INVESTIGATION OF BIOLOGICALLY IMPORTANT SMALL MOLECULES:
QUANTUM CHEMICAL AND MOLECULAR DYNAMICS CALCULATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

E. DENİZ TEKİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
PHYSICS

AUGUST 2010

Approval of the thesis:

INVESTIGATION OF BIOLOGICALLY IMPORTANT SMALL MOLECULES:

QUANTUM CHEMICAL AND MOLECULAR DYNAMICS CALCULATIONS

submitted by **E. DENİZ TEKİN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Physics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**   ———————

Prof. Dr. Sinan Bilikmen
Head of Department, **Physics**   ———————

Prof. Dr. Şakir Erkoç
Supervisor, **Physics, METU**   ———————

Examining Committee Members:

Prof. Dr. Ahmet Palazoğlu
Chemical Engineering and Materials Science, UC-Davis   ———————

Prof. Dr. Şakir Erkoç
Physics Department, METU   ———————

Prof. Dr. Roland Faller
Chemical Engineering and Materials Science, UC-Davis   ———————

Prof. Dr. Nesrin Hasırcı
Chemistry Department, METU   ———————

Prof. Dr. Sibel Başkal
Physics Department, METU   ———————

Date: ———————

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: E. DENİZ TEKİN

Signature   :

ABSTRACT

INVESTIGATION OF BIOLOGICALLY IMPORTANT SMALL MOLECULES:
QUANTUM CHEMICAL AND MOLECULAR DYNAMICS CALCULATIONS

Tekin, E. Deniz

Ph.D, Department of Physics

Supervisor : Prof. Dr. Şakir Erkoç

August 2010, 126 pages

In this thesis, six small molecules (S-allylcysteine, S-allyl mercaptocysteine, allicin, methyl propyl disulfide, allyl methyl sulfide and dipropylsulfide) that are found in garlic and onion, and are known to be beneficial for human health were studied using molecular mechanics, semi-empirical methods, ab-initio (Restricted Hartree Fock), and density functional theory. Using the same methods, a synthetic pyrethroid pesticide molecule, called cyfluthrin, was also studied. Structural, vibrational and electronic properties of these molecules were found. These theoretical studies could clarify the role of these molecules on human health before they are commercially developed and used. In addition, unfolding dynamics of small peptide sequences (DDATKTFT and its variants) in immunoglobulin G-binding protein G was investigated. Protein folding and unfolding is one of the most important unsolved problems in molecular biology. Because of the large number of atoms involved in protein folding, it is a massive computational problem. The hope is that, one could understand this mechanism with the help of molecular dynamics simulation on small peptides. One of our findings is that the location of the hydrogen bonds is important for the stability of the peptide.

ÖZ

BİYOLOJİK ÖNEME SAHİP KÜÇÜK MOLEKÜLLERİN KUANTUM KİMYASAL VE MOLEKÜLER DİNAMİK HESAPLAMALARLA İNCELENMESİ

Tekin, E. Deniz

Doktora, Fizik Bölümü

Tez Yöneticisi   : Şakir Erkoç

Ağustos 2010, 126 sayfa

Bu tezde moleküler mekanik, yarı-ampirik yöntemler, ab-initio (Sınırlandırılmış Hartree Fock) ve yoğunluk fonksiyoneli kuramı yöntemleri kullanılarak, insan sağlığına faydalı olduğu bilinen, sarımsak ve soğanda bulunan moleküllerin küçük bir kısmı olan, S-alilsistein, S-alil merkaptosistein, alisin, metil propil disülfit, alil metil sülfit and dipropilsülfit molekülleri ile tarımda zararlı böceklerin yok edilmesinde kullanılan, yapay pretroid içeren cyfluthrin molekülünün geometrisi, titreşim özellikleri ve elektronik yapıları ayrıntılı bir şekilde incelendi. Elde edilen kuramsal sonuçlar, bu moleküllerin ticari olarak geliştirilip, kullanılmasından önce onlar hakkındaki belirsizlikleri açıklamaya yardımcı olacaktır. Ayrıca bu tezde, moleküler biyolojinin çözülememiş problemlerinden biri olan protein katlanması ve açılması problemine de değinilmiştir. Problem, çok fazla sayıda atomun dinamiğinin anlaşılması ile çözüleceği için, son derece yüklü bir hesaptır. Bu sebeple, hesabı daha kolay olan, imünogloblin G-bağlanma proteini G' nin küçük dizileri incelenmiştir (DDATKTFT ve türevleri). Molekül dinamiği benzetişiminin, bu dinamiği anlamada etkili olduğu umut edilmektedir. Bulduğumuz sonuçlardan biri, hidrojen bağının yerinin kararlılık üzerinde etkili olduğudur.

Anahtar Kelimeler: Hesaplamalı Kimya, Moleküler Dinamik, Yoğunluk Fonksiyoneli Kuramı, Biyolojik Moleküller.

To my loving family

ACKNOWLEDGMENTS

It has been a long and difficult but overall a satisfying journey for me. Along the way, many people gave me a hand, offered their guidance and most importantly their friendship.

