select (type
O4 .OR. type O1) END
      IF ?MIND .lt. @cutoff1 THEN
       set atemp ?MIND
      define wat select (type C5 .around. 4.0) .and. type OH2 end
      set nwat ?NSEL
      IF @nwat .gt. 0 THEN
      set var1 1
      label loop1
            COOR MINDistance select (type H51 .OR. type H52) END
SELECT wat .subset. @var1 END
            IF ?MIND .lt. @cutoff2 THEN
                  set btemp ?MIND
                  IF @btemp .lt. @b THEN
                        set waterid ?SELRESI
                        IF keto1 .eq. 2 THEN
                              COOR MINDistance SELECT (resi @waterid
.and. (type H1 .or. type H2)) END select (type O4) END
                        else
                              COOR MINDistance SELECT (resi @waterid
.and. (type H1 .or. type H2)) END select (type O1) END
                        ENDIF
                        IF ?MIND .lt. @cutoff3 THEN
                        set ctemp ?MIND
                        IF @ctemp .lt. @c THEN
                        set a @atemp
                              set b @btemp
                              set c @ctemp
                  ENDIF
                  ENDIF
            ENDIF
         ENDIF
           incr var1 by 1
```

```
              if @var1 .le. @nwat goto loop1
        !open unit 98 write form name hbond@q.pdb
          !coor write pdb unit 98
          !incr q
          echo @1 @a @b @c
          calc interval = @1 - @d
         IF @interval .gt. 10000 THEN
         IF @b .lt. 9999 THEN
         IF @c .lt. 9999 THEN
          set d @1
          incr m
          !set count 1
        open write formatted unit 96 name struc@m.psf
           write psf card unit 96
          !open unit 96 write form name struc@m.pdb
          !coor write pdb unit 96
           open unit 96 write form name struc@m.crd
           coor write card unit 96
         set a2 @a
         set b2 @b
         set c2 @c
         set a 9999
         set b 9999
         set c 9999
     ENDIF
     ENDIF
       ELSE
        !incr count
        IF @a .le. @a2 THEN
        IF @b .le. @b2 THEN
        IF @c .lt. @c2 THEN
        open write formatted unit 96 name struc@m.psf
           write psf card unit 96
           !open unit 96 write form name struc@m.pdb
           !coor write pdb unit 96
           open unit 96 write form name struc@m.crd
           coor write card unit 96
         set a2 @a
         set b2 @b
         set c2 @c
         set a 9999
         set b 9999
         set c 9999
       ELSE
            set a 9999
            set b 9999
            set c 9999
     ENDIF
       ELSE
            set a 9999
            set b 9999
            set c 9999
     ENDIF
       ELSE
            set a 9999
```

```
            set b 9999
            set c 9999
     ENDIF
     ENDIF
     ENDIF
     ENDIF
       if @1 .lt. @termination goto loop

! protein - solvent:
! Note: assume that the trajectory is RECENTERED (protein near the
box center)
! do not use verbose option, since we are not interested in
individual waters here.
! Note that the trajectory is not closed between these commands, and
can be reused
coor contact CUTHB @cutoff1 verbose select type HG1  end select type
O1 .OR. type O4  end -
     firstu @trajfirst nunit @diff iunit 99

stop
```

Table 44: Script for reading trajectory and saving coordinates of reactant precursors based on pre-determined criteria

```
* CPD
* Run minimization for replica along reaction coordinate
*

bomlev -2


stream datadir.def

open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

! Topology and param file of CPD
if @?str eq 1 stream @str

set struc struc1
! Identify key atoms of catalytic water and CPD
set hrid 429
set hsid WT4
set alphah1 H52
set alphah2 H51
set keto O1

! Read PSF file of solvated CPD
read psf card name ../@struc.psf

! Read coordinates of product
open coor card unit 12 name p@struc.crd
read coor card unit 12

coor copy COMP

! Read coordinates of reactant
open coor card unit 12 name r@struc.crd
read coor card unit 12

!---------- Needed to define Q-Chem env. vars. ----------
 envi qchemcnt  "qchem-replica.inp"
 envi qcheminp  "q1.inp"
 envi qchemexe  "qchem"
```

```
 envi qchemout   "qchem.out"
 ENVI QCHEMPWD   "@9"
!---------------------------------------------------------

! Define QM atoms
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end

define buffer select .not. qm  end

REPLicate A NREPlica 16 SELECT qm END

! Need to redefine atom selection after relicate command
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end

define buffer select .not. qm .and. .not. segid A*  end

COOR orient rms mass select .not. (qm .or. segid A*) end

! Fix 'buffer' region:
cons fix select buffer end


!--------------------------------------------------------

! { setup initial guess coordinates for all intermediates }

set 1 1
set a 0.0
label loop
 coor duplicate sele qm end sele segid A@1 end
 coor duplicate sele qm end sele segid A@1 end comp
 coor average fact @a sele segid A@1 end
 incr a by 0.066667  ! 1/15 (since a starts at 0 and ends at 15)
 incr 1 by 1
if @1 .lt. 16.5 goto loop
!--------------------------------------------------------------------
------------

DELEte ATOM SELEct  qm  END
DEFIne replicas SELEct SEGId A* END

! Fix 'buffer' region:
define buffer select .not. segid A*  end
cons fix select none end
cons fix select buffer end

! { average the non active reactant and product atoms }
COOR average sele .not. replicas end
COOR copy comp

! { set up an appropriate interaction matrix }
BLOCK 2
 CALL 2 SELEct replicas END
 COEF 1 1 1.0
```

207

```
 COEF 2 2 0.0625   ! 16
 COEF 2 1 0.0625   ! 16
END

!#################################################

 scalar wmain set 1.0
 scalar wmain set 0.0 sele type h* end
 SCALAR wmain set 4 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type h*) .or. (resn cpd .and. type @alphah1) end
 SCALAR wmain set 2 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type o*) end
 SCALAR wmain set 1.5 sele (resn cpd .and. type @keto) .or. (resn
hocl .and. type o*) .or. (resn cpd .and. type c5) end
 SCALAR wmain set 1.5 sele (resn hocl .and. type h*) .or. (resn cpd
.and. type @alphah2) end

 coor copy comp
 coor copy comp weight

! write psf for next script/step
write psf card name replica.psf

! Run the Conventional Replica/Path Method
  rpath krms 600.0 rmax 0.25 kangle 100.0 kmax 600.0 cosmax 0.90
rota -
        tran weight

!#################################################

! { fix the endpoints }
cons fix select none end
cons fix sele segid a1 .or. segid a16 .or. buffer end

qchem RPATh parallel 1 remove sele replicas end

!----------------------------------------------------------
open write formatted unit 10 name replica_ini.crd
write coor card unit 10

MINI SD NSTEP 50 NPRINT 50
open write formatted unit 10 name replicaSD50.crd
write coor card unit 10


set 1 100
label loop22
MINI ABNR NSTEP 100 NPRINT 10
open write formatted unit 10 name replicaABNR@1.crd
write coor card unit 10

incr 1 by 100
if @1 .le. 500 goto loop22
!----------------------------------------------------------
```

```
! { plot energy as a function of the path }
open write card unit 20 name energy.dat
set 1 1
label eloop
BLOCK 2
 CALL 1 sele all end
 CALL 2 sele replicas .and. .not. segid A@1 end
 COEF 1 1 1.0
 COEF 2 1 0.0
 COEF 2 2 0.0
END

ENERGY

write title unit 20
* @1 ?ener
*
incr 1 by 1
if @1 .lt. 16.5 goto eloop

STOP
```

Table 45: Script for mapping out minimum energy path using Replica Path method with all MM waters fixed

```
* CPD
* Run minimization for replica along reaction coordinate
* AND locate Transition State using CI-NEB
*

bomlev -2


stream datadir.def

open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

! Topology and param file of CPD
if @?str eq 1 stream @str
```

```
set struc struc1
! Identify key atoms of catalytic water and CPD
set hrid 429
set hsid WT4
set alphah1 H52
set alphah2 H51
set keto O1

! Read PSF file of solvated CPD
read psf card name ../../@struc.psf

! Read coordinates of product
open coor card unit 12 name ../p@struc.crd
read coor card unit 12

coor copy COMP

! Read coordinates of reactant
open coor card unit 12 name ../r@struc.crd
read coor card unit 12

!---------- Needed to define Q-Chem env. vars. ----------
 envi qchemcnt  "qchem-replica.inp"
 envi qcheminp  "q1.inp"
 envi qchemexe  "qchem"
 envi qchemout  "qchem.out"
 ENVI QCHEMPWD  "@9"
!---------------------------------------------------------

! Define QM atoms
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end

define buffer select .not. qm  end

REPLicate A NREPlica 16 SELECT qm END

! Need to redefine atom selection after relicate command
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end

define buffer select .not. qm .and. .not. segid A*  end

COOR orient rms mass select .not. (qm .or. segid A*) end

! Fix 'buffer' region:
cons fix select buffer end

!---------------------------------------------------------

! { setup initial guess coordinates for all intermediates }

set 1 1
set a 0.0
label loop
```

```
 coor duplicate sele qm end sele segid A@1 end
 coor duplicate sele qm end sele segid A@1 end comp
 coor average fact @a sele segid A@1 end
 incr a by 0.066667  ! 1/15 (since a starts at 0 and ends at 15)
 incr 1 by 1
if @1 .lt. 16.5 goto loop
!------------------------------------------------------------------
------------

DELEte ATOM SELEct  qm  END

! Read PSF file of solvated CPD
read psf card name replica.psf
! Read coordinates of previous replicas
open coor card unit 12 name ci-nebsd50.crd
!read coor card comp unit 12
read coor card unit 12


!------------------------------------------------------------------
------------

DEFIne replicas SELEct SEGId A* END

! Fix 'buffer' region:
define buffer select .not. segid A*  end
cons fix select none end
cons fix select buffer end

! { average the non active reactant and product atoms }
COOR average sele .not. replicas end
COOR copy comp

! { set up an appropriate interaction matrix }
BLOCK 2
 CALL 2 SELEct replicas END
 COEF 1 1 1.0
 COEF 2 2 0.0625  ! 16
 COEF 2 1 0.0625  ! 16
END

!##################################################

 scalar wmain set 1.0
 scalar wmain set 0.0 sele type h* end
 SCALAR wmain set 4 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type h*) .or. (resn cpd .and. type @alphah1) end
 SCALAR wmain set 2 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type o*) end
 SCALAR wmain set 1.5 sele (resn cpd .and. type @keto) .or. (resn
hocl .and. type o*) .or. (resn cpd .and. type c5) end
 SCALAR wmain set 1.5 sele (resn hocl .and. type h*) .or. (resn cpd
.and. type @alphah2) end

 coor copy comp
 coor copy comp weight
```

```
! Run the Conventional Replica/Path Method
  rpath krms 600.0 rmax 0.25 kangle 100.0 kmax 600.0 cosmax 0.90
rota -
        tran weight nebf etan cimg ppmf

!###################################################

! { fix the endpoints }
cons fix select none end
cons fix sele segid a1 .or. segid a16 .or. buffer end

qchem RPATh parallel 1 remove sele replicas end

!----------------------------------------------------------
MINI SD NSTEP 50 NPRINT 50
open write formatted unit 10 name ci-nebSD100.crd
write coor card unit 10


set 1 200
label loop22
MINI SD NSTEP 100 NPRINT 10
open write formatted unit 10 name ci-nebsd@1.crd
write coor card unit 10

incr 1 by 100
if @1 .le. 500 goto loop22
!----------------------------------------------------------

! { plot energy as a function of the path }
open write card unit 20 name energy.dat
set 1 1
label eloop
BLOCK 2
 CALL 1 sele all end
 CALL 2 sele replicas .and. .not. segid A@1 end
 COEF 1 1 1.0
 COEF 2 1 0.0
 COEF 2 2 0.0
END

ENERGY

write title unit 20
* @1 ?ener
*
incr 1 by 1
if @1 .lt. 16.5 goto eloop

STOP
```

Table 46: Script for locating transition state of system using CI-NEB with all MM waters fixed

```
* CPD
* Run minimization for replica along reaction coordinate
* keeping a hydration shell flexible arounf the QM region
*

bomlev -2



stream datadir.def

open unit 10 read form name @0top_all22_prot.rtf
read rtf card unit 10
close unit 10
open unit 10 read form name @0par_all22_prot.prm
read param card flex unit 10
close unit 10
open unit 10 read card name @0top_all36_cgenff_remas.rtf
read rtf card unit 10 append
close unit 10
open unit 10 read card name @0par_all36_cgenff_remas.prm
read para card flex unit 10 append
close unit 10

! Topology and param file of CPD
if @?str eq 1 stream @str

set struc struc1
! Identify key atoms of catalytic water and CPD
set hrid 429
set hsid WT4
set alphah1 H52
set alphah2 H51
set keto O1

! Read PSF file of solvated CPD
read psf card name ../min-struc-shell/@struc.psf

! Read coordinates of product
open coor card unit 12 name ../min-struc-shell/t@struc.crd
read coor card unit 12


coor copy COMP

! Read coordinates of reactant
open coor card unit 12 name ../min-struc-shell/r@struc.crd
read coor card unit 12

!---------- Needed to define Q-Chem env. vars. ----------
 envi qchemcnt   "qchem-replica.inp"
 envi qcheminp   "q1.inp"
 envi qchemexe   "qchem"
 envi qchemout   "qchem.out"
 ENVI QCHEMPWD   "@9"
!--------------------------------------------------------
```

```
! Define QM atoms
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end


define buffer1 select .not. qm  end

! Define flexible hydration shell
define water sele ((segid WT4 .and. (resid 11 .or. resid 13 .or.
resid 62 .or. resid 147 .or. resid 216 .or. resid 241 .or. resid
422)) .or. (segid WT6 .and. resid 131)) -
show end

REPLicate A NREPlica 16 SELECT (qm .or. water) END

! Need to redefine atom selection after relicate command
define qm sele segid CPD .or. segid hocl .or. (resid @hrid .and.
segid @hsid) end

define buffer1 select .not. qm .and. .not. segid A*  end

define water sele ((segid WT4 .and. (resid 11 .or. resid 13 .or.
resid 62 .or. resid 147 .or. resid 216 .or. resid 241 .or. resid
422)) .or. (segid WT6 .and. resid 131)) -
show end

COOR orient rms mass select .not. (qm .or. segid A* .or. water) end

define buffer select buffer1 .and. .not. water end

! Fix 'buffer' region:
cons fix select buffer end

!--------------------------------------------------------

! { setup initial guess coordinates for all intermediates }

set 1 1
set a 0.0
label loop1
 coor duplicate sele (qm .or. water) end sele segid A@1 end
 coor duplicate sele (qm .or. water) end sele segid A@1 end comp
 coor average fact @a sele segid A@1 end
 incr a by 0.1428571  ! 1/7 (since a starts at 0 and ends at 15)
 incr 1 by 1
if @1 .lt. 8.5 goto loop1

open coor card unit 12 name ../min-struc-shell/p@struc.crd
read coor card unit 12


coor copy COMP

! Read coordinates of reactant
open coor card unit 12 name ../min-struc-shell/t@struc.crd
read coor card unit 12
```