I owe a lot to my advisor Prof. Dr. Şakir Erkoç from whom I learned a great deal. His dilligence, patience and comforting approach at tough times made my life easier. Prof. Dr. Figen Erkoç has a very special place in my heart: I consulted her not only about biology but about many things in my life.

I thank all my collaborators for the work we did together, especially Prof. Ahmet Palazoğlu, Prof. Roland Faller and his group who have been more than collaborators. If the year I spent in Davis was wonderful, Prof. Faller and Prof. Palazoğlu certainly has contributed to that. Allison Dickey really deserves more than words, a big hug: she helped me through Gromacs and more. She has been a "bigger sister" to me. I would like to thank Professors Nesrin Hasırcı and Sibel Başkal for being my thesis observervation committee members and also for their constructive comments about my thesis studies. I thank Assist. Prof. Dr. Hande Toffoli for her DFT lectures that I sit in, lecture notes and for critically reading the DFT part but more importantly for her true friendship.

I made friends for a life time in Davis. They are always in my mind: Hasina, Emel, Sezen, Nazar, Volkan, Alper, Ali, İlknur, Müge and Cengiz. I spent the most enjoyable year of my life in Davis with their friendship. I have plans to unite them all in a city. Özge, Özlem, Serap and Songül, what would I do without you here? I find comfort when I see you, when I talk to you. Nazım, my friend, you have survived my endless questions: if this file does not compile, I know that you are two meters away. Thank you for your friendship. Emre, Barış and Aytun thank you for your help, for the good times and laughter we had in the lab. Aysun Altınay Tecirli, Candan Dözen, Hakan Karaağaç, İpek Güler, Arife İmer, Kıvanç Uyanık, Elif Beklen,

Özlem Aytekin, Zeynep Eke, Gülşen Parlak and Sevim Aygar thank you for your support and friendship.

My father, my mother, my brother, my husband and more recently my little baby girl Elif Ada have been the joy of my life. My father and mother left their home to stay with me for the good part of my PhD work. Without their help, I would not have done it. My husband Bayram has supported me all the way. The life became more beautiful, enjoyable and meaningful with him. The best moment of my years in the METU physics department was meeting him. Now a promise to my little girl: now that I am done, I will not be the woman who leaves in the morning and comes back in the evening, I am at your disposal.

I wrote the good part of this thesis in Starbucks at Bilkent Park: I would like to thank the friendly people working there.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In drug design, natural products have played significant role in recent years. Natural products provide a rich variety of chemical compounds and biological activities, therefore they have been used in antibiotics [1, 2, 3, 4, 5], in anticancer agents [6, 7, 8, 9, 10, 11, 12] and in drugs for cardiovascular diseases [13, 14, 15, 16, 17] etc. In the future, one can expect that research on drug design based on natural products will intensify. Although there are a number of drugs from plant material, mixtures with unknown and potentially unexpected effects are not suitable for veterinary and medicinal use. Only pure chemicals with known pharmacological actions and potency can be developed for medicinal use; a rather long process.

Computational chemistry is used as an effective method with computers to which drug design is based. Here the goal is to understand if and how a molecule will bind to a target molecule and how the two molecules will interact. Computational chemistry is also used as a tool for structure-activity relationship studies and facilitates extrapolation of biological effects of functional groups, molecule geometry, design of enzyme inhibitors and agonists.

Computer-aided drug design uses molecular mechanics (MM), molecular dynamics (MD) and semi-empirical, ab-initio methods and density functional theory (DFT) at different levels. MM and MD are used to find the conformational changes of the target molecule and the small molecule that one uses as an ingredient of the drug. Quantum chemistry methods, such as DFT provide optimized parameters for the MD simulations. Understanding the electronic properties of the molecule would help us to know its binding affinity and hence provide important information on drug design. One drawback is: one cannot get the final form of the drug from

simulation results, and simulation at best, reduces the number of iterations for the final design.

In this thesis, with the help of MM, semi-empirical methods, ab-initio (Restricted Hartree Fock (RHF)), and DFT, we studied the structural properties, vibrational modes and electronic structure of seven small molecules (whose details will be given below). Separately, we have also studied the problem of protein folding with the help of MD. The results of our work have been published [18, 19, 20, 21, 22].

We studied various components of fresh garlic that recently has attracted much public attention due to beneficial health effects such as lowering high blood pressure, namely allicin, methyl propyl disulfide (MPD) and aged garlic extract (AGE) including S-allylcysteine (SAC), S-allyl mercaptocysteine (SAMC), and volatile metabolites of allicin that is allyl methyl sulfide (AMS) and dipropylsulfide (DPS) molecules. Garlic and onion contain more than 100 biologically useful secondary metabolites. The above molecules are only a small representative group. The nature and origin of onion and garlic flavor compounds have been studied since the 1940s. Both garlic and onion have sulfur-containing compounds which make them beneficial for health and also give them their strong odor. Onion contains chromium which improves cellular response to insulin. It also has high quercetin content, a known antioxidant with positive cardiovascular protective role. Other antioxidants such as beta-carotene, vitamins C and B, potassium and selenium are abundant in garlic and onion. Both also contain traces of prostaglandins A1 and E which lower blood pressure and cholesterol levels. Onion also has been shown to help build strong bones. Because of all these properties and some other that we have not mentioned, we have studied the six molecules found in garlic and onion, in detail. Next, we discuss the properties of the six molecules in some more detail.

S-allyl cysteine (SAC) and S-allyl mercaptocysteine (SAMC) are bioavailable water-soluble organosulfur compounds (OSC) from aged garlic extract (AGE). AGE contains antioxidant photochemicals, which prevents oxidant damage. Oxidant damage causes reactions such as oxidative modification of DNA, proteins and lipids [23, 24, 25]. Beneficial health effects attributed to AGE are: reducing the risk of cardiovascular disease, stroke, cancer and aging, including the oxidant-mediated brain cell damage that is implicated in Alzheimer's disease.

Maldonado et al. [26] have shown SAC to prevent tubular and glomerular adverse effects induced by the antibiotic gentamicin due to antioxidant activities. Allicin was incubated with cysteine at physiological temperature and pH, under conditions mimicking those in the intestinal tract, and was found to react with cysteine in less than one minute, yielding two moles of SAMC per mol of allicin. Other garlic-derived compounds, including diallyl trisulfide, diallyl disulfide, and ajoene, were also transformed to SAMC in this in vitro model system. Furthermore, allicin can also react rapidly with glutathione- or thiol-containing proteins to produce SAMC. These results suggest that after the consumption of garlic, SAMC can be a major metabolic product in the intestinal tract and could, therefore, reach a high local concentration in the intestinal mucosa, thus enhancing its effects on colon cancer prevention. Xiao et al. [27] concluded that SAMC exerts antiproliferative effects by binding directly to tubulin and disrupting the microtubule (MT) assembly, thus arresting cells in mitosis and triggering c-Jun NH2-terminal kinase (JNK1) and caspase-3 signaling pathways that lead to apoptosis. Their proposed mechanism of action was by forming disulfide bonds with thiol-containing amino acids (such as cysteine residues) in tubulin. Alicin, was shown to inhibit the proliferation of human breast, endometrial, and colon cancer cells. It is rapidly metabolized, both in vitro and in vivo. Therefore SAMC seems to be a bioavailable natural antitumor agent.

When raw garlic bulb is crushed/damaged, alliinase is released from the tissue and produces allicin by cleaving alliin from garlic oil [28]. Allicin (diallythiosulphinate), another organosulfur compound (OSC), a major ingredient of fresh garlic extract is in turn the precursor to several sulfur-containing compounds responsible for the flavor, odor and pharmacological properties of garlic and is rapidly converted to diallyldisulfide and others. Some of the reported beneficial biological effects of allicin include: broad spectrum of antimicrobial activity, antihyperlipidaemic and antihypertensive effects, immunmodulating effects with promising therapeutic use for chronic inflammatory diseases, antioxidant activity, anti-thrombotic, anti-atherosclerotic and anti-cancer activities [29, 30, 31, 32, 33, 34, 35, 36].

Methyl propyl disulfide (MPD) is an organosulfur compound (OSC) present in garlic and onion. It has low solubility in water and mainly used as a flavoring agent (FAO No: 565) [37, 38]. Studies of chemopreventive action of MPD in multi-organ carcinogenesis bioassay showed significant inhibitory effects on the development of glutathione S-transferase placen-

tal positive foci in rats [24, 39, 40]. In addition to the inhibitory effects mentioned above for the ITO test, a decrease in the formation of 8-hydroxydeoxyguanosine, a DNA adduct generated by reactive (activated) oxygen species. More recently, studies on its potential for development as a cancer preventive agent added to food as a flavoring agent [41] due to its lipid peroxidation inhibitory actions are published.

Allyl methyl sulfide (AMS) is a low molecular weight volatile thiosulfinate, one of 21 compounds from garlic flavors, in garlic oil. Its amount in garlic oil changes depending on the extraction methods [42, 43]. AMS was reported as the main metabolite of allicin ($\geq$90%) and reached maximum levels in four hours and persisted over 30 hours indicating that it is a product of systemic metabolism. Allicin does not carry a methyl group, therefore the source of methyl group in AMS is the S-adenosylmethionine of liver. AMS, the physiologically active metabolite, is detected in breath together with acetone [44].

Dipropylsulfide (DPS) molecule is a lipid-soluble organosulfur which is a volatile component of both onion and garlic. This molecule metabolizes to the corresponding sulfoxide and sulfone. Works on the sulfoxide and sulfone show that they are relatively stable under physiological conditions [45].

OSCs have a wide range of medicinal properties and pharmacological effects attributed to phytochemical content. The OSCs have been shown to have effects on plasma lipids and lipid and cholesterol metabolism; blood pressure and vascular resistance; fibrinolysis and blood coagulation; platelet adhesion, platelet aggregation and thromboxane formation. In addition, antioxidant and radical scavenging effects; antifungal, antibacterial, and antiviral effects; organospecific antitoxic effects; antitumor and antimutagenic effects; and hormonal effects have been reported [29, 41, 25, 24, 39].

In another work reported in section 3.4 we have studied a pesticide molecule called Cyfluthrin [21]. It is a synthetic pyrethroid with a cyano group. It is effective as a stomach poisson and via contact. It is used on a wide range of insects ranging from ants, cockroaches to cabbage-worm. It has been widely used on the insects that chew on the crops such as cotton, peanut

4

etc. It is also used for pest control in the context of public health, such as malaria control.