```
set a 0.0
label loop2
 incr a by 0.125  ! 1/8 (since a starts at 9 and ends at 16)
 coor duplicate sele (qm .or. water) end sele segid A@1 end
 coor duplicate sele (qm .or. water) end sele segid A@1 end comp
 coor average fact @a sele segid A@1 end
 incr 1 by 1
if @1 .lt. 16.5 goto loop2
!--------------------------------------------------------------------
------------

DELEte ATOM SELEct  (qm .or. water) END

! Read PSF file of solvated CPD
read psf card name replica.psf
! Read coordinates of previous replicas
open coor card unit 12 name replicaabnr600.crd
!read coor card comp unit 12
read coor card unit 12

DEFIne replicas SELEct SEGId A* END
DEFIne qm SELEct (SEGId A* .and. (resid @hrid .or. resname cpd .or.
resname hocl)) show END

! Fix 'buffer' region:
define buffer select .not. segid A*  end
cons fix select none end
cons fix select buffer end

! { average the non active reactant and product atoms }
COOR average sele .not. replicas end
COOR copy comp

! { set up an appropriate interaction matrix }
BLOCK 2
 CALL 2 SELEct replicas END
 COEF 1 1 1.0
 COEF 2 2 0.0625  ! 16
 COEF 2 1 0.0625  ! 16
END

!#####################################################

 scalar wmain set 1.0
 scalar wmain set 0.0 sele type h* end
 SCALAR wmain set 4 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type h*) .or. (resn cpd .and. type @alphah1) show end
 SCALAR wmain set 2 sele (resid @hrid .and. (segid @hsid .or. segid
A*) .and. type o*) show end
 SCALAR wmain set 1.5 sele (resn cpd .and. type @keto) .or. (resn
hocl .and. type o*) .or. (resn cpd .and. type c5) show end
 SCALAR wmain set 1.5 sele (resn hocl .and. type h*) .or. (resn cpd
.and. type @alphah2) show end

 coor copy comp
```

```
 coor copy comp weight

! Run the Conventional Replica/Path Method
rpath krms 600.0 rmax 0.25 kangle 100.0 kmax 600.0 cosmax 0.90 rota
-
         tran weight nebf etan cimg ppmf

!####################################################

! { fix the endpoints }
cons fix select none end
cons fix sele segid a1 .or. segid a16 .or. buffer end

qchem RPATh QCLJ parallel 1 remove sele qm end

!----------------------------------------------------------
set 1 100
label loop22
MINI SD NSTEP 100 NPRINT 10
open write formatted unit 10 name ci-nebSD@1.crd
write coor card unit 10

incr 1 by 100
if @1 .le. 500 goto loop22
!----------------------------------------------------------

! { plot energy as a function of the path }
open write card unit 20 name energy/energy-ci_500.dat
set 1 1
label eloop
BLOCK 2
 CALL 1 sele all end
 CALL 2 sele replicas .and. .not. segid A@1 end
 COEF 1 1 1.0
 COEF 2 1 0.0
 COEF 2 2 0.0
END

ENERGY

write title unit 20
* @1 ?ener
*
incr 1 by 1
if @1 .lt. 16.5 goto eloop

stop
```

Table 47: Script for locating transition state using CI-NEB for system with flexible hydration shell

216

```
$comment
qchem control file needed for the
qchem-replica/ci-neb.inp
$end

$rem
exchange            B3LYP
basis               6-31+G**
qm_mm               true
jobtype             force
symmetry            off
sym_ignore          true
print_input         false
qmmm_print          true
SCF_ALGORITHM       DIIS_GDM
SCF_GUESS_ALWAYS    true
THRESH              14
SCF_CONVERGENCE     9
MAX_SCF_CYCLES      200
$end

$molecule
0 1
$end
```

Table 48: QM script read by CHARMM to create actual script to be run by QChem. Used in Replica Path method and CI-NEB method

```
$external_charges
   -12.66269371    -3.72031255    -0.82295700    -0.83400000
0.00024239      3.15057423
   -12.46711514    -4.08620449    -1.68400236     0.41700000
0.00007331      0.40001352
   -13.03480408    -2.85588642    -1.16791530     0.41700000
0.00007331      0.40001352
   -12.82328583    -0.88125814    -6.57305158    -0.83400000
0.00024239      3.15057423
   -12.43442879    -0.05557931    -6.15039695     0.41700000
0.00007331      0.40001352
   -12.32563613    -0.88686097    -7.40489782     0.41700000
0.00007331      0.40001352
    -8.21232033    -2.74258280     5.04212284    -0.83400000
0.00024239      3.15057423
    -8.40833473    -1.77126777     4.86145020     0.41700000
0.00007331      0.40001352
    -8.27885723    -2.76478148     6.04030132     0.41700000
0.00007331      0.40001352
   -11.67862480     8.36726152    -9.01817243    -0.83400000
0.00024239      3.15057423
   -11.98195331     9.23895475    -8.73214214     0.41700000
0.00007331      0.40001352
```

```
  -12.57279938       7.96145526      -9.25817983       0.41700000
0.00007331       0.40001352
    3.44952856      -0.70802351     -11.04037974      -0.83400000
0.00024239       3.15057423
    4.19502234      -0.41123434     -10.54358408       0.41700000
0.00007331       0.40001352
    2.76111905      -1.03030439     -10.34711382       0.41700000
0.00007331       0.40001352
    1.74631441      -7.57735586      -7.82149363      -0.83400000
0.00024239       3.15057423
    1.88608599      -7.97766495      -6.95973730       0.41700000
0.00007331       0.40001352
    0.95238292      -8.00864410      -8.19403648       0.41700000
0.00007331       0.40001352
   -8.90489964       8.30332777      -4.33863510      -0.83400000
0.00024239       3.15057423
   -9.47780470       9.03232691      -4.47938116       0.41700000
0.00007331       0.40001352
   -8.57478909       8.57199309      -3.45566349       0.41700000
0.00007331       0.40001352
   -7.58822541       9.37925360      -1.99477844      -0.83400000
0.00024239       3.15057423
   -7.33173518       9.14004347      -1.09726529       0.41700000
0.00007331       0.40001352
   -8.25068621      10.13018725      -1.80135827       0.41700000
0.00007331       0.40001352
   -8.04161835      -1.21567106      -4.70295620      -0.83400000
0.00024239       3.15057423
   -8.63120365      -1.92100966      -5.04181099       0.41700000
0.00007331       0.40001352
   -8.11833286      -1.30447316      -3.73433805       0.41700000
0.00007331       0.40001352
   -4.79387236       2.43332100      -7.43530464      -0.83400000
0.00024239       3.15057423
   -4.88113356       1.47559214      -7.60975504       0.41700000
0.00007331       0.40001352
   -5.71967506       2.66631365      -7.72223759       0.41700000
0.00007331       0.40001352
   -0.78158152       3.00790405      -9.29691982      -0.83400000
0.00024239       3.15057423
   -0.92151082       2.10098171      -8.98417664       0.41700000
0.00007331       0.40001352
   -0.55826366       3.04506445     -10.22206688       0.41700000
0.00007331       0.40001352
  -15.28397678       2.52064809      -9.35715883      -0.83400000
0.00024239       3.15057423
  -15.33644496       3.37625131      -8.87684839       0.41700000
0.00007331       0.40001352
  -15.90883388       2.12835606      -8.72084635       0.41700000
0.00007331       0.40001352
    2.06923771       7.52523041       0.79183286      -0.83400000
0.00024239       3.15057423
    2.85905099       7.62222481       0.25203398       0.41700000
0.00007331       0.40001352
    2.15169120       6.52610636       0.92274112       0.41700000
```

218

```
0.00007331      0.40001352
     -4.75245047      3.31913185   -10.09435272    -0.83400000
0.00024239      3.15057423
     -5.01067591      3.26426482    -9.13492489     0.41700000
0.00007331      0.40001352
     -4.87542582      4.27081680   -10.23380089     0.41700000
0.00007331      0.40001352
    -11.10755942      6.90023927    -1.86551757    -0.83400000
0.00024239      3.15057423
    -11.67873881      6.29674725    -2.37477489     0.41700000
0.00007331      0.40001352
    -11.42030928      7.74367647    -2.26106115     0.41700000
0.00007331      0.40001352
     -9.95538367     -7.17085660    -1.88822378    -0.83400000
0.00024239      3.15057423
    -10.81655134     -6.88799012    -2.17057241     0.41700000
0.00007331      0.40001352
     -9.58955277     -6.44168390    -2.42013372     0.41700000
0.00007331      0.40001352
     -8.20581011     11.30084059     1.41081138    -0.83400000
0.00024239      3.15057423
     -7.73376469     10.50504038     1.12831684     0.41700000
0.00007331      0.40001352
     -8.92558913     10.99959776     1.94595142     0.41700000
0.00007331      0.40001352
      0.16168584     -2.01792860    -7.50541592    -0.83400000
0.00024239      3.15057423
      0.34712616     -2.90599823    -7.07846880     0.41700000
0.00007331      0.40001352
     -0.60477573     -1.85593629    -6.95776033     0.41700000
0.00007331      0.40001352
     -2.12588978      3.02616072    -6.71686792    -0.83400000
0.00024239      3.15057423
     -3.03456879      2.88080072    -6.92847157     0.41700000
0.00007331      0.40001352
     -1.76671541      3.39728117    -7.53661919     0.41700000
0.00007331      0.40001352
     -6.14285946      7.88858700    -5.11608028    -0.83400000
0.00024239      3.15057423
     -7.08336496      7.83981323    -4.87488365     0.41700000
0.00007331      0.40001352
     -5.97302198      6.92667627    -5.08902025     0.41700000
0.00007331      0.40001352
      1.81689727      5.69858503    -7.30674553    -0.83400000
0.00024239      3.15057423
      1.59058917      4.86141920    -7.68436670     0.41700000
0.00007331      0.40001352
      2.70646715      5.43788433    -6.92526436     0.41700000
0.00007331      0.40001352
     -6.23954790      5.53375944    -9.18444555    -0.83400000
0.00024239      3.15057423
     -6.95070474      6.09495386    -8.81293361     0.41700000
0.00007331      0.40001352
     -6.49491709      4.66287192    -8.81128567     0.41700000
0.00007331      0.40001352
```

```
     -9.15484623      11.43463251      -1.47783427      -0.83400000
0.00024239      3.15057423
     -8.91523747      12.08645365      -0.77841572       0.41700000
0.00007331      0.40001352
    -10.12292866      11.60295698      -1.57418231       0.41700000
0.00007331      0.40001352
     -4.96940088       7.32975769       5.59079790      -0.83400000
0.00024239      3.15057423
     -4.97037077       8.25142002       5.25582504       0.41700000
0.00007331      0.40001352
     -4.40525436       6.91930676       4.96578026       0.41700000
0.00007331      0.40001352
     -0.26684002      10.00807020     -10.06604914      -0.83400000
0.00024239      3.15057423
      0.03129144       9.58866331      -9.17263989       0.41700000
0.00007331      0.40001352
     -0.95131163      10.56023714      -9.62617354       0.41700000
0.00007331      0.40001352
    -11.15907404       5.46717755       0.61493973      -0.83400000
0.00024239      3.15057423
    -11.00866053       5.88342504      -0.24188514       0.41700000
0.00007331      0.40001352
    -10.59244559       5.96438555       1.28520636       0.41700000
0.00007331      0.40001352
     -4.83281708      -6.92203426      -1.25783682      -0.83400000
0.00024239      3.15057423
     -4.91476154      -6.10188293      -0.64053756       0.41700000
0.00007331      0.40001352
     -3.88136363      -6.92576265      -1.46580851       0.41700000
0.00007331      0.40001352
     -1.64866905       9.96153948      -6.94652442      -0.83400000
0.00024239      3.15057423
     -1.67398457      10.90520594      -7.14385967       0.41700000
0.00007331      0.40001352
     -1.66991572      10.06001780      -5.95639495       0.41700000
0.00007331      0.40001352
     -0.23070198       5.22515583      -2.17702746      -0.83400000
0.00024239      3.15057423
     -0.68355227       5.50536346      -3.00527048       0.41700000
0.00007331      0.40001352
      0.65467203       5.28325462      -2.50035000       0.41700000
0.00007331      0.40001352
      0.89187348       0.68845546      -4.60773897      -0.83400000
0.00024239      3.15057423
      1.56052935       1.42618001      -4.32142544       0.41700000
0.00007331      0.40001352
      0.97568953       0.66868222      -5.61498117       0.41700000
0.00007331      0.40001352
    -10.85094322      -1.28269304      -8.74101127      -0.83400000
0.00024239      3.15057423
    -11.16956712      -0.75009549      -9.47861684       0.41700000
0.00007331      0.40001352
    -10.24647237      -0.64405835      -8.33378614       0.41700000
0.00007331      0.40001352
     -3.79904056      -0.90402341      -9.47147846      -0.83400000
```

```
0.00024239      3.15057423
     -4.06642818    -1.54297495    -8.79240990     0.41700000
0.00007331      0.40001352
     -4.66850042    -0.62892956    -9.77049541     0.41700000
0.00007331      0.40001352
     -6.40939454    12.73173353    -2.70289318    -0.83400000
0.00024239      3.15057423
     -6.89332251    11.97487089    -2.36274592     0.41700000
0.00007331      0.40001352
     -5.60239176    12.65092108    -2.20562856     0.41700000
0.00007331      0.40001352
      0.18268676    12.24280665    -2.97719757    -0.83400000
0.00024239      3.15057423
     -0.32673602    12.72819779    -2.30513470     0.41700000
0.00007331      0.40001352
     -0.16427783    12.64990208    -3.74899320     0.41700000
0.00007331      0.40001352
      1.70999646     2.95553684    -8.22914314    -0.83400000
0.00024239      3.15057423
      1.62235367     2.07860875    -7.79374266     0.41700000
0.00007331      0.40001352
      0.96318227     2.97343135    -8.90891552     0.41700000
0.00007331      0.40001352
    -10.94326019    -5.29858589     0.67193198    -0.83400000
0.00024239      3.15057423
    -11.55381584    -4.77557135     0.18947107     0.41700000
0.00007331      0.40001352
    -11.55175591    -5.99943066     1.06350756     0.41700000
0.00007331      0.40001352
    -14.95601198    -0.71559520    -5.01344677    -0.83400000
0.00024239      3.15057423
    -14.32886336    -0.70913930    -5.71967311     0.41700000
0.00007331      0.40001352
    -15.76153776    -0.45086856    -5.52604957     0.41700000
0.00007331      0.40001352
     -5.94910002     9.58284092     3.54445243    -0.83400000
0.00024239      3.15057423
     -5.84677458     8.84972477     2.95315051     0.41700000
0.00007331      0.40001352
     -6.83834124     9.41940403     3.83400202     0.41700000
0.00007331      0.40001352
     -1.56935549     0.35127306    -9.18744278    -0.83400000
0.00024239      3.15057423
     -2.36397362    -0.17355590    -9.39029121     0.41700000
0.00007331      0.40001352
     -1.75811040     0.50599879    -8.16943741     0.41700000
0.00007331      0.40001352
    -14.07500622    -3.92943330     1.44328885    -0.83400000
0.00024239      3.15057423
    -13.79086134    -3.88089319     0.51646833     0.41700000
0.00007331      0.40001352
    -13.71641657    -3.12310978     1.84860949     0.41700000
0.00007331      0.40001352
    -14.20119565    -0.15306178    -9.65109794    -0.83400000
0.00024239      3.15057423
```