The final part of this thesis is devoted to the study of folding problem in small peptides. One of the most important unsolved problems in molecular biology is the protein folding problem which is basically understanding and predicting the three dimensional native structure of the protein and its intermediates from the information of its amino acid sequence. To understand both the protein folding and unfolding dynamics, Molecular Dynamics (MD) simulation is a powerful tool that provides us with a detailed analysis of the interactions between the atoms of the protein and the solvent [46]. In atomistic MD, simulation of unfolding as opposed to folding is normally the more viable route [47, 48, 49, 50, 51, 52, 53, 54, 55] as proteins fold on timescales from microseconds to minutes. Even with the fastest computers, such a simulation would take years of computer time. Therefore, one has to speed up the simulation which is possible by studying unfolding at high temperatures. Another reason to favor the simulation of unfolding is that one starts with the well-defined native state (usually taken from the Protein Data Bank (PDB) [56]) in contrast to folding where one may have difficulty in finding proper initial states out of virtually infinitely many. There are, however, a few cases of short chain peptides where folding can be observed. Ho and Dill simulated the folding of a number of very short peptides [57] which provide a good starting point for selecting fragments that reliably fold into well defined structures. There are at least two reasons for choosing a short peptide: The nanosecond simulation time scale is expected to match the real life folding and unfolding of a peptide and, some peptides tend to mimic proteins as far as the folding/unfolding properties are concerned [58, 59, 60]. Here, we studied the short sequence (DDATKTFT) of the Immunoglobulin G-Binding Protein G, reported to fold into a hairpin-turn structure [57] although there is some debate [61]. Here, the two C-terminal residues are dangling, i.e., do not appear to have interaction partners in other sections of the strand marking the minimum of sequence required for inducing a stable reversal of the peptide backbone. We again identify the harmonic modes and additionally study their dependence on temperature and sequence length from 6 to 8 amino acids. Thus, we are studying not only this sequence but also two shorter variants (that is DDATKT and DDATKTF). Our work is in part based on the Karhunen-Loeve (KL) expansion [62, 63, 64] which is a special type of principal component analysis (PCA) to reduce the dimensionality of the data. PCA is an extremely useful tool to extract meaningful spatiotemporal information from the data. In our

case, PCA helps us identify the relevant, small number of modes or degrees of freedom of the atomic fluctuations of the protein out of the entire simulation. In the literature, there have been several studies along this line [55, 65, 66, 67, 68]. The details of the PCA are given in Appendix B.

CHAPTER 2

METHODS OF CALCULATION

2.1 Classical Methods

2.1.1 Molecular Mechanics

In the study of systems with many particles, such as hundreds or thousands of atoms, one cannot use analytical techniques. In fact, it is well-known that for any realistic potentials (such as Coulomb type) analytic solution of a 3-body problem for long times is not possible. So, that is why one resorts to numerical techniques. But, even in the case of numerical techniques, one has to make many simplifying assumptions to study, for example, large molecules. Molecular Mechanics (MM) [69] is one such approximation: one assumes that each atom is modeled by a single particle bonded with the other atoms via springs. The electrons are simply assumed to be distributed over the nuclei in a way consistent with energy-minimization. Therefore, what is explicitly studied is the motion of nuclei in MM.

Perhaps the most crucial ingredient in MM calculations is the choice of the potential energy function or the so called force field. The choice of the potential energy or the force field depends on the purpose of the calculation: therefore there are many choices. In our calculations the HyperChem package program [70] has been used, which offers four MM force fields: MM+ [71], AMBER [72, 73, 74], BIO+(CHARMM) [75] and OPLS [76, 77]. In the computations MM+ force field has been used, which is an extension of MM2. The latter was developed by Allinger [71, 78] which is efficient for small organic molecules.

The force field in MM+ consists of the following terms:

$$E = E_{str} + E_{ang} + E_{tor} + E_{oop} + E_{vdW} + E_{el} + E_{cross}. \qquad (2.1)$$

In what follows, we will explain briefly what each term represents. The first two terms denote respectively the bond-stretching and angle-bending energy terms which are, in their general form, modeled as harmonic (Hooke) type potentials. More explicitly, bond-stretching energy is given as

$$E_{str} = \sum_{bonds} \frac{1}{2} K_r (r - r_0)^2, \qquad (2.2)$$

where K_r denotes the force constant and r denotes the bond distance. Clearly a choice about the origin is made here: r_0 denotes the equilibrium value which is usually taken from an experiment or from some quantum chemical calculation. Similarly the angle-bending energy reads

$$E_{ang} = \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2, \qquad (2.3)$$

where θ denotes the bond angle and the rest is analogous to the stretching-energy case.