```
    -13.31239956      -0.01980678    -10.02959602      0.41700000
0.00007331      0.40001352
    -14.23207479       0.72270402     -9.13877456      0.41700000
0.00007331      0.40001352
     -4.26881552       7.71750784     -1.41377366     -0.83400000
0.00024239      3.15057423
     -4.41654110       8.23268509     -2.18154502      0.41700000
0.00007331      0.40001352
     -4.30856514       8.49301434     -0.75255018      0.41700000
0.00007331      0.40001352
     -4.74750484      -3.37516130    -15.79960478     -0.83400000
0.00024239      3.15057423
     -4.63706268      -2.38356889    -15.73057712      0.41700000
0.00007331      0.40001352
     -4.68567718      -3.63342202    -14.85764683      0.41700000
0.00007331      0.40001352
     -4.04387140      -3.49535775     -3.62332439     -0.83400000
0.00024239      3.15057423
     -3.56032968      -3.52109051     -2.79814959      0.41700000
0.00007331      0.40001352
     -3.76096344      -2.59369636     -3.92540073      0.41700000
0.00007331      0.40001352
     -4.61556101      -1.24008286      4.84261656     -0.83400000
0.00024239      3.15057423
     -5.14179945      -0.44616830      4.54327583      0.41700000
0.00007331      0.40001352
     -5.03048182      -1.94250596      4.28915596      0.41700000
0.00007331      0.40001352
     -1.35502505       2.50174856     -4.15784693     -0.83400000
0.00024239      3.15057423
     -0.55688095       2.08537936     -4.51790762      0.41700000
0.00007331      0.40001352
     -1.81425250       2.67176390     -5.01293898      0.41700000
0.00007331      0.40001352
     -4.02193737       3.84920478      1.39791763     -0.83400000
0.00024239      3.15057423
     -3.83969712       4.04143286      0.46111229      0.41700000
0.00007331      0.40001352
     -3.27061915       4.43042898      1.79557204      0.41700000
0.00007331      0.40001352
     -4.14192867       4.14322472      6.12727976     -0.83400000
0.00024239      3.15057423
     -3.47189999       4.11018562      6.82111549      0.41700000
0.00007331      0.40001352
     -3.87384987       4.94566154      5.64689922      0.41700000
0.00007331      0.40001352
     -6.97204511       4.99274978    -12.98970979     -0.83400000
0.00024239      3.15057423
     -7.11722868       5.04745182    -13.95340984      0.41700000
0.00007331      0.40001352
     -6.07540624       5.26901039    -12.69357722      0.41700000
0.00007331      0.40001352
    -11.15246794      -0.81246515     -1.01006361     -0.83400000
0.00024239      3.15057423
    -11.65789435      -0.15057273     -0.48734994      0.41700000
```

```
0.00007331      0.40001352
   -10.98611185    -1.53026434    -0.37222715     0.41700000
0.00007331      0.40001352
   -10.49845339    -2.72003377    -4.31350005    -0.83400000
0.00024239      3.15057423
   -10.03901196    -3.47029699    -3.88716806     0.41700000
0.00007331      0.40001352
   -11.21078218    -3.11810888    -4.82077945     0.41700000
0.00007331      0.40001352
   -10.43672371    -2.58069944     0.94477940    -0.83400000
0.00024239      3.15057423
   -10.76564312    -2.61942196     1.87792277     0.41700000
0.00007331      0.40001352
   -10.24156666    -3.48802614     0.75041306     0.41700000
0.00007331      0.40001352
     0.91760345     9.73693964    -7.59769754    -0.83400000
0.00024239      3.15057423
     1.23726268    10.49232313    -7.09701375     0.41700000
0.00007331      0.40001352
    -0.02289228     9.75407335    -7.26275258     0.41700000
0.00007331      0.40001352
    -5.99805737     0.84667850     3.87888217    -0.83400000
0.00024239      3.15057423
    -5.94073153     1.76076257     4.15010500     0.41700000
0.00007331      0.40001352
    -6.92460632     0.94163722     3.55427098     0.41700000
0.00007331      0.40001352
    -2.01993370    -2.41292572     4.75738716    -0.83400000
0.00024239      3.15057423
    -2.73513818    -1.84698534     5.13407993     0.41700000
0.00007331      0.40001352
    -2.52957535    -3.24370980     4.59012318     0.41700000
0.00007331      0.40001352
    -5.43073265    10.00289938    -6.50841932    -0.83400000
0.00024239      3.15057423
    -5.61131529     9.02838728    -6.27480845     0.41700000
0.00007331      0.40001352
    -4.66234718    10.16112539    -5.88907926     0.41700000
0.00007331      0.40001352
   -11.03087748    -7.08904613    -4.87601007    -0.83400000
0.00024239      3.15057423
   -10.90402795    -6.16890395    -4.59643758     0.41700000
0.00007331      0.40001352
   -10.80614556    -7.09166301    -5.84391130     0.41700000
0.00007331      0.40001352
    -3.63956523    -3.36325860    -0.87198746    -0.83400000
0.00024239      3.15057423
    -3.94332099    -2.65172744    -0.25449085     0.41700000
0.00007331      0.40001352
    -3.69248080    -4.07021904    -0.17554607     0.41700000
0.00007331      0.40001352
   -10.08251974     5.72401301     4.71755896    -0.83400000
0.00024239      3.15057423
    -9.62992594     4.95266090     4.32068686     0.41700000
0.00007331      0.40001352
```

```
    -10.86702463        5.21120922        5.15894978        0.41700000
0.00007331        0.40001352
     -4.13924782       11.85287878       -1.60033063       -0.83400000
0.00024239        3.15057423
     -4.31139831       11.72317812       -0.65656475        0.41700000
0.00007331        0.40001352
     -4.23945564       10.91515658       -1.76205759        0.41700000
0.00007331        0.40001352
     -4.39291442       -4.17281164      -13.18898035       -0.83400000
0.00024239        3.15057423
     -3.98017515       -3.53256524      -12.49230824        0.41700000
0.00007331        0.40001352
     -3.69411672       -4.87217391      -13.18280781        0.41700000
0.00007331        0.40001352
      1.20678234       -5.24383259       -8.92506409       -0.83400000
0.00024239        3.15057423
      1.35684121       -6.17150831       -8.64550972        0.41700000
0.00007331        0.40001352
      1.96638489       -5.15202427       -9.56377506        0.41700000
0.00007331        0.40001352
     -9.16294765        1.08909702       -3.71860242       -0.83400000
0.00024239        3.15057423
     -8.78751564        0.20517330       -3.61345363        0.41700000
0.00007331        0.40001352
    -10.02747917        0.90368420       -4.07745409        0.41700000
0.00007331        0.40001352
     -0.67135024        2.72109532        6.92797995       -0.83400000
0.00024239        3.15057423
     -0.64405292        2.37443161        7.87056541        0.41700000
0.00007331        0.40001352
     -0.21395230        3.57453251        7.07770300        0.41700000
0.00007331        0.40001352
    -10.68309562        3.89056596      -14.85120957       -0.83400000
0.00024239        3.15057423
    -11.01002043        4.37402638      -14.05275243        0.41700000
0.00007331        0.40001352
    -11.09756201        4.42782291      -15.53460645        0.41700000
0.00007331        0.40001352
     -1.02736950       -3.80126142        1.77678549       -0.83400000
0.00024239        3.15057423
     -0.28711194       -4.12946701        2.31694460        0.41700000
0.00007331        0.40001352
     -1.66594553       -3.93872166        2.52733207        0.41700000
0.00007331        0.40001352
      1.71680295        8.01037121       -4.37776756       -0.83400000
0.00024239        3.15057423
      2.63868952        8.24894619       -4.36792231        0.41700000
0.00007331        0.40001352
      1.45998514        8.57782555       -3.65151715        0.41700000
0.00007331        0.40001352
      1.53728318        9.93886471       -2.37570119       -0.83400000
0.00024239        3.15057423
      1.00594270       10.64303112       -2.76150298        0.41700000
0.00007331        0.40001352
      2.34552383       10.47169781       -2.33500862        0.41700000
```

```
0.00007331      0.40001352
   -10.53553773     -6.64651454     -7.58977521     -0.83400000
0.00024239     3.15057423
   -10.46895374     -5.78948797     -8.05549563      0.41700000
0.00007331      0.40001352
   -11.13511670     -7.07794488     -8.25828088      0.41700000
0.00007331      0.40001352
   -15.01429571     -3.79450239     -4.26328529     -0.83400000
0.00024239     3.15057423
   -15.81602348     -3.94996084     -4.77125896      0.41700000
0.00007331      0.40001352
   -15.03899062     -4.56679166     -3.66648496      0.41700000
0.00007331      0.40001352
    -5.38185132     -6.40675128    -10.09477533     -0.83400000
0.00024239     3.15057423
    -5.30530752     -6.91480697     -9.30909980      0.41700000
0.00007331      0.40001352
    -5.07502664     -7.15343488    -10.65338577      0.41700000
0.00007331      0.40001352
    -0.06777705     -1.54545724      3.20003581     -0.83400000
0.00024239     3.15057423
    -0.67194760     -1.83165503      3.93613744      0.41700000
0.00007331      0.40001352
    -0.39186412     -2.25184846      2.60409546      0.41700000
0.00007331      0.40001352
   -10.30944066     10.89215967     -4.46063272     -0.83400000
0.00024239     3.15057423
   -10.69502835     11.15455458     -3.60946570      0.41700000
0.00007331      0.40001352
    -9.70807843     11.63192484     -4.68163048      0.41700000
0.00007331      0.40001352
     0.61740971     -4.68050623     -6.34114838     -0.83400000
0.00024239     3.15057423
     0.78671414     -5.15730524     -7.22121811      0.41700000
0.00007331      0.40001352
     0.98086965     -5.31479073     -5.70090866      0.41700000
0.00007331      0.40001352
    -8.64982002      6.60391446     -9.13670509     -0.83400000
0.00024239     3.15057423
    -8.39039152      7.40284273     -8.62461298      0.41700000
0.00007331      0.40001352
    -8.59633653      6.94826933    -10.04805630      0.41700000
0.00007331      0.40001352
     3.02723942      5.99347262    -10.25929568     -0.83400000
0.00024239     3.15057423
     2.08526926      6.02522592    -10.30489561      0.41700000
0.00007331      0.40001352
     3.18437706      6.67438893     -9.59243223      0.41700000
0.00007331      0.40001352
    -2.38154745     -2.35321260    -11.34691048     -0.83400000
0.00024239     3.15057423
    -2.93011117     -1.79772305    -10.80638409      0.41700000
0.00007331      0.40001352
    -1.74729598     -2.70422769    -10.63128185      0.41700000
0.00007331      0.40001352
```

```
    -12.44829056        1.53020648      -2.10830064       -0.83400000
0.00024239        3.15057423
    -12.23021910        0.81421022      -2.72519775        0.41700000
0.00007331        0.40001352
    -13.35727379        1.31450180      -1.91980024        0.41700000
0.00007331        0.40001352
      3.43269563       -2.69687724      -0.10293386       -0.83400000
0.00024239        3.15057423
      2.54266810       -3.08579445       0.04740809        0.41700000
0.00007331        0.40001352
      3.14372587       -1.90356135      -0.56189650        0.41700000
0.00007331        0.40001352
      2.08547902        2.45265818       2.48892140       -0.83400000
0.00024239        3.15057423
      2.08475614        1.52860451       2.68032646        0.41700000
0.00007331        0.40001352
      1.23709583        2.70122695       2.87923908        0.41700000
0.00007331        0.40001352
     -0.62606150        1.70599651       4.35170889       -0.83400000
0.00024239        3.15057423
     -1.01177561        1.91791201       5.24454880        0.41700000
0.00007331        0.40001352
      0.11095714        1.07372582       4.50905991        0.41700000
0.00007331        0.40001352
     -8.72354126        0.06846704       4.47919989       -0.83400000
0.00024239        3.15057423
     -8.62431431        0.15579160       5.45167637        0.41700000
0.00007331        0.40001352
     -9.00101471        0.95212418       4.19547892        0.41700000
0.00007331        0.40001352
     -8.08821886        8.80652806      -7.76312845       -0.83400000
0.00024239        3.15057423
     -8.08538168        8.96548381      -6.83112448        0.41700000
0.00007331        0.40001352
     -8.69755762        9.43573841      -7.99723547        0.41700000
0.00007331        0.40001352
      1.32351474        8.99528412     -11.79206106       -0.83400000
0.00024239        3.15057423
      1.19973521        8.11135249     -11.39570257        0.41700000
0.00007331        0.40001352
      0.69346599        9.48115878     -11.20606730        0.41700000
0.00007331        0.40001352
     -2.38850546       -4.94034481       3.80965996       -0.83400000
0.00024239        3.15057423
     -1.75714445       -5.59468269       4.09257984        0.41700000
0.00007331        0.40001352
     -3.20031834       -5.54584837       3.98892403        0.41700000
0.00007331        0.40001352
      4.39702923        3.76032346     -11.27697775       -0.83400000
0.00024239        3.15057423
      3.74555446        3.04366396     -11.42017958        0.41700000
0.00007331        0.40001352
      3.90806332        4.49140419     -10.96205828        0.41700000
0.00007331        0.40001352
     -4.85812473       -1.44872844       0.77312088       -0.83400000
```

```
0.00024239       3.15057423
     -4.36791420     -0.76636475      1.20949137      0.41700000
0.00007331       0.40001352
     -5.61824656     -1.56889963      1.34408581      0.41700000
0.00007331       0.40001352
     -9.74353950      9.42327854    -10.71633022     -0.83400000
0.00024239       3.15057423
    -10.20953052      8.74405369    -10.18987100      0.41700000
0.00007331       0.40001352
     -9.10131710      8.90507975    -11.18247001      0.41700000
0.00007331       0.40001352
      2.53637052     -1.17628205     -2.58882594     -0.83400000
0.00024239       3.15057423
      2.01252079     -0.63970524     -2.00847268      0.41700000
0.00007331       0.40001352
      2.36895680     -0.67201531     -3.42464590      0.41700000
0.00007331       0.40001352
      1.44744551     -5.98655796     -1.29462314     -0.83400000
0.00024239       3.15057423
      1.17277491     -5.18924332     -0.91189516      0.41700000
0.00007331       0.40001352
      2.32334924     -5.80902910     -1.66169631      0.41700000
0.00007331       0.40001352
     -6.33511591     -0.01520389    -10.34225082     -0.83400000
0.00024239       3.15057423
     -7.22906637      0.03534056     -9.94759083      0.41700000
0.00007331       0.40001352
     -6.47330046     -0.67812967    -11.02531242      0.41700000
0.00007331       0.40001352
     -3.70592165     -5.71443033     -6.87193918     -0.83400000
0.00024239       3.15057423
     -3.29449630     -5.88120127     -5.99348688      0.41700000
0.00007331       0.40001352
     -4.52918863     -6.21176195     -6.74052525      0.41700000
0.00007331       0.40001352
     -7.11906910     -4.18849802      1.37515175     -0.83400000
0.00024239       3.15057423
     -7.52234268     -5.03166389      0.95673203      0.41700000
0.00007331       0.40001352
     -6.63077593     -3.80992174      0.57103395      0.41700000
0.00007331       0.40001352
     -8.28528807      6.87247042      6.68962078     -0.83400000
0.00024239       3.15057423
     -7.75805018      6.03710632      6.71104602      0.41700000
0.00007331       0.40001352
     -8.92442248      6.46918635      6.10256557      0.41700000
0.00007331       0.40001352
     -0.97797883     -3.78068852     -9.42269993     -0.83400000
0.00024239       3.15057423
     -0.10935444     -4.10967207     -9.11183834      0.41700000
0.00007331       0.40001352
     -1.52983975     -4.60139513     -9.34214687      0.41700000
0.00007331       0.40001352
     -7.38302422      3.12178922     -7.92789936     -0.83400000
0.00024239       3.15057423
```

```
    -8.27246284        3.35031962      -8.32844925       0.41700000
0.00007331       0.40001352
    -7.70795059        2.82963705      -7.04745626       0.41700000
0.00007331       0.40001352
    -7.88662446       -0.96222890     -14.28459013      -0.83400000
0.00024239       3.15057423
    -7.52294076       -1.23327173     -13.45680607       0.41700000
0.00007331       0.40001352
    -7.24290098       -0.25180829     -14.55416072       0.41700000
0.00007331       0.40001352
   -12.82983320        4.12625536      -8.06429594      -0.83400000
0.00024239       3.15057423
   -12.13068145        4.72974762      -7.78295343       0.41700000
0.00007331       0.40001352
   -13.16165631        4.46152433      -8.95243662       0.41700000
0.00007331       0.40001352
    -4.70759682       12.63666985       0.90494724      -0.83400000
0.00024239       3.15057423
    -5.47761945       12.64068434       1.48255868       0.41700000
0.00007331       0.40001352
    -3.96486107       12.72035143       1.59445711       0.41700000
0.00007331       0.40001352
    -1.24187362        4.21707201       0.07136906      -0.83400000
0.00024239       3.15057423
    -2.14307785        4.44525146      -0.25578910       0.41700000
0.00007331       0.40001352
    -0.66912919        4.69755125      -0.56244379       0.41700000
0.00007331       0.40001352
    -7.26906395        4.95614481       2.09572673      -0.83400000
0.00024239       3.15057423
    -7.13686609        4.41562796       2.87547636       0.41700000
0.00007331       0.40001352
    -7.08530760        4.16864300       1.46855402       0.41700000
0.00007331       0.40001352
     3.01701570       -4.49612761     -10.81339359      -0.83400000
0.00024239       3.15057423
     2.50256014       -3.82992887     -11.23882484       0.41700000
0.00007331       0.40001352
     2.92421103       -5.16518164     -11.55401802       0.41700000
0.00007331       0.40001352
   -12.21378527       -5.08707012      -3.07905973      -0.83400000
0.00024239       3.15057423
   -12.14048893       -4.51026452      -3.82017339       0.41700000
0.00007331       0.40001352
   -12.40157674       -5.90906490      -3.49503053       0.41700000
0.00007331       0.40001352
   -10.05143850        1.94935616     -10.18584316      -0.83400000
0.00024239       3.15057423
    -9.50594919        2.15841588     -10.95662229       0.41700000
0.00007331       0.40001352
   -10.07889145        2.79037722      -9.62124269       0.41700000
0.00007331       0.40001352
    -8.82584394       -9.10530961      -4.30116571      -0.83400000
0.00024239       3.15057423
    -9.49099816       -8.40333713      -4.14527286       0.41700000
```

```
0.00007331       0.40001352
     -8.32374943     -8.85117400     -5.12318147      0.41700000
0.00007331       0.40001352
      1.37454128      0.25868979     -7.27763844     -0.83400000
0.00024239       3.15057423
      2.33860111      0.10850611     -7.38796854      0.41700000
0.00007331       0.40001352
      0.99464637     -0.53343081     -7.71353245      0.41700000
0.00007331       0.40001352
    -11.18325753     -4.23153699     -8.70698727     -0.83400000
0.00024239       3.15057423
    -10.78524161     -4.06742681     -9.57790888      0.41700000
0.00007331       0.40001352
    -11.25484593     -3.36790002     -8.24308670      0.41700000
0.00007331       0.40001352
     -7.12915611      8.89290428      0.73002899     -0.83400000
0.00024239       3.15057423
     -6.33329964      8.44900608      1.14337349      0.41700000
0.00007331       0.40001352
     -7.88682985      8.37209320      1.08534729      0.41700000
0.00007331       0.40001352
      5.38012457     -0.54735911     -2.27368593     -0.83400000
0.00024239       3.15057423
      4.51870298     -0.15049253     -1.96952999      0.41700000
0.00007331       0.40001352
      5.07495451     -0.94536340     -3.09382677      0.41700000
0.00007331       0.40001352
     -3.46732783      4.77845144     -3.82093167     -0.83400000
0.00024239       3.15057423
     -3.38476062      5.57009029     -4.33427954      0.41700000
0.00007331       0.40001352
     -2.67004585      4.25333261     -3.99563122      0.41700000
0.00007331       0.40001352
    -14.55496531      4.38153252     -5.58311479     -0.83400000
0.00024239       3.15057423
    -13.80156510      3.96126088     -6.13437097      0.41700000
0.00007331       0.40001352
    -14.31428247      5.30897459     -5.42915075      0.41700000
0.00007331       0.40001352
     -2.72425508      6.60446930      1.62740326     -0.83400000
0.00024239       3.15057423
     -2.32749271      7.46156549      1.71987069      0.41700000
0.00007331       0.40001352
     -2.76245975      6.48467875      0.67839438      0.41700000
0.00007331       0.40001352
      3.33107523      8.43439345     -8.54588148     -0.83400000
0.00024239       3.15057423
      2.40179531      8.72522597     -8.24057886      0.41700000
0.00007331       0.40001352
      3.17592756      8.79769282     -9.48630545      0.41700000
0.00007331       0.40001352
     -1.44419694     -0.15008463      0.90180939     -0.83400000
0.00024239       3.15057423
     -0.93585372      0.05823188      1.70593786      0.41700000
0.00007331       0.40001352
```

229

```
    -2.40071201      -0.01339579      1.11363053      0.41700000
0.00007331      0.40001352
   -14.20629809      0.83022694      2.10712771     -0.83400000
0.00024239      3.15057423
   -14.36266777      1.76458744      1.91583304      0.41700000
0.00007331      0.40001352
   -15.13752863      0.53048877      2.00985317      0.41700000
0.00007331      0.40001352
    -6.00994361     -3.64904512     -9.74807466     -0.83400000
0.00024239      3.15057423
    -6.02327836     -4.62967027     -9.87929333      0.41700000
0.00007331      0.40001352
    -6.41130699     -3.47952855     -8.90855659      0.41700000
0.00007331      0.40001352
    -9.72279089      4.08046731     -8.47114532     -0.83400000
0.00024239      3.15057423
   -10.14164751      4.07075199     -7.58354585      0.41700000
0.00007331      0.40001352
    -9.45448654      4.93826684     -8.67747467      0.41700000
0.00007331      0.40001352
     2.36111903      3.35652876     -1.40611970     -0.83400000
0.00024239      3.15057423
     1.52693748      2.91525888     -1.57490945      0.41700000
0.00007331      0.40001352
     2.59196639      3.60032487     -2.29153204      0.41700000
0.00007331      0.40001352
    -8.00023096      7.26906427    -11.82199018     -0.83400000
0.00024239      3.15057423
    -7.72904127      6.32404229    -12.08138340      0.41700000
0.00007331      0.40001352
    -7.87810343      7.72599202    -12.70340793      0.41700000
0.00007331      0.40001352
   -10.47480318      9.94727044      2.94826369     -0.83400000
0.00024239      3.15057423
   -11.02077982      9.21269707      2.60963969      0.41700000
0.00007331      0.40001352
    -9.88222429      9.54147629      3.63387061      0.41700000
0.00007331      0.40001352
    -2.71831846      7.26807117     -4.91059446     -0.83400000
0.00024239      3.15057423
    -1.83354998      6.95340204     -5.23628330      0.41700000
0.00007331      0.40001352
    -2.42383814      8.02316856     -4.36898947      0.41700000
0.00007331      0.40001352
    -3.22734318     11.43752596     -4.74619107     -0.83400000
0.00024239      3.15057423
    -3.88539916     11.19878885     -4.05698187      0.41700000
0.00007331      0.40001352
    -3.16146670     12.40673325     -4.64846705      0.41700000
0.00007331      0.40001352
   -12.79879639      8.49569516      1.97440962     -0.83400000
0.00024239      3.15057423
   -13.00420162      9.23681168      1.39012246      0.41700000
0.00007331      0.40001352
   -13.20748398      7.75041036      1.56776909      0.41700000
```

```
0.00007331       0.40001352
     -0.05857183       9.00719643       0.01041870      -0.83400000
0.00024239       3.15057423
      0.63110948       8.64094925       0.64670825       0.41700000
0.00007331       0.40001352
      0.56038725       9.08142948      -0.73362136       0.41700000
0.00007331       0.40001352
     -7.09304714      -3.84997964      -1.75053632      -0.83400000
0.00024239       3.15057423
     -6.13900280      -3.86057234      -1.52202857       0.41700000
0.00007331       0.40001352
     -7.32928896      -2.90759778      -1.78680813       0.41700000
0.00007331       0.40001352
      0.51643757       6.08165411     -11.06615755      -0.83400000
0.00024239       3.15057423
     -0.45160194       5.91537253     -10.87063524       0.41700000
0.00007331       0.40001352
      0.61144309       5.56039111     -11.89289114       0.41700000
0.00007331       0.40001352
     -0.30671766       2.19889760      -1.54712915      -0.83400000
0.00024239       3.15057423
     -0.69333565       2.51956201      -2.39870834       0.41700000
0.00007331       0.40001352
     -0.66089958       2.84127760      -0.89206445       0.41700000
0.00007331       0.40001352
     -5.22856943       9.59043715      -3.37400214      -0.83400000
0.00024239       3.15057423
     -5.26549594       9.04799483      -4.17350060       0.41700000
0.00007331       0.40001352
     -6.07296972       9.35296080      -2.90561656       0.41700000
0.00007331       0.40001352
     -5.75866365      -5.57629251      -4.23175621      -0.83400000
0.00024239       3.15057423
     -5.78828192      -6.01294804      -3.36458659       0.41700000
0.00007331       0.40001352
     -5.03877687      -4.88105726      -4.05672073       0.41700000
0.00007331       0.40001352
     -0.35178292       6.88002110      -5.96708536      -0.83400000
0.00024239       3.15057423
      0.05929075       6.69217491      -6.82010269       0.41700000
0.00007331       0.40001352
      0.36811858       7.45860863      -5.52579689       0.41700000
0.00007331       0.40001352
     -2.61745942      -9.06752909      -7.69998706      -0.83400000
0.00024239       3.15057423
     -2.17233766      -9.12680710      -6.82903446       0.41700000
0.00007331       0.40001352
     -1.82279123      -9.02951897      -8.24847616       0.41700000
0.00007331       0.40001352
    -11.48333380       4.13632177      -5.11629863      -0.83400000
0.00024239       3.15057423
    -11.97418425       4.47260411      -4.33130355       0.41700000
0.00007331       0.40001352
    -10.71379969       4.80816965      -5.07707115       0.41700000
0.00007331       0.40001352
```

```
    -11.71560955     -0.36630845     -3.79901981     -0.83400000
0.00024239     3.15057423
    -11.89199448     -0.63953209     -2.91745734      0.41700000
0.00007331     0.40001352
    -11.41184425     -1.17987430     -4.14404202      0.41700000
0.00007331     0.40001352
     -3.86429667      0.73954272      2.21400285     -0.83400000
0.00024239     3.15057423
     -4.68699312      0.75686336      2.73384047      0.41700000
0.00007331     0.40001352
     -3.70628762      1.68097520      2.28645182      0.41700000
0.00007331     0.40001352
     -8.15234089      2.87343121     -5.31772280     -0.83400000
0.00024239     3.15057423
     -8.38430595      2.13287497     -4.75802040      0.41700000
0.00007331     0.40001352
     -8.59313869      3.51260829     -4.68990803      0.41700000
0.00007331     0.40001352
    -13.90095207     -1.92863603     -2.64111324     -0.83400000
0.00024239     3.15057423
    -14.04746172     -2.68678375     -3.22457357      0.41700000
0.00007331     0.40001352
    -14.29936955     -1.32203813     -3.31445737      0.41700000
0.00007331     0.40001352
     -7.81349421     -8.34208012      1.75187039     -0.83400000
0.00024239     3.15057423
     -8.14758301     -7.58741093      1.27109015      0.41700000
0.00007331     0.40001352
     -8.31437016     -8.25825024      2.58634329      0.41700000
0.00007331     0.40001352
     -2.40375209     -0.54637957     -6.65046597     -0.83400000
0.00024239     3.15057423
     -3.09874248     -1.17887616     -6.82425451      0.41700000
0.00007331     0.40001352
     -2.29158998     -0.65922529     -5.67838287      0.41700000
0.00007331     0.40001352
    -12.12910626      2.02758436      4.74833201     -0.83400000
0.00024239     3.15057423
    -11.90799639      2.35200636      3.82883370      0.41700000
0.00007331     0.40001352
    -12.94300863      2.47341697      5.02678840      0.41700000
0.00007331     0.40001352
     -8.13464069      1.90512717      1.44615233     -0.83400000
0.00024239     3.15057423
     -8.76016331      1.75818563      0.71094149      0.41700000
0.00007331     0.40001352
     -7.80995941      0.98576325      1.45561659      0.41700000
0.00007331     0.40001352
      4.77623844      1.17368972      2.88786459     -0.83400000
0.00024239     3.15057423
      4.95773411      1.47990453      3.83131814      0.41700000
0.00007331     0.40001352
      3.92742062      1.64190423      2.80566359      0.41700000
0.00007331     0.40001352
      0.74211888      3.98345270    -12.58798620     -0.83400000
```