As is expected for large deformations, Hooke's law and the harmonic assumption will break and one has to modify both equations 2.2 and 2.3 to some more realistic potentials, such as a Morse function. But Morse function potentials are expensive to simulate, therefore they are not generally used in MM+. Instead the following expression is used

$$E_{str} = 143.88 \sum_{bonds} \frac{1}{2} K_r (r - r_0)^2 \{1 - \text{switch}(a_1, a_2, a_3) \text{CS}(r - r_0)\}, \qquad (2.4)$$

where "switch" is the cut-off function. With $a_1 \equiv r - r_0$, $a_2 \equiv -CS/3$, $a_3 \equiv -4CS/3$ and taking $CS = -2.0$, the cut-off function is assigned the following properties:

$$\text{switch}(x, a, b) \equiv 1 \qquad \text{for } x \leq a$$

$$\text{switch}(x, a, b) \equiv 0 \qquad \text{for } x \geq b$$

$$\text{switch}(x, a, b) \equiv \frac{(b-x)2(b+2x-3a)}{(b-a)3} \qquad \text{for } a < x < b$$

$$E_{ang} = 0.043828 \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \left\{ 1 + \text{SF}(\theta - \theta_0)^4 \right\}, \tag{2.5}$$

where the default value of the scale factor is $SF = 7.0 \times 10^{-8}$. Note that these expressions are taken from the HyperChem manual [70].

In Eq. 2.1, the third term represents the torsional (or dihedral) energy that arises from the rotation about a bond. Specifically, consider Fig. 2.1 consisting of 4 bonded-atoms.
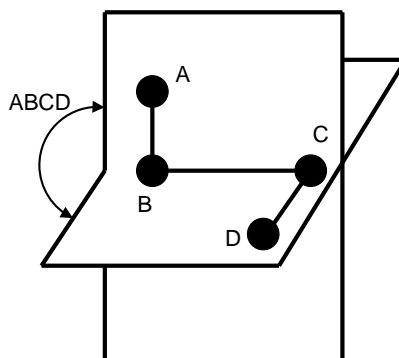
Figure 2.1: Dihedral (torsional) motions

E_{tor} gives the energy associated with a change of rotation about the BC bond. The energy expression reads

$$E_{tor} = \sum_{dihedrals} \left\{ \frac{V_1}{2}(1 + \cos\phi) + \frac{V_2}{2}(1 - \cos 2\phi) + \frac{V_3}{2}(1 + \cos 3\phi) \right\}. \tag{2.6}$$

Here V_1, V_2, and V_3 are in units of energy and set the barrier energy for rotation around the BC bond.

In Eq. 2.1 the fourth term, E_{oop}, represents the energy of out of the plane bending. In certain cases, four bonded atoms prefer to be on the same plane, thus bending out of the plane costs energy (see Fig. 2.2). This energy, in the harmonic approximation, can be written as

$$E_{oop} = \sum_i \frac{1}{2} k_i \, d_i^2, \tag{2.7}$$

where the sum runs over the atoms which do out of plane bending and k_i is the oop-bending constant.
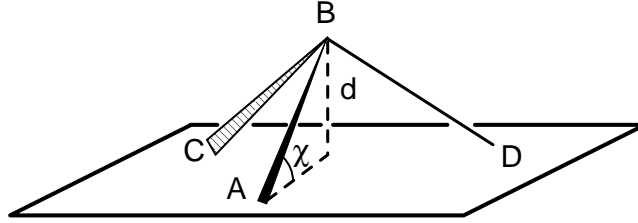
Figure 2.2: Out-of-plane variable definitions

The fifth term in Eq. 2.1 represents the van der Waals type interactions between the atoms that are not directly bonded to each other. Such an interaction necessarily has two components; one attractive, dominant over large separations and one repulsive, dominant over small separations. Usually, the "6-12 Lennard-Jones" potential is chosen to model this force:

$$E_{vdW} = \sum_i \sum_{j>i} \epsilon_{ij} \left[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6 \right], \tag{2.8}$$

where ϵ_{ij} and σ_{ij} are dimensionful parameters with dimensions of energy and length, respectively. Both terms in this expression have their origins in quantum mechanics: the attractive $-1/r^6$ term comes from the polarization of the atoms and is equivalent to induced dipole-dipole interactions. Experimentally van der Waals was the first person to show this force, and theoretically London [79] derived it from quantum mechanics. And in fact, later on, Casimir and Polder [80] and Casimir [81] have shown that such an attractive interaction arises between any neutral object due to fluctuations in the vacuum. The repulsive term, on the other hand,

comes from the Pauli Exclusion Principle, which very broadly speaking, says that electrons (fermions) cannot be in the same state. In fact, Pauli-exclusion principle does not fix the form of the potential to be $1/r^{12}$: this was assumed for the sake of simplicity and it proved to be so in many computations especially in rare gases (but unfortunately, it does not work in hydrocarbons). Thus in some sense, $1/r^{12}$ should be considered as an empirical potential.

There is a problem in the L-J potential: it is long-range (in fact infinite range) and so not very efficient for a simulation. One, therefore, needs to find a way to put a cut-off. In MM+ the following modified form of the potential is used

$$E_{vdW} = \sum_{ij \in vdW} \epsilon_{ij} \left(2.9 \times 10^5 e^{-12.5\rho_{ij}} - \frac{2.25}{\rho_{ij}^6} \right). \tag{2.9}$$

Here the parameters are as follows: $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and ϵ_i refers to the hardness parameter that determines the depth of the well. ρ_{ij} is a dimensionless parameter formed from the van der Waals radii (r_i) of each atoms and is given as

$$\rho_{ij} = \frac{R_{ij}}{r_{ij}}, \tag{2.10}$$

where $r_{ij} = r_i + r_j$ and R_{ij} is a numerical constant with dimensions of length.