```
0.00024239      3.15057423
      0.60185914       4.04047788     -13.53514406       0.41700000
0.00007331      0.40001352
      0.97192983       3.06695114     -12.54137203       0.41700000
0.00007331      0.40001352
     -0.76769174      -7.02604163     -11.16328187      -0.83400000
0.00024239      3.15057423
     -0.38186277      -7.78137744     -10.63009954       0.41700000
0.00007331      0.40001352
     -1.22818960      -6.46307004     -10.47652889       0.41700000
0.00007331      0.40001352
      4.37003708       2.76025224      -6.92355156      -0.83400000
0.00024239      3.15057423
      3.92463827       3.15142441      -7.70465279       0.41700000
0.00007331      0.40001352
      4.50914335       3.59152150      -6.37731504       0.41700000
0.00007331      0.40001352
     -8.30797672      -1.39353228      -1.91907835      -0.83400000
0.00024239      3.15057423
     -9.26729870      -1.47033036      -2.11512351       0.41700000
0.00007331      0.40001352
     -8.23423767      -0.78818732      -1.15559924       0.41700000
0.00007331      0.40001352
     -2.28016353      -7.50444031      -1.97101736      -0.83400000
0.00024239      3.15057423
     -2.13940763      -8.43739891      -2.08793688       0.41700000
0.00007331      0.40001352
     -1.92836106      -7.39228487      -1.08472371       0.41700000
0.00007331      0.40001352
     -1.29852414      -2.00593233      -1.39815152      -0.83400000
0.00024239      3.15057423
     -2.06059861      -2.63118434      -1.22234380       0.41700000
0.00007331      0.40001352
     -1.35011625      -1.50407732      -0.53287137       0.41700000
0.00007331      0.40001352
     -1.25620568       9.07398510      -3.14377642      -0.83400000
0.00024239      3.15057423
     -0.98548353       8.96021366      -2.24652028       0.41700000
0.00007331      0.40001352
     -1.43802559       9.99148083      -3.28466463       0.41700000
0.00007331      0.40001352
     -2.59129238       7.64402485      -8.00587845      -0.83400000
0.00024239      3.15057423
     -2.31897163       8.41683292      -7.51239729       0.41700000
0.00007331      0.40001352
     -3.05295825       7.16875267      -7.28372812       0.41700000
0.00007331      0.40001352
    -11.90737121       0.19829425     -11.09515112      -0.83400000
0.00024239      3.15057423
    -11.97140186       0.54969701     -12.00583546       0.41700000
0.00007331      0.40001352
    -11.10978763       0.66691789     -10.84755104       0.41700000
0.00007331      0.40001352
     -6.82864778       1.72701606     -13.89898937      -0.83400000
0.00024239      3.15057423
```

```
    -7.48352640        2.39244613       -13.71178930         0.41700000
0.00007331       0.40001352
    -6.28011625        1.74514827       -13.09398358         0.41700000
0.00007331       0.40001352
     0.55798095       -1.10797775       -10.47128677        -0.83400000
0.00024239       3.15057423
    -0.24944159       -0.74474925        -9.97930431         0.41700000
0.00007331       0.40001352
     0.48203656       -2.05238032       -10.16823483         0.41700000
0.00007331       0.40001352
    -6.76074375       -5.44309247       -12.43423335        -0.83400000
0.00024239       3.15057423
    -6.75596727       -6.04231561       -11.68501709         0.41700000
0.00007331       0.40001352
    -5.81740344       -5.09057487       -12.50117249         0.41700000
0.00007331       0.40001352
    -4.91705132       -5.96508503        3.87170434         -0.83400000
0.00024239       3.15057423
    -5.13534880       -6.33181906        4.74279737          0.41700000
0.00007331       0.40001352
    -5.50008392       -5.18575287        3.93029785          0.41700000
0.00007331       0.40001352
   -10.94529121        6.49569544        -7.19758909        -0.83400000
0.00024239       3.15057423
   -10.64577310        7.17182794        -6.54275530         0.41700000
0.00007331       0.40001352
   -11.34200495        7.16338750        -7.77787702         0.41700000
0.00007331       0.40001352
     3.16204190        1.88890493        -4.03848839        -0.83400000
0.00024239       3.15057423
     3.97982430        1.96310699        -3.55065823         0.41700000
0.00007331       0.40001352
     3.45212984        2.54667330        -4.69871569         0.41700000
0.00007331       0.40001352
    -9.39131472        3.34740822        3.65268476         -0.83400000
0.00024239       3.15057423
   -10.35637018        3.26801233        3.31626825          0.41700000
0.00007331       0.40001352
    -8.87785456        2.97510366        2.96392123          0.41700000
0.00007331       0.40001352
     5.33299685       -0.72178423        1.01310968         -0.83400000
0.00024239       3.15057423
     4.98679018        0.00726878        1.51931453          0.41700000
0.00007331       0.40001352
     5.07496309       -0.48181382        0.13991784          0.41700000
0.00007331       0.40001352
    -6.10084486       -3.68351793        -6.15237331        -0.83400000
0.00024239       3.15057423
    -6.13695145       -4.61886549        -5.89592266         0.41700000
0.00007331       0.40001352
    -5.82811546       -3.25457215        -5.30410385         0.41700000
0.00007331       0.40001352
     0.76032954       -3.48379612        -0.22682017        -0.83400000
0.00024239       3.15057423
     0.42061335       -2.77902484        -0.75983793         0.41700000
```

```
0.00007331       0.40001352
    -0.05648462      -3.70602298       0.31795514       0.41700000
0.00007331       0.40001352
     4.12509823       5.38506174      -5.62322903      -0.83400000
0.00024239      3.15057423
     4.29394341       5.72260857      -4.74928808       0.41700000
0.00007331       0.40001352
     4.66634512       5.95089626      -6.21641684       0.41700000
0.00007331       0.40001352
    -2.69925857      -5.80725384      -9.22448254      -0.83400000
0.00024239      3.15057423
    -3.01268291      -5.70234489      -8.25645638       0.41700000
0.00007331       0.40001352
    -3.55955219      -5.70203733      -9.68144321       0.41700000
0.00007331       0.40001352
     5.03205967       6.04832220      -3.16290331      -0.83400000
0.00024239      3.15057423
     5.07969189       6.09124041      -2.19273758       0.41700000
0.00007331       0.40001352
     5.24470472       6.95099831      -3.37862182       0.41700000
0.00007331       0.40001352
    -9.36898136       5.41011095      -4.26578951      -0.83400000
0.00024239      3.15057423
    -8.72502232       6.13347483      -4.48604059       0.41700000
0.00007331       0.40001352
    -9.69041061       5.74940777      -3.38067007       0.41700000
0.00007331       0.40001352
   -11.61799833       3.08080570       2.14025084      -0.83400000
0.00024239      3.15057423
   -11.03162596       3.63134016       1.63971142       0.41700000
0.00007331       0.40001352
   -12.49525187       3.27264903       1.73668007       0.41700000
0.00007331       0.40001352
    -2.81103130       8.71794228     -12.89315531      -0.83400000
0.00024239      3.15057423
    -1.88338275       9.02276854     -12.98277447       0.41700000
0.00007331       0.40001352
    -2.73918958       8.32624678     -11.99997732       0.41700000
0.00007331       0.40001352
   -11.25334570      -2.65388866       3.42953793      -0.83400000
0.00024239      3.15057423
   -12.22408507      -2.47943588       3.28228919       0.41700000
0.00007331       0.40001352
   -11.15817663      -2.37865348       4.29516773       0.41700000
0.00007331       0.40001352
   -14.00799530       7.01025053      -9.47427433      -0.83400000
0.00024239      3.15057423
   -14.57171910       7.13068714      -8.68792837       0.41700000
0.00007331       0.40001352
   -14.04438238       6.06720861      -9.69673174       0.41700000
0.00007331       0.40001352
   -12.33667443       4.92322360       5.91163210      -0.83400000
0.00024239      3.15057423
   -12.77645323       5.80331640       5.76974515       0.41700000
0.00007331       0.40001352
```

```
   -13.07686398      4.25352408      5.70845715      0.41700000
0.00007331      0.40001352
    -4.50106001      9.82244968      0.20328268     -0.83400000
0.00024239      3.15057423
    -3.83117175      9.63473892      0.85173446      0.41700000
0.00007331      0.40001352
    -5.16375017     10.24154758      0.76081568      0.41700000
0.00007331      0.40001352
    -2.04785728      5.68396378    -10.46375847     -0.83400000
0.00024239      3.15057423
    -2.31726146      4.81767797    -10.26839066      0.41700000
0.00007331      0.40001352
    -2.11540723      6.24663305     -9.61080551      0.41700000
0.00007331      0.40001352
   -15.56483099      3.95638359     -3.07604356     -0.83400000
0.00024239      3.15057423
   -15.40316412      4.89086066     -3.05194088      0.41700000
0.00007331      0.40001352
   -15.20639679      3.72684411     -3.95270773      0.41700000
0.00007331      0.40001352
     1.06297350     -0.06558885     -0.39115554     -0.83400000
0.00024239      3.15057423
     0.63488048      0.47885665     -1.04606795      0.41700000
0.00007331      0.40001352
     0.32820570     -0.27375838      0.18539919      0.41700000
0.00007331      0.40001352
    -3.86939263     -2.79157066     -7.23537540     -0.83400000
0.00024239      3.15057423
    -3.52083158     -3.67018366     -7.43383360      0.41700000
0.00007331      0.40001352
    -4.79131699     -3.02509856     -7.00481272      0.41700000
0.00007331      0.40001352
     3.49416780      1.04973090     -0.55044836     -0.83400000
0.00024239      3.15057423
     3.15805387      1.95264065     -0.61162758      0.41700000
0.00007331      0.40001352
     2.66715837      0.52932620     -0.47689819      0.41700000
0.00007331      0.40001352
    -2.26489282      9.15015507      1.43185747     -0.83400000
0.00024239      3.15057423
    -1.52316463      9.17018604      0.77595973      0.41700000
0.00007331      0.40001352
    -2.15654683     10.03418350      1.85911751      0.41700000
0.00007331      0.40001352
    -7.63225603     -0.85161811      1.89244413     -0.83400000
0.00024239      3.15057423
    -7.57100630     -1.26256263      2.78277183      0.41700000
0.00007331      0.40001352
    -8.24349785     -1.45958865      1.43706441      0.41700000
0.00007331      0.40001352
     4.56507969      8.72389984     -4.05687284     -0.83400000
0.00024239      3.15057423
     5.49787426      8.77483273     -4.12891150      0.41700000
0.00007331      0.40001352
     4.51406193      9.59997559     -3.65144944      0.41700000
```

```
0.00007331      0.40001352
    -9.34895897      6.58639002      2.19171357     -0.83400000
0.00024239      3.15057423
    -9.33700466      6.57118511      3.15167809      0.41700000
0.00007331      0.40001352
    -8.59110069      6.00167227      1.94855523      0.41700000
0.00007331      0.40001352
    -9.33462048      1.23601043     -0.92071909     -0.83400000
0.00024239      3.15057423
   -10.27926826      1.31667566     -1.02176487      0.41700000
0.00007331      0.40001352
    -9.16207790      1.35826623     -1.84440970      0.41700000
0.00007331      0.40001352
    -1.01151693     -3.85261488     -4.38280582     -0.83400000
0.00024239      3.15057423
    -0.47555986     -3.88193035     -5.18132019      0.41700000
0.00007331      0.40001352
    -1.17485571     -2.93691993     -4.36469316      0.41700000
0.00007331      0.40001352
     2.05352044      5.26270914     -3.81057954     -0.83400000
0.00024239      3.15057423
     2.30209565      5.01257896     -4.75035810      0.41700000
0.00007331      0.40001352
     2.00577831      6.25401068     -3.82367873      0.41700000
0.00007331      0.40001352
    -5.97430611     -3.35284114      3.66271663     -0.83400000
0.00024239      3.15057423
    -6.34943104     -3.33122945      2.73790097      0.41700000
0.00007331      0.40001352
    -6.79119682     -3.41774416      4.22118616      0.41700000
0.00007331      0.40001352
    -4.07392979     -5.21918535      1.08563566     -0.83400000
0.00024239      3.15057423
    -3.44744134     -5.98561859      1.17205524      0.41700000
0.00007331      0.40001352
    -4.25162745     -5.11320877      2.05839086      0.41700000
0.00007331      0.40001352
    -2.21267319     -6.27845049     -4.61125135     -0.83400000
0.00024239      3.15057423
    -1.91711199     -5.39501381     -4.35645103      0.41700000
0.00007331      0.40001352
    -2.53220248     -6.62919712     -3.75691271      0.41700000
0.00007331      0.40001352
    -9.53280926     -4.88076448      3.80372620     -0.83400000
0.00024239      3.15057423
    -8.99044991     -4.07709312      3.85303903      0.41700000
0.00007331      0.40001352
   -10.20314503     -4.51203346      3.21993613      0.41700000
0.00007331      0.40001352
    -8.07159270     -7.82660449     -6.74893344     -0.83400000
0.00024239      3.15057423
    -8.77082718     -7.65391792     -7.43743528      0.41700000
0.00007331      0.40001352
    -7.36448348     -7.98657335     -7.31359351      0.41700000
0.00007331      0.40001352
```

```
   -11.19498127       5.76199588      -12.77836721       -0.83400000
0.00024239       3.15057423
   -10.77039545       6.31348988      -12.11654918        0.41700000
0.00007331       0.40001352
   -11.05010145       6.19977089      -13.65626829        0.41700000
0.00007331       0.40001352
    -3.46119308       5.24027777       -1.12712348       -0.83400000
0.00024239       3.15057423
    -3.91522670       6.10867834       -1.21982348        0.41700000
0.00007331       0.40001352
    -3.42272019       4.99315119       -2.04138613        0.41700000
0.00007331       0.40001352
     3.24086380      -3.51059532       -3.72262669       -0.83400000
0.00024239       3.15057423
     2.88096380      -2.71320820       -3.18550754        0.41700000
0.00007331       0.40001352
     3.92413592      -3.01082754       -4.34126472        0.41700000
0.00007331       0.40001352
     1.75788546       0.10917673        4.03318119       -0.83400000
0.00024239       3.15057423
     1.86394334      -0.17879789        4.93115330        0.41700000
0.00007331       0.40001352
     1.20584977      -0.65019161        3.76357555        0.41700000
0.00007331       0.40001352
    -4.31579109       1.46899585      -12.09816191       -0.83400000
0.00024239       3.15057423
    -5.03805347       1.03099184      -11.60303042        0.41700000
0.00007331       0.40001352
    -4.27556987       2.34382228      -11.59347937        0.41700000
0.00007331       0.40001352
    -8.67339897      -5.38950539       -3.40987897       -0.83400000
0.00024239       3.15057423
    -8.08180046      -5.60862160       -4.11813736        0.41700000
0.00007331       0.40001352
    -8.06923866      -4.93459272       -2.73378325        0.41700000
0.00007331       0.40001352
    -4.98947447       6.34384319      -11.47277468       -0.83400000
0.00024239       3.15057423
    -4.38051813       6.99816831      -11.06174009        0.41700000
0.00007331       0.40001352
    -5.62951296       6.29497203      -10.76720922        0.41700000
0.00007331       0.40001352
    -8.27862644      -6.22921848        0.11456603       -0.83400000
0.00024239       3.15057423
    -9.20361233      -5.98716402        0.10752918        0.41700000
0.00007331       0.40001352
    -8.23515320      -6.65712786       -0.77678341        0.41700000
0.00007331       0.40001352
   -12.20935461       7.45031266        4.53264741       -0.83400000
0.00024239       3.15057423
   -11.38798830       6.89944844        4.40146951        0.41700000
0.00007331       0.40001352
   -12.55618403       7.59442238        3.63910709        0.41700000
0.00007331       0.40001352
    -6.23802818      -2.05695356      -12.25023545       -0.83400000
```