The sixth term in Eq. 2.1 is the electrostatic energy coming from the dipole-dipole interactions of polar bonds (not to be confused with the above induced dipole moment that we described). Fig. 2.3 shows this interaction:

In MM+ this interaction is taken as

$$E_{el} = 14.39418\epsilon \sum_{i,j \in polarbonds} \mu_i \mu_j \left\{ \frac{\cos\chi - 3\cos\alpha_i \cos\alpha_j}{R_{ij}^3} \right\}. \tag{2.11}$$

Here ϵ is the dielectric constant of the medium, R_{ij} is the separation vector between the centers of the dipoles (the center is taken to be the mid-point for each dipole). α_i and α_j are the angles between the dipoles and the R_{ij} vector and χ is the angle between two dipoles.

Finally, the last term in Eq. 2.1 represents all the couplings between the first three terms. This idea was put forward by Allinger [71, 78] which gave much more accurate results in
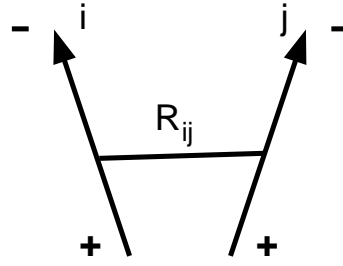
Figure 2.3: Bond dipoles

the simulations. MM+ force field does not include all the possible couplings and just takes the most significant one that is the coupling between bond-stretching and angle-bending. The form of the energy is taken as

$$E_{str-bend} = 2.51118 \sum_{angles} K_{sb}(\theta - \theta_0)_{ikj} \left\{ (r - r_0)_{ik} + (r - r_0)_{jk} \right\}. \tag{2.12}$$

Here, k is the central atom and the angle is defined between the i, j atoms that are on the left and on the right of the k'th atom. Then the above expression couples the stretching of the ik and jk bonds. The numerical factor in front converts the units from millidynes per radian to kcal per degree and K_{sb} is the stretch-bend force constant.

2.1.2 Molecular Dynamics

Molecular Dynamics (MD) is used to study the details of molecular systems ranging from ideal gases, liquids to biomolecules. It is now over a half century old and was first introduced by Alder and Wainright [82] who used it to understand the "phase transition for a hard sphere system". In the context of protein simulations, the first paper using MD appeared in 1977 which was devoted to the study of the bovine pancreatic trypsin inhibitor (BPTI) [83].

MD method uses the Newton's second law to calculate the trajectory of N interacting particles. Let $\vec{r}_i$ be the position of the i'th particle, then Newton's equation reads

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = -\vec{\nabla}_i V_i(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N). \tag{2.13}$$

V_i is the potential energy and $\vec{F}_i$ is the force acting on the i'th particle. To be able to solve this equation, besides the force or the potential energy, one must be given the initial positions and velocities. In the simulations, there are various choices: one can take the positions to be the positions of some lattice site or from pre-existing data, such as a simulation or from some experiment [84, 85]. In our case, we took the initial positions from the Protein Data Bank (PDB) [56]. As for the initial velocities, one can assume some kind of statistical distribution such as Gaussian or Maxwell-Boltzaman [85, 86]. In our case, all the initial velocities were set to zero. The potential energy function will come from force fields that we described in the molecular mechanics section.

Once, the initial configuration and the velocities as well as the forces are given, one integrates the Newton's equation in small incremental times (δt). Clearly there could be many different ways to do this integration, since some kind of approximation has to be adopted. We used the Verlet algorithm [87, 88], which is one of the most commonly used integrator. It uses one forward and one backward time steps in the following way:

$$\vec{r}(t + \delta t) = \vec{r}(t) + \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)(\delta t)^2 + \frac{1}{6}\vec{b}(t)(\delta t)^3 + O((\delta t)^4), \tag{2.14}$$

$$\vec{r}(t - \delta t) = \vec{r}(t) - \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)(\delta t)^2 - \frac{1}{6}\vec{b}(t)(\delta t)^3 + O((\delta t)^4). \tag{2.15}$$

The linear and third terms are cancelled when we add these two terms to get

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)(\delta t)^2 + O((\delta t)^4). \tag{2.16}$$

Clearly, the error in this algorithm is of the order of $(\delta t)^4$. Velocities in this algorithm are not used, but for some purposes they may be needed. For example, to check the conservation of total energy, the kinetic energies must be computed. Therefore, one needs to find the velocities, which can be computed as

$$\vec{v}(t) = \frac{\vec{r}(t + \delta t) - \vec{r}(t - \delta t)}{2\delta t}, \tag{2.17}$$

which has an error of the order of $(\delta t)^2$.

2.1.2.1   MD Simulation Details

In this section, we will present some details of the MD simulation. More specifically, we will first discuss what macroscopic, bulk properties, such as pressure and temperature, are fixed and which ensemble is used, then we will mention how these quantities are fixed with the corresponding algorithms. Finally, we will discuss the periodic boundary conditions relevant to our simulations.

In a MD simulation, total energy is conserved (for a closed system), therefore from the ensembles of statistical mechanics, the most suitable one would be to consider the micro-canonical ensemble. In such an ensemble, NVE (number of particles, volume and total energy) are fixed. But, this ensemble sometimes turns out to be not too useful when one wants to compare the simulation results with the experiments. Therefore, one is free to chose an ensemble that best fits the physical situation [89]. In our simulation, we kept the particle number (N), the pressure (P) and the temperature (T) constant. This ensemble is called the NPT ensemble. During the simulation, in the NPT ensemble, one has to introduce ways to keep the temperature and pressure constant. Clearly, due to rounding errors and the external (or frictional) forces, one needs to readjust these values. To keep the temperature constant, the system can be coupled to a thermostat which scales the velocities. In our simulation, we used the Berendsen thermostat [90]. As for keeping the pressure, one has an option to scale the size of the simulation box and the atomic coordinates, in three different ways: isotropically, semi-isotropically and anisotropically. We used the Berendsen barostat [90] and the anisotropic scaling. More explicitly, in each step, velocities are scaled by a scale factor $\lambda(t)$ given as

$$\lambda = \left[1 + \frac{\delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{\delta t}{2})} - 1 \right\} \right]^{1/2}. \tag{2.18}$$

In the case of anisotropic scaling, that we used, the coordinates and the sytem box vectors are

14

scaled by the following matrix

$$\mu' = \begin{pmatrix} \mu_{xx} & \mu_{xy} + \mu_{yx} & \mu_{xz} + \mu_{zx} \\ 0 & \mu_{yy} & \mu_{yz} + \mu_{zy} \\ 0 & 0 & \mu_{zz} \end{pmatrix}. \tag{2.19}$$

Note that in this case, we do not touch the velocities.

An important ingredient of the simulation is the choice of proper time steps. Larger time steps, would yield a large data but clearly miss certain motions, such as bond vibrations in flexible molecules. On the other hand, too small time steps will necessarily cost more computer time [89]. In the cases where bond-effects are not of particular relevance, there is a way to take them into account with larger time steps [86, 89]. "Bond constraint" algorithms such as SHAKE [91] and LINCS [92] can be used. In our simulation, we used the LINCS algorithm. In this case one can take the time-step to be 2 fs, eventhough the bond effects are relevant at 1 fs.

In a simulation, the system to be simulated has to be in some sense isolated. But, one usually does not want to deal with surface effects. Therefore, it is common to consider the system to be a box in an infinite array of periodic boxes. This gives periodic boundary conditions (PBC) [89, 93]. One then has to define interactions of the particles in our box with the other particles in the neighboring boxes. Note that we keep the number of particles in our box to be constant. In principle one can define many different ways of interactions with the neighbouring boxes but according to the "minimal image convention" a given particle in the box of interest interact with the closest images of the other $N - 1$ particles. To get an accurate account of interactions, the box size has to be taken to be at least $2R$, where R is the cut-off distance for interactions [86].

2.2 Quantum-Chemical Methods

2.2.1 Wave Function Based Methods

Our main goal is to solve the time-independent non-relativistic Schrödinger equation (SE), which in the general form reads

$$\hat{H}\psi = E\psi. \tag{2.20}$$

More specifically, The Hamiltonian operator for a molecule which consist of M nuclei (located at $\vec{R}_k$) and N electrons (located at $\vec{r}_i$) can be written as

$$
\begin{aligned}
\hat{H} = \quad & -\sum_{k=1}^{M}\frac{\hbar^2}{2M_k}\nabla_{\vec{R}_k}^2 - \sum_{i=1}^{N}\frac{\hbar^2}{2m_e}\nabla_{\vec{r}_i}^2 + \frac{1}{2}\sum_{k_1\neq k_2=1}^{M}\frac{1}{4\pi\epsilon_0}\frac{Z_{k_1}Z_{k_2}e^2}{|\vec{R}_{k_1}-\vec{R}_{k_2}|} \\
& +\frac{1}{2}\sum_{i_1\neq i_2=1}^{N}\frac{1}{4\pi\epsilon_0}\frac{e^2}{|\vec{r}_{i_1}-\vec{r}_{i_2}|} - \sum_{k=1}^{M}\sum_{i=1}^{N}\frac{1}{4\pi\epsilon_0}\frac{Z_k e^2}{|\vec{R}_k-\vec{r}_i|},
\end{aligned}
\tag{2.21}
$$

where the first and second terms denote the kinetic energies of the nuclei and the electrons, respectively. The third and the fourth terms represent the Coulomb interactions among the nuclei and among the electrons, respectively. In both of these terms, divergent self-interaction parts are excluded and a factor $\frac{1}{2}$ is introduced to remove double-counting. Finally, the last term gives the nucleus-electron interactions.