```
0.00024239      3.15057423
    -6.13116277    -2.92388261    -11.77491231     0.41700000
0.00007331      0.40001352
    -5.61272014    -2.27334703    -12.91469634     0.41700000
0.00007331      0.40001352
    -6.09605598     3.33998847     4.70277500    -0.83400000
0.00024239      3.15057423
    -6.38206244     4.09658241     5.27146816     0.41700000
0.00007331      0.40001352
    -5.20117903     3.37160611     4.91095448     0.41700000
0.00007331      0.40001352
    -5.52061605     7.17474270     2.26865196    -0.83400000
0.00024239      3.15057423
    -6.20627928     6.43898106     2.35506892     0.41700000
0.00007331      0.40001352
    -4.70384932     6.67641544     2.00695515     0.41700000
0.00007331      0.40001352
   -13.65620129     7.45130690    -5.26471823    -0.83400000
0.00024239      3.15057423
   -14.23529415     7.70012941    -5.97015684     0.41700000
0.00007331      0.40001352
   -12.97290819     8.15358284    -5.38182562     0.41700000
0.00007331      0.40001352
    -0.15509634    -2.72927952     6.89051485    -0.83400000
0.00024239      3.15057423
    -0.82812691    -3.00526023     7.53619385     0.41700000
0.00007331      0.40001352
    -0.73481238    -2.54001141     6.12987900     0.41700000
0.00007331      0.40001352
   -12.08586591     9.46859130    -6.14232176    -0.83400000
0.00024239      3.15057423
   -12.37590473    10.00505051    -6.89231603     0.41700000
0.00007331      0.40001352
   -11.43954294     9.98039016    -5.61479585     0.41700000
0.00007331      0.40001352
   -13.77178452    -1.77177949     3.01155768    -0.83400000
0.00024239      3.15057423
   -13.99905369    -1.46797081     3.87259972     0.41700000
0.00007331      0.40001352
   -13.96769497    -0.96654649     2.47607974     0.41700000
0.00007331      0.40001352
    -9.07558553     3.02668914    -12.63456457    -0.83400000
0.00024239      3.15057423
    -9.34321707     3.15119704    -13.57319062     0.41700000
0.00007331      0.40001352
    -8.43977182     3.74490556    -12.58055227     0.41700000
0.00007331      0.40001352
   -12.64775536     9.15957074    -2.72061105    -0.83400000
0.00024239      3.15057423
   -12.22316954    10.00026898    -2.50208421     0.41700000
0.00007331      0.40001352
   -13.41526100     9.49654822    -3.12602325     0.41700000
0.00007331      0.40001352
   -14.14414046     6.12802414     0.60974650    -0.83400000
0.00024239      3.15057423
```

```
    -13.50532505        5.49088292         0.29932980         0.41700000
0.00007331       0.40001352
    -14.30748865        5.77580695         1.50289587         0.41700000
0.00007331       0.40001352
     -3.58556791        4.65147798       -13.03767273        -0.83400000
0.00024239       3.15057423
     -2.69634958        5.00402955       -12.93018744         0.41700000
0.00007331       0.40001352
     -4.19563766        5.29573004       -12.55840990         0.41700000
0.00007331       0.40001352
    -12.75514815        3.79680629        -0.63400074        -0.83400000
0.00024239       3.15057423
    -12.09488985        4.44156812        -0.37973495         0.41700000
0.00007331       0.40001352
    -12.15647814        3.19973974        -1.14255233         0.41700000
0.00007331       0.40001352
     -2.82846212        5.86562014         4.46032476        -0.83400000
0.00024239       3.15057423
     -1.94455612        5.60928631         4.17307854         0.41700000
0.00007331       0.40001352
     -2.69034266        6.77059746         4.29284573         0.41700000
0.00007331       0.40001352
     -0.74459024       -1.47454067       -13.34422085        -0.83400000
0.00024239       3.15057423
     -0.96498151       -0.51782318       -13.23167774         0.41700000
0.00007331       0.40001352
     -1.36986490       -1.90669199       -12.73271248         0.41700000
0.00007331       0.40001352
    -11.42049568        1.74410542        -6.20227831        -0.83400000
0.00024239       3.15057423
    -11.45701950        2.53933820        -5.56972616         0.41700000
0.00007331       0.40001352
    -11.59960716        2.22951516        -7.01957052         0.41700000
0.00007331       0.40001352
      2.00765657       -5.96039009        -4.22972727        -0.83400000
0.00024239       3.15057423
      1.66777205       -5.67990160        -3.34176111         0.41700000
0.00007331       0.40001352
      2.59278393       -5.23373747        -4.40355873         0.41700000
0.00007331       0.40001352
     -1.80957465       13.41002819        -1.08127956        -0.83400000
0.00024239       3.15057423
     -2.67182962       12.98799917        -1.07766084         0.41700000
0.00007331       0.40001352
     -2.11009364       14.30828163        -0.96596484         0.41700000
0.00007331       0.40001352
      3.06266475       -2.00935292        -8.57498169        -0.83400000
0.00024239       3.15057423
      2.29461861       -2.45595694        -8.22233963         0.41700000
0.00007331       0.40001352
      3.88176322       -2.37163520        -8.21736526         0.41700000
0.00007331       0.40001352
      5.41639662        6.98136330        -7.56164074        -0.83400000
0.00024239       3.15057423
      4.51864910        7.33842087        -7.54427862         0.41700000
```

```
0.00007331       0.40001352
      5.68595123       7.35090637      -8.42472267       0.41700000
0.00007331       0.40001352
    -12.95901034       5.69611732      -3.22328277      -0.83400000
0.00024239       3.15057423
    -13.77139923       5.68432896      -2.69825788       0.41700000
0.00007331       0.40001352
    -13.34648249       6.33859949      -3.87178178       0.41700000
0.00007331       0.40001352
      2.86162949       4.83362484       1.04894245      -0.83400000
0.00024239       3.15057423
      2.94785142       4.43282461       0.15239438       0.41700000
0.00007331       0.40001352
      2.96798277       4.03178787       1.60738480       0.41700000
0.00007331       0.40001352
     -0.37394977       3.92270803       2.80023980      -0.83400000
0.00024239       3.15057423
     -0.68419307       3.90205503       1.88045883       0.41700000
0.00007331       0.40001352
     -0.54673797       3.04112530       3.16742992       0.41700000
0.00007331       0.40001352
     -1.33274293      -1.08798683      -4.14268970      -0.83400000
0.00024239       3.15057423
     -0.58509195      -0.49760401      -4.26913738       0.41700000
0.00007331       0.40001352
     -1.57219243      -0.88222152      -3.26186395       0.41700000
0.00007331       0.40001352
      0.51386356      -4.55409622       4.08526278      -0.83400000
0.00024239       3.15057423
      0.54449826      -5.05407763       4.86631298       0.41700000
0.00007331       0.40001352
      1.17675662      -3.82966375       4.26508665       0.41700000
0.00007331       0.40001352
     -9.03044796       0.01655403      -7.12107897      -0.83400000
0.00024239       3.15057423
     -8.20575523      -0.06905016      -6.59003067       0.41700000
0.00007331       0.40001352
     -9.60659695       0.42270315      -6.47083044       0.41700000
0.00007331       0.40001352
     -5.23969854      -8.46711982      -8.28344024      -0.83400000
0.00024239       3.15057423
     -5.52945674      -9.31833971      -7.88173354       0.41700000
0.00007331       0.40001352
     -4.30740798      -8.55647839      -8.23987974       0.41700000
0.00007331       0.40001352
      1.25795798       1.03646688     -12.03802130      -0.83400000
0.00024239       3.15057423
      0.44199638       0.96703128     -12.53276369       0.41700000
0.00007331       0.40001352
      1.25303463       0.28819088     -11.38923762       0.41700000
0.00007331       0.40001352
     -3.86322170       8.42595885     -10.16992955      -0.83400000
0.00024239       3.15057423
     -4.20128625       9.35292456     -10.00118308       0.41700000
0.00007331       0.40001352
```

```
     -3.28508107        8.33410094        -9.41658883        0.41700000
0.00007331       0.40001352
     -1.39618202        1.11947588       -12.98610136       -0.83400000
0.00024239       3.15057423
     -1.43460079        2.01820723       -13.32191441        0.41700000
0.00007331       0.40001352
     -2.25057884        1.17017154       -12.53048346        0.41700000
0.00007331       0.40001352
    -11.98060472       -4.45980895        -5.87099279       -0.83400000
0.00024239       3.15057423
    -12.82667083       -4.14411701        -6.23088182        0.41700000
0.00007331       0.40001352
    -11.49747241       -4.45754302        -6.66067851        0.41700000
0.00007331       0.40001352
    -15.00323031        2.41336863        -0.69267030       -0.83400000
0.00024239       3.15057423
    -14.16930363        2.92504148        -0.74159809        0.41700000
0.00007331       0.40001352
    -15.41205666        2.73811428        -1.49574371        0.41700000
0.00007331       0.40001352
     -6.21218586        5.15519667        -4.94121981       -0.83400000
0.00024239       3.15057423
     -6.72395086        4.43384933        -5.30777979        0.41700000
0.00007331       0.40001352
     -5.42491817        4.68438816        -4.57948351        0.41700000
0.00007331       0.40001352
     -2.29626624       11.47527525        -9.14929299       -0.83400000
0.00024239       3.15057423
     -2.29822628       12.39635966        -8.84270148        0.41700000
0.00007331       0.40001352
     -3.22399269       11.25381300        -9.08112816        0.41700000
0.00007331       0.40001352
     -5.04480140       11.05734179        -9.18205981       -0.83400000
0.00024239       3.15057423
     -5.45388369       11.94699976        -9.04231505        0.41700000
0.00007331       0.40001352
     -5.47470669       10.61978934        -8.38914208        0.41700000
0.00007331       0.40001352
    -12.99631064        4.39289984       -11.04582184       -0.83400000
0.00024239       3.15057423
    -12.52935117        5.03446301       -11.59578340        0.41700000
0.00007331       0.40001352
    -12.39318173        3.59869632       -11.06841629        0.41700000
0.00007331       0.40001352
    -14.18057526       -3.40689767        -7.67142881       -0.83400000
0.00024239       3.15057423
    -14.31170643       -3.11911500        -8.59677041        0.41700000
0.00007331       0.40001352
    -14.29993022       -2.60616792        -7.08094276        0.41700000
0.00007331       0.40001352
      1.48241453       -3.17238231       -13.02713272       -0.83400000
0.00024239       3.15057423
      0.96281199       -2.33442588       -12.99476597        0.41700000
0.00007331       0.40001352
      0.82027153       -3.64223047       -13.57238743        0.41700000
```

```
0.00007331      0.40001352
$end


$lj_parameters
  1      0.00003187      3.91995436
  2      0.00019123      3.02905564
  3      0.00009562      3.59923082
  4      0.00005578      2.38760856
  5      0.00005578      2.38760856
  6      0.00009562      3.59923082
  7      0.00005578      2.38760856
  8      0.00005578      2.38760856
  9      0.00003187      3.91995436
 10      0.00019123      3.02905564
 11      0.00009562      3.59923082
 12      0.00005578      2.38760856
 13      0.00005578      2.38760856
 14      0.00007331      0.40001352
 15      0.00030613      3.14487248
 16      0.00054661      3.40323310
 17      0.00024239      3.15057423
 18      0.00007331      0.40001352
 19      0.00007331      0.40001352
$end


$comment
qchem control file needed for the
scan.inp
$end


$rem
igdesp               699
symmetry              off
sym_ignore            true
exchange             b3lyp
basis                6-31+g**
qm_mm                true
jobtype              force
symmetry             off
sym_ignore           true
print_input          false
qmmm_print           true
scf_algorithm        diis_gdm
scf_guess_always     true
thresh               14
scf_convergence      9
max_scf_cycles       200
$end


$molecule
0 1
 C        -4.5302529979      0.3474980463     -3.4482620388
 O        -4.2821842065     -0.6443653521     -4.1159852232
 C        -4.2589833280      0.4728237024     -1.9548600563
 H        -3.1853144153      0.4014444707     -1.7546327052
```

```
 H          -4.7318365026      -0.3880481106      -1.4758713928
 C          -4.8803182970       1.8151120642      -1.5142897172
 H          -4.1078092431       2.5797624642      -1.3630062200
 H          -5.4671954870       1.7724772590      -0.5921306735
 C          -5.7213707362       2.2742501489      -2.6949769955
 O          -6.6676775358       3.0452749043      -2.6486124046
 C          -5.2295935503       1.5944700901      -3.9544916891
 H          -4.4735430443       2.2400199363      -4.4323231722
 H          -5.9951562143       1.4005382003      -4.7094519761
 H          -7.6213871300       3.9841879802      -1.5545514805
 O          -8.1913255525       4.6144490141      -1.0423151450
CL          -7.3445458593       6.1039254073      -1.2278823717
 O          -5.9399657407      -0.4077018344      -6.4716743463
 H          -5.1471612627      -0.5542510367      -5.9306930634
 H          -6.6122922685      -0.9632560473      -6.0326838892
$end
```

Table 49: Script generated by CHARMM to be run by QChem for each replica in Replica Path method and CI-NEB method

Appendix B: Topology and parameter files

B1   Topology and parameter files for sugar moleclues.