The wave function depends on the spins ($\vec{\sigma}_i$) and positions of the electrons as well as the positions of the nuclei: so $\Psi = \Psi(\vec{r}_1,\vec{r}_2,...\vec{r}_N,\vec{R}_1,\vec{R}_2,...\vec{R}_M,\vec{\sigma}_1,\vec{\sigma}_2,...\vec{\sigma}_N)$. For a complicated system, such as a molecule, there is virtually no hope to solve the SE exactly; therefore, one resorts to some reasonable approximations. The first approximation that comes into mind is that of the Born-Oppenheimer (BO)[94] approximation which uses the fact that the motion of the nuclei is negligible compared to the much lighter (about 2000 times) electrons. This naturally leads to a separation of the nuclear wave function from the electronic one. Denoting the nuclear coordinates collectively as $\vec{R}$ and the the electronic coordinates collectively as $\vec{r}$, and skipping the spins, the BO approximation amounts to writing

$$\psi(\vec{R},\vec{r}) = \psi_n(\vec{R})\psi_e(\vec{R};\vec{r}), \tag{2.22}$$

which then gives the explicit form of the electronic SE as

$$\left\{-\sum_{i=1}^{N}\frac{\hbar^2}{2m_e}\nabla_{\vec{r}_i}^2 + \frac{1}{2}\sum_{i_1\neq i_2=1}^{N}\frac{1}{4\pi\epsilon_0}\frac{e^2}{|\vec{r}_{i_1}-\vec{r}_{i_2}|} - \sum_{k=1}^{M}\sum_{i=1}^{N}\frac{1}{4\pi\epsilon_0}\frac{Z_k e^2}{|\vec{R}_k-\vec{r}_i|}\right\}\psi_e(\vec{R};\vec{r}) = E_e(\vec{R})\psi_e(\vec{R};\vec{r}).$$
$$\tag{2.23}$$

Observe that the nuclear coordinates appear as a parameter in the electronic SE equation, which reflects the fact that for different locations of the nuclei, both the electronic energies and the wave functions will change. The fine point here is that the nuclear coordinates enter through the electron-nucleus interaction, which can be interpreted as the electrons moving in the external field created by the nuclei. The nuclei now are basically classical point particles. Once this approximation is made, one adopts the following procedure to get the numerical solution:

Step 1: positions of the nuclei are chosen (potential due to the nuclei is calculated),

Step 2: the electronic SE is solved,

Step 3: the total energy is calculated.

This procedure is repeated until the improvement after these steps becomes smaller than a desired accuracy.

In fact, actual execution of the above steps is highly non-trivial, we will still need to do several other approximations. But before that, introducing new units (that is the atomic units) somewhat simplifies (at least in terms of appearance) the Hamiltonian. We take $\hbar = 1$, $m_e = 1$, $4\pi\epsilon_0 = 1$, $e = 1$. Of course, at the end one can always recover these constants by dimensional analysis. So, in the atomic units, the electronic SE reads

$$\left\{ -\sum_{i=1}^{N} \frac{1}{2}\nabla_{\vec{r}_i}^2 + \frac{1}{2}\sum_{i\neq j=1}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{k=1}^{M}\sum_{i=1}^{N} \frac{Z_k}{|\vec{R}_k - \vec{r}_i|} \right\} \psi_e(\vec{R};\vec{r}) = E_e(\vec{R})\psi_e(\vec{R};\vec{r}). \qquad (2.24)$$

To rewrite the equation in a more compact way, it is a standard procedure to split up the Hamiltonian to one and two electron operators defined as

$$\left\{ \sum_{i=1}^{N} \hat{h}_1(\vec{r}_i) + \frac{1}{2}\sum_{i\neq j=1}^{N} \hat{h}_2(\vec{r}_i,\vec{r}_j) \right\} \psi_e(\vec{r}) = E_e\psi_e(\vec{r}), \qquad (2.25)$$

where

$$\hat{h}_1(\vec{r}_i) = -\frac{1}{2}\nabla_{\vec{r}_i}^2 - \sum_{k=1}^{M} \frac{Z_k}{|\vec{R}_k - \vec{r}_i|}, \qquad (2.26)$$

17

$$\hat{h}_2(\vec{r}_i, \vec{r}_j) = \frac{1}{|\vec{r}_i - \vec{r}_j|}. \tag{2.27}$$

Eq. 2.25 is our main equation to solve. Because of the electron-electron interactions, it is still a very complicated equation: one has to assume a form of the electronic wave function to proceed further. Next we discuss, without going into detail, some well-established approximation techniques to this electronic wave function. This is a standard textbook material which can be found for example in [95, 96].

2.2.1.1 Hartree Method

Neglecting that the electrons are indistinguishable fermions obeying the Pauli exclusion principle, Hartree [97, 98, 99, 100, 101] assumed that the total electronic wave function is in the form of simply the product of individual orthonormal atomic orbitals (AO) $\phi_i(\vec{r}_i)$, namely

$$\Phi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)...\phi_N(\vec{r}_N). \tag{2.28}$$

This form of the many electron wave function is called as the Hartree product. Clearly, this is a somewhat crude approximation.

2.2.1.2 Restricted Hartree-Fock Method (RHF)

For a system with no magnetic properties that is with paired up-down electrons, it is easy to remedy the lack of anti-symmetry in the Hartree approximation by taking the following determinantal wave function (instead of the direct product above)

$$\Phi(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_2(\vec{r}_1) & ... & \phi_N(\vec{r}_1) \\ \phi_1(\vec{r}_2) & \phi_2(\vec{r}_2) & ... & \phi_N(\vec{r}_2) \