```
RESI AXYS           0.000    ! alpha-D-xylopyranos monomer
                             ! can be used for beta L arabinose(ARB)
GROU
ATOM C1   CC3162     0.340
ATOM H1   HCA1       0.090
ATOM O1   OC311     -0.650
ATOM HO1  HCP1       0.420
ATOM C5   CC3263     0.020
ATOM H51  HCA2       0.090
ATOM H52  HCA2       0.090
ATOM O5   OC3C61    -0.400
GROU
ATOM C2   CC3161     0.140
ATOM H2   HCA1       0.090
ATOM O2   OC311     -0.650
ATOM HO2  HCP1       0.420
GROU
ATOM C3   CC3161     0.140
ATOM H3   HCA1       0.090
ATOM O3   OC311     -0.650
ATOM HO3  HCP1       0.420
GROU
ATOM C4   CC3161     0.140
ATOM H4   HCA1       0.090
ATOM O4   OC311     -0.650
ATOM HO4  HCP1       0.420

BOND C1   O1      C1   H1      O1   HO1     C1   O5      C1   C2
BOND C2   H2      C2   O2      O2   HO2     C2   C3      C3   H3
BOND C3   O3      O3   HO3     C3   C4      C4   H4      C4   O4
BOND O4   HO4     C4   C5      C5   H51     C5   H52
BOND C5   O5
DONO BLNK HO1
DONO BLNK HO2
DONO BLNK HO3
DONO BLNK HO4
ACCE O1
ACCE O2
ACCE O3
ACCE O4
ACCE O5
!    I    J    K    L      R(IK)    T(IKJ)    PHI     T(JKL)   R(KL)
IC   O1   C2   *C1  H1     1.4115  105.82 -120.26  110.81   1.0905
IC   O1   O5   *C1  C2     1.4115  113.00 -118.92  111.32   1.5218
IC   O2   C3   *C2  H2     1.4190  110.99 -116.97  108.77   1.0892
```

```
IC   O2   C1   *C2  C3       1.4190  111.74 -123.77  110.04  1.5167
IC   O3   C4   *C3  H3       1.4198  107.51  119.38  108.28  1.0957
IC   O3   C2   *C3  C4       1.4198  111.19  119.03  110.08  1.5102
IC   O4   C5   *C4  H4       1.4163  107.94 -120.61  109.07  1.0972
IC   O4   C3   *C4  C5       1.4163  111.43 -118.91  108.83  1.5171
IC   O5   C1   C2   C3       1.4059  111.32   53.56  110.04  1.5167
IC   C1   C2   C3   C4       1.5218  110.04  -54.55  110.08  1.5102
IC   C2   C3   C4   C5       1.5167  110.08   57.15  108.83  1.5171
IC   C3   C4   C5   O5       1.5102  108.83  -58.25  109.77  1.4384
IC   C4   C5   O5   C1       1.5171  109.77   59.83  114.27  1.4059
IC   C5   O5   C1   C2       1.4384  114.27  -57.30  111.32  1.5218
IC   O5   C1   O1   HO1      1.4059  113.00   67.18  108.04  0.9634
IC   C1   C2   O2   HO2      1.5218  111.74  -44.11  105.37  0.9665
IC   C2   C3   O3   HO3      1.5167  111.19  -51.80  106.26  0.9641
IC   C3   C4   O4   HO4      1.5102  111.43   47.58  105.89  0.9645
PATC FIRS NONE LAST NONE


RESI BXYS           0.000  ! beta-D-xylopyranos monomer


GROU
ATOM C1   CC3162    0.340
ATOM H1   HCA1      0.090
ATOM O1   OC311    -0.650
ATOM HO1  HCP1      0.420
ATOM C5   CC3163    0.020
ATOM H51  HCA2      0.090
ATOM H52  HCA2      0.090
ATOM O5   OC3C61   -0.400
GROU
ATOM C2   CC3161    0.140
ATOM H2   HCA1      0.090
ATOM O2   OC311    -0.650
ATOM HO2  HCP1      0.420
GROU
ATOM C3   CC3161    0.140
ATOM H3   HCA1      0.090
ATOM O3   OC311    -0.650
ATOM HO3  HCP1      0.420
GROU
ATOM C4   CC3161    0.140
ATOM H4   HCA1      0.090
ATOM O4   OC311    -0.650
ATOM HO4  HCP1      0.420
GROU


BOND C1   O1       C1   H1       O1   HO1      C1   O5       C1   C2
BOND C2   H2       C2   O2       O2   HO2      C2   C3       C3   H3
BOND C3   O3       O3   HO3      C3   C4       C4   H4       C4   O4
BOND O4   HO4      C4   C5       C5   H51      C5   H52
BOND C5   O5
DONO BLNK HO1
DONO BLNK HO2
DONO BLNK HO3
DONO BLNK HO4
ACCE O1
```

```
ACCE O2
ACCE O3
ACCE O4
ACCE O5
!      I     J     K     L      R(IK)   T(IKJ)    PHI    T(JKL)   R(KL)
IC     O1    C2    *C1   H1     1.3890  105.75  114.54  108.17  1.0950
IC     O1    O5    *C1   C2     1.3890  111.55  117.06  110.06  1.5340
IC     O2    C3    *C2   H2     1.4154  112.27 -118.21  108.23  1.0919
IC     O2    C1    *C2   C3     1.4154  110.87 -125.56  111.08  1.5253
IC     O3    C4    *C3   H3     1.4157  110.61  120.65  108.81  1.1068
IC     O3    C2    *C3   C4     1.4157  108.09  120.77  109.86  1.5177
IC     O4    C5    *C4   H4     1.4252  110.90 -120.61  108.35  1.1024
IC     O4    C3    *C4   C5     1.4252  108.31 -122.08  111.17  1.5287
IC     O5    C1    C2    C3     1.4254  110.06   54.09  111.08  1.5253
IC     C1    C2    C3    C4     1.5340  111.08  -51.23  109.86  1.5177
IC     C2    C3    C4    C5     1.5253  109.86   53.25  111.17  1.5288
IC     C3    C4    C5    O5     1.5177  111.17  -57.46  108.69  1.4274
IC     C4    C5    O5    C1     1.5288  108.69   62.25  113.77  1.4254
IC     C5    O5    C1    C2     1.4274  113.77  -60.97  110.06  1.5340
IC     C1    C2    O2    HO2    1.5340  110.87 -100.51  112.13  0.9638
IC     C2    C3    O3    HO3    1.5253  108.09 -165.88  112.08  0.9730
IC     C3    C4    O4    HO4    1.5177  108.31  134.18  106.97  0.9713


PATC  FIRS NONE LAST NONE

RESI BNAG            0.000  ! N-Acytyl-Glucosamine

GROU
ATOM C1    CC3162     0.340
ATOM H1    HCA1       0.090
ATOM O1    OC311     -0.650
ATOM HO1   HCP1       0.420
ATOM C5    CC3163     0.110
ATOM H5    HCA1       0.090
ATOM O5    OC3C61    -0.400
GROU
ATOM C2    CC3161     0.070
ATOM H2    HCA1       0.090
ATOM NL    NH1       -0.47
ATOM HL    H          0.31
GROUP
ATOM CL    CT3       -0.27
ATOM HL1   HA         0.09
ATOM HL2   HA         0.09
ATOM HL3   HA         0.09
GROUP
ATOM CLP   C          0.51
ATOM OL    O         -0.51
GROUP
GROU
ATOM C3    CC3161     0.140
ATOM H3    HCA1       0.090
ATOM O3    OC311     -0.650
ATOM HO3   HCP1       0.420
GROU
```

```
ATOM C4     CC3161     0.140
ATOM H4     HCA1       0.090
ATOM O4     OC311     -0.650
ATOM HO4    HCP1       0.420
GROU
ATOM C6     CC321      0.050
ATOM H61    HCA2       0.090
ATOM H62    HCA2       0.090
ATOM O6     OC311     -0.650
ATOM HO6    HCP1       0.420


BOND C1    O1      C1    H1      O1    HO1     C1    O5      C1    C2
BOND C2    H2      C2    NL      NL    HL      NL    CLP     C2    C3
BOND C3    O3      O3    HO3     C3    C4      C4    H4      C4    O4
BOND O4    HO4     C4    C5      C5    H5      C5    C6      C6    H61
BOND C6    H62     C6    O6      O6    HO6     C5    O5      C3    H3
BOND CL    HL1     CL    HL2     CL    HL3     CL    CLP
DOUBLE  CLP OL
IMPR CLP CL NL OL
DONO BLNK HO1
DONO BLNK HO3
DONO BLNK HO4
DONO BLNK HO6
ACCE O1
ACCE O3
ACCE O4
ACCE O5
ACCE O6
!      I      J      K      L      R(IK)     T(IKJ)     PHI     T(JKL)    R(KL)
IC     O1     C2     *C1    H1     1.3890    105.75    114.54   108.17   1.0950
IC     O1     O5     *C1    C2     1.3890    111.55    117.06   110.06   1.5340
IC     O3     C4     *C3    H3     1.4157    110.61    120.     108.81   1.1068
IC     O3     C2     *C3    C4     1.4157    108.09    120.     109.86   1.5177
IC     O4     C5     *C4    H4     1.4252    110.90   -120.     108.35   1.1024
IC     O4     C3     *C4    C5     1.4252    108.31   -120.     111.17   1.5287
IC     C6     O5     *C5    H5     1.5099    108.10    118.69   109.65   1.1042
IC     C6     C4     *C5    O5     1.5099    111.57    119.10   108.69   1.4274
IC     O6     H62    *C6    H61    1.4132    110.47   -120.32   107.85   1.0945
IC     O6     C5     *C6    H62    1.4132    110.45   -121.53   108.99   1.0959
IC     O5     C1     C2     C3     1.4254    110.06     54.09   111.08   1.5253
IC     C1     C2     C3     C4     1.5340    111.08    -51.23   109.86   1.5177
IC     C2     C3     C4     C5     1.5253    109.86     53.25   111.17   1.5288
IC     C3     C4     C5     O5     1.5177    111.17    -57.46   108.69   1.4274
IC     C4     C5     O5     C1     1.5288    108.69     62.25   113.77   1.4254
IC     C5     O5     C1     C2     1.4274    113.77    -60.97   110.06   1.5340
IC     C4     C5     C6     O6     1.5287    111.57    180.     110.45   1.4132
IC     O5     C1     O1     HO1    1.4254    111.55    -60.     107.83   0.9684
IC     C1     C2     O2     HO2    1.5340    110.87    -60.     112.13   0.9638
IC     C2     C3     O3     HO3    1.5253    108.09     60.     112.08   0.9730
IC     C3     C4     O4     HO4    1.5177    108.31     60.     106.97   0.9713
IC     C5     C6     O6     HO6    1.5099    110.44    -60.     107.72   0.9641
ic     c2     clp    *nl    hl     0.0 0.0   180.0    0.0 0.0
ic     nl     cl     *clp   ol     0.0 0.0   180.0    0.0 0.0
ic     cl     clp    nl     c2     0.0 0.0   180.0    0.0 0.0
ic     ol     clp    nl     c2     0.0 0.0     0.0    0.0 0.0
```

```
ic hl1 cl  clp nl   0.0 0.0  180.0   0.0 0.0
ic hl2 cl  clp nl   0.0 0.0   60.0   0.0 0.0
ic hl3 cl  clp ol   0.0 0.0  120.0   0.0 0.0
ic c2  clp *nl hl   0.0 0.0  180.0   0.0 0.0
ic hl1 hl2 *cl hl3  0.0 0.0  240.0   0.0 0.0
ic hl  nl  c2  h2   0.0 0.0    0.0   0.0 0.0


PATC   FIRS NONE LAST NONE

RESI PCA           0.00  ! first resi in CPO (Pyroglutamic acid)

GROUP
ATOM CB   CT2    -0.18
ATOM HB1  HA      0.09
ATOM HB2  HA      0.09
GROUP
ATOM CG   CT2    -0.28
ATOM HG1  HA      0.09
ATOM HG2  HA      0.09
ATOM CD   CC      0.72
ATOM OE   OC     -0.69
ATOM N    NH1    -0.40
ATOM HN   H       0.31
ATOM CA   CT1     0.07
ATOM HA   HB      0.09
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51


BOND CD N
BOND CB CA   CG CB   CD CG
BOND N   HN   N  CA C    CA
BOND C   +N   CA HA  CB HB1 CB   HB2 CG   HG1
BOND CG HG2
DOUBLE O   C    CD   OE
IMPR N    -C CA   HN   C CA +N O
DONOR HN N
ACCEPTOR OE   CD
ACCEPTOR O C
IC -C    CA   *N    HN    1.3471 124.4500  180.0000 113.9900  0.9961
IC -C    N    CA    C     1.3471 124.4500  180.0000 107.2700  1.5216
IC N     CA   C     +N    1.4512 107.2700  180.0000 117.2500  1.3501
IC +N    CA   *C    O     1.3501 117.2500  180.0000 121.0700  1.2306
IC CA    C    +N    +CA   1.5216 117.2500  180.0000 124.3000  1.4530
IC N     C    *CA   CB    1.4512 107.2700  121.9000 111.7100  1.5516
IC N     C    *CA   HA    1.4512 107.2700 -118.0600 107.2600  1.0828
IC N     CA   CB    CG    1.4512 111.0400  180.0000 115.6900  1.5557
IC CG    CA   *CB   HB1   1.5557 115.6900  121.2200 108.1600  1.1145
IC CG    CA   *CB   HB2   1.5557 115.6900 -123.6500 109.8100  1.1131
IC CA    CB   CG    CD    1.5516 115.6900  180.0000 115.7300  1.5307
IC CD    CB   *CG   HG1   1.5307 115.7300  117.3800 109.5000  1.1053
IC CD    CB   *CG   HG2   1.5307 115.7300 -121.9600 111.0000  1.1081
```

Table 50: Topology file for sugars, for n-acytyl-glucosamine and for pyroglutamic acid, which are found in CPO, and for which parameters were created by analogy

```
PRES SSO4          0.000    ! patch between MAN-MAN( C1-O4 )

DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO4
BOND 1C1 2O4
GROU
ATOM 1C1    CC3162    0.240
ATOM 1H1    HCA1      0.090
ATOM 1C5    CC3163    0.080
ATOM 1H5    HCA1      0.090
ATOM 1O5    OC3C61   -0.500
GROU
ATOM 2C4    CC3161    0.360
ATOM 2H4    HCA1      0.090
ATOM 2O4    OC311    -0.450

PRES XSO4          0.000    ! patch between xys-sugar( C1-O4 )

DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO4
BOND 1C1 2O4
GROU
ATOM C1     CC3162    0.210
ATOM H1     HCA1      0.090
ATOM C5     CC3263    0.020
ATOM H51    HCA2      0.090
ATOM H52    HCA2      0.090
ATOM O5     OC3C61   -0.500
GROU
ATOM 2C4    CC3161    0.360
ATOM 2H4    HCA1      0.090
ATOM 2O4    OC311    -0.450


PRES SSO3          0.000    ! patch between MAN-MAN( C1-O3 )

DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO3
BOND 1C1 2O3
GROU
ATOM 1C1    CC3162    0.240
ATOM 1H1    HCA1      0.090
ATOM 1C5    CC3163    0.080
ATOM 1H5    HCA1      0.090
ATOM 1O5    OC3C61   -0.500
GROU
ATOM 2C4    CC3161    0.360
ATOM 2H4    HCA1      0.090
ATOM 2O4    OC311    -0.450

PRES XSO3          0.000    ! patch between XYS-MAN( C1-O3 )
```

250

```
DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO3
BOND 1C1 2O3
GROU
ATOM C1    CC3162    0.210
ATOM H1    HCA1      0.090
ATOM C5    CC3263    0.020
ATOM H51   HCA2      0.090
ATOM H52   HCA2      0.090
ATOM O5    OC3C61   -0.500
GROU
ATOM 2C4   CC3161    0.360
ATOM 2H4   HCA1      0.090
ATOM 2O4   OC311    -0.450


PRES SSO6          0.000    ! patch between MAN-MAN( C1-O6 )

DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO6
BOND 1C1 2O6
GROU
ATOM 1C1   CC3162    0.240
ATOM 1H1   HCA1      0.090
ATOM 1C5   CC3163    0.080
ATOM 1H5   HCA1      0.090
ATOM 1O5   OC3C61   -0.500
GROU
ATOM 2C6   CC321     0.250
ATOM 2H61  HCA2      0.090
ATOM 2H62  HCA2      0.090
ATOM 2O6   OC311    -0.430

PRES XSO6          0.000    ! patch between XYS-MAN( C1-O6 )

DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HO6
BOND 1C1 2O6
GROU
ATOM C1    CC3162    0.210
ATOM H1    HCA1      0.090
ATOM C5    CC3263    0.020
ATOM H51   HCA2      0.090
ATOM H52   HCA2      0.090
ATOM O5    OC3C61   -0.500
GROU
ATOM 2C6   CC321     0.250
ATOM 2H61  HCA2      0.090
ATOM 2H62  HCA2      0.090
ATOM 2O6   OC311    -0.430

PRES SASN          0.000    ! patch between MAN-ASN( C1-ND2 )
```

```
DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HD22
BOND 1C1 2ND2
GROU
ATOM C1    CC3162    0.240
ATOM H1    HCA1      0.090
ATOM C5    CC3163    0.080
ATOM H5    HCA1      0.090
ATOM O5    OC3C61   -0.500
GROUP
ATOM 2ND2  NH2      -0.32
ATOM 2HD21 H         0.32


PRES STHR          0.000 ! patch between MAN-THR( C1-OG1 )


DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HG1
BOND 1C1 2OG1
GROU
ATOM C1    CC3162    0.240
ATOM H1    HCA1      0.090
ATOM C5    CC3163    0.080
ATOM H5    HCA1      0.090
ATOM O5    OC3C61   -0.500
GROUP
ATOM 2CB   CT1       0.17
ATOM 2HB   HA        0.09
ATOM 2OG1  OH1      -0.26


PRES SSER          0.000 ! patch between MAN-SER( C1-OG )


DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HG1
BOND 1C1 2OG
GROU
ATOM C1    CC3162    0.240
ATOM H1    HCA1      0.090
ATOM C5    CC3163    0.080
ATOM H5    HCA1      0.090
ATOM O5    OC3C61   -0.500
GROUP
ATOM 2CB   CT2       0.08
ATOM 2HB1  HA        0.09
ATOM 2HB2  HA        0.09
ATOM 2OG   OH1      -0.26


PRES XSER          0.000  ! patch between AXYS-SER( C1-OG )


DELETE ATOM 1O1
DELETE ATOM 1HO1
DELETE ATOM 2HG1
```

```
BOND 1C1 2OG
GROU
ATOM C1    CC3162    0.210
ATOM H1    HCA1      0.090
ATOM C5    CC3263    0.020
ATOM H51   HCA2      0.090
ATOM H52   HCA2      0.090
ATOM O5    OC3C61   -0.500
GROUP
ATOM 2CB   CT2       0.08
ATOM 2HB1  HA        0.09
ATOM 2HB2  HA        0.09
ATOM 2OG   OH1      -0.26


PRES PCYS          0.00 ! Patch for HEME to CYS link
                        ! Patch residues must be 1-CYS,2-HEME


DELETE ANGLE 2NA 2FE 2NC  2NB 2FE 2ND
DELETE ATOM 1HG1


GROUP
ATOM CB    CT2      -0.05
ATOM HB1   HA        0.09
ATOM HB2   HA        0.09
ATOM SG    S        -0.13


BOND 1SG 2FE
ANGLE 1CB 1SG 2FE   1SG 2FE 2NA  1SG 2FE 2NB
ANGLE 1SG 2FE  2NC  1SG 2FE  2ND
DIHE 1CB  1SG  2FE  2NA

IC 1CB 1SG 2FE  2NA   1.8180  100.60    0.0000  90.000  2.0000
IC 1CB 1SG 2FE  2NB   1.8180  100.60    90.000  90.000  2.0000
IC 1CB 1SG 2FE  2NC   1.8180  100.60    180.00  90.000  2.0000
IC 1CB 1SG 2FE  2ND   1.8180  100.60   -90.000  90.000  2.0000
IC 1CA 1CB 1SG  2FE   1.5380  112.50    0.2000  100.60  2.4500
```

Table 51: Patches used to attach sugar-sugar molecules, sugar-amino acids and heme-cystine


```
* $Id: par_allxx_sugar.inp,v 1.40 2008/07/08 18:06:46 ganesh Exp $
*>>>>>>>>>>>> All-hydrogen parameters used in the <<<<<<<<<<<<<<<
*>>>>> development of the CHARMM carbohydrate  force field<<<<<<<<
*>>>>>>>>>>>>>>>>>>>>>>>> February 2006 <<<<<<<<<<<<<<<<<<<<<<<<<<<
*>>>>>>>> Direct comments to Alexander D. MacKerell Jr. <<<<<<<<<<
*>>>>>>>> 410-706-7442 or email:
* Initial sugar parameter file
*


! par22: par_all22_prot.inp
! alex:  Alex MacKerell
! sng:   Shannon Greene
```

```
! og:    Olgun Guvench
! viv:   Igor Vorobyov
! erh:   Elizabeth Hatcher
!

BONDS
!V(bond) = Kb(b - b0)**2
!Kb: kcal/mole/A**2
!b0: A
!
! atom types     Kb      b0
!
! monosaccharide CO
CC3162  OH1    428.00  1.400

! monosaccharide CN
CC3161  NC2D1    320.00  1.430
CC3161  NH1     320.00  1.430
CC3162  NH2     320.00  1.430

!***************** ANGLES ************
ANGLES
!V(angle) = Ktheta(Theta - Theta0)**2
!V(Urey-Bradley) = Kub(S - S0)**2
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
! atom types              Ktheta  Theta0   Kub    S0
!
!
! monosaccharide CCN
CC3161   CC3161   NH1  70.000  113.50
CC3161   CC3162   NH2  70.000  113.50
CC3162   CC3161   NH1  70.000  113.50
!
! monosaccharide CNC
CC3161   NH1    C  50.000  120.00

! monosaccharide NCO
OC3C61   CC3162   NH2    110.0   112.0
!
! monosaccharide NCH
NH1    CC3161   HCA1    48.000  108.00
NH2    CC3161   HCA1    48.000  108.00
NH2    CC3162   HCA1    48.000  108.00
!
! monosaccharide HNC
CC3161   NH1    H  35.000  117.00
!
! monosaccharide CCO
CC3161  CC3162  OH1     75.700  107.00

! monosaccharide OCO
```

```
OH1      CC3162   OC3C61   45.000  116.50
!
! monosaccharide COC
CC3162  OH1      CT1      95.00   109.70
!
! monosaccharide HCO
HCA1     CC3162   OH1      55.000  108.89


! ******** V(dihedral) = Kchi(1 + cos(n(chi) - delta))
DIHEDRALS

! monosaccharide CCCO
!  C3-C2-C1(anomeric)-O5
CC3161  CC3161  CC3162  OH1   0.31  3  180.0

! monosaccharide CCCN
NH1     CC3161  CC3162  NH2   0.20  3    0.0
NH1     CC3161  CC3161  CC3161  0.20  3    0.0
NH2     CC3162  CC3161  CC3161  0.20  3    0.0

! monosaccharide CCOC
!  C2-C1-O5-C5
CC3161  CC3162  OH1   CT1   0.20  3    0.0
CC3161  CC3162  OH1   CT2   0.20  3    0.0
CC3163  CC3162  OC311  CC3162  0.20  3    0.0
CC3263  CC3161  OC311  CC3162  0.20  3    0.0
CC3161  CC3161  OC311  CC3162  0.20  3    0.0
CC3161  CC3162  OC311  CC3161  0.20  3    0.0

! C1-O-C4-C5/C3
CC3162  OC311  OC3C61  CC3163  0.20  3    0.0
CC3162  OC311  OC3C61  CC3161  0.20  3    0.0

! monosaccharide OCCO
!  (H)O-C(anomeric)-C-O(H)
OH1   CC3162  CC3161  OC311   2.65  1  180.0
OH1   CC3162  CC3161  OC311   0.00  2    0.0
OH1   CC3162  CC3161  OC311   0.13  3  180.0

! monosaccharide HCCO
HCA1     CC3161  CC3162  OH1   0.14  3    0.0

! monosaccharide HCCN
HCA1     CC3162  CC3161  NH1   0.20  3    0.0
HCA1     CC3161  CC3162  NH2   0.20  3    0.0
HCA1     CC3161  CC3161  NH1   0.20  3    0.0

! monosaccharide OCCN
OC311    CC3161  CC3161  NH1   0.20  3    0.0
OC3C61  CC3162  CC3161  NH1   0.20  3    0.0
OC311  CC3162  CC3161  NH1   0.20  3    0.0

! monosaccharide OCNC
OC3C61  CC3162  NH2  CC     0.0     3        0.0
OC3C61  CC3162  NH2  H      0.0     3        0.0
```

```
CC3161  NH1   C    O     2.50  2  180.0
CC3162  NH2   CC   O     2.50  2  180.0


! monosaccharide CNCC
CC3161  NH1   C    CT3    1.60  1    0.0
CC3162  NH2   CC   CT2    1.60  1    0.0
CC    NH2   CC3162  CC3161   0.50  1  180.0
C   NH1   CC3161  CC3162   0.50  1  180.0
C   NH1   CC3161  CC3161   0.50  1  180.0


! monosaccharide HCOC
HCA1    CC3162  OH1  CT1   0.284 3    0.0
HCA1    CC3162  OH1  CT2   0.284 3    0.0
HCA1    CC3161  OC311  CC3162   0.284 3    0.0
HCA1    CC3162  OC311  CC3161   0.284 3    0.0


! monosaccharide HNCC
CC3162  CC3161  NH1   H    0.00  1    0.0
CC3161  CC3162  NH2   H    0.00  1    0.0
H    NH1   CC3161  CC3161   0.00  1    0.0


! monosaccharide HNCH
HCA1    CC3161  NH1   H    0.00  1    0.0
HCA1    CC3162  NH2   H    0.00  1    0.0


! monosaccharide CNCH
HCA1    CC3161  NH1   C    0.00  1    0.0
HCA1    CC3162  NH2   CC   0.00  1    0.0


! OCOC
OC3C61  CC3162  OH1  CT2   0.2  3    0.0
OC3C61  CC3162  OC311  CC3161  0.2  3    0.0
OH1   CC3162  OC3C61  CC3163    0.41  1  180.0
OH1   CC3162  OC3C61  CC3163    0.89  2    0.0
OH1   CC3162  OC3C61  CC3163    0.05  3    0.0


!*************** CPO - HEME Parameter entry*************
NH1  CC    370.000     1.3450
FE   NPH   270.2000    2.0000
FE   S     300.0000    2.4500


CT2  S    FE    50.0       100.6
S    FE   NPH   100.0        90.0
CT1  NH1  CC    50.000   120.0000
H    NH1  CC    34.000   123.0000
NH1  CC   CT2   80.000   116.5000
OC   CC   NH1   80.000   122.5000


NPH  FE   S    CT2      0.0400  4     0.00
C    CT1  NH1  CC       0.2000  1   180.00
CT2  CC   NH1  CT1      1.6000  1     0.00
CT2  CC   NH1  CT1      2.5000  2   180.00
CT2  CT1  NH1  CC       1.8000  1     0.00
H    NH1  CC   CT2      2.5000  2   180.00
```

```
HB    CT1  NH1  CC         0.0000  1     0.00
OC    CC   NH1  CT1        2.5000  2   180.00
OC    CC   NH1  H          2.5000  2   180.00
```

Table 52: Lines appended by Cassian to CHARMM parameter file in order to include parameters developed for entries in topology file in Appendix B1

VITA

CASSIAN D'CUNHA

1995 – 1999	Bachelor of Engineering: Chemical Engineering University of Mumbai Mumbai, India
2000 – 2003	Master of Science: Computer Science Florida International University Miami, Florida, USA
2004	Cyberbridges Scholarship Florida International University Miami, Florida, USA
2005 – 2010	Teaching Assistant Florida International University Miami, Florida, USA
2011	Guest lecturer: Physical Chemistry II Florida International University Miami, Florida, USA
	Dissertation Year Fellowship Florida International University Miami, Florida, USA

PUBLICATIONS AND PRESENTATIONS

D. C. Chatfield, A. Augsten, C. D'Cunha, and S. E. Wong, "Methyl dynamics in crystalline amino acids: MD and NMR," J. Comput. Chem. **24**, 1052-1058, 2003.

C. D'Cunha, D. C. Chatfield, A. Augsten, "Correlation times and adiabatic barriers for methyl rotations in proteins," FAME, Orlando, Summer 2004.

D. C. Chatfield, A. Augsten, and C. D'Cunha, "Correlation times and adiabatic barriers for methyl rotations in proteins," J. Biomol. NMR. 1-9, 2004.

C. D'Cunha, A. Gairola, A. Augsten, and D. C. Chatfield, "MD/NMR Characterization of internal motions of peptide toxins," American Chemical Society meeting, Philadelphia, Fall 2004.

D. C. Chatfield, A. Augsten, C. D'Cunha, E. Lewandowska, and S. F. Wnuk, "Theoretical and Experimental Study of Regioselectivity of Michael Additions," Eur. J. Org. Chem. 313-322, 2004.

C. D'Cunha, D. Chatfield, H. L. Alvarez, D. A. Cox, E. Crumpler, R. Gutierrez, J. Ibarra, E. Johnson, K. Kumar, T. Milledge, G. Narasimhan, S. M. Sadjadi, C. Zhang, "CyberBridges A Model Collaboration Infrastructure for e-Science," SuperComputing 2006, Tampa, Fall 2006 (oral presentation).

H. L. Alvarez, D. Chatfield, D. A. Cox, E. Crumpler, C. D'Cunha, R. Gutierrez, J. Ibarra, E. Johnson, K. Kumar, T. Milledge, G. Narasimhan, S. M. Sadjadi, C. Zhang, "CyberBridges A Model Collaboration Infrastructure for e-Science," Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'07), Rio de Janeiro, Brazil, May 2007.

C. D'Cunha and D. C. Chatfield, "Study of the halogenation mechanism catalyzed by Chloroperoxidase," American Chemical Society meeting, Philadelphia, Fall 2008.

C. D'Cunha and D. C. Chatfield, "Study of the halogenation mechanism catalyzed by Chloroperoxidase," 3rd CMM 'Users Meet Developers' Workshop on QM/MM Simulations, Philadelphia, Fall 2008.

A. N. Morozov, C. D'Cunha, C. A. Alvarez, and D. C. Chatfield "Enantiospecificity of Chloroperoxidase-Catalyzed Epoxidation: Biased Molecular Dynamics Study of a Cis-β-Methylstyrene/Chloroperoxidase-Compound I Complex," Biophys. J. **100(4)**, 1066-1075, 2011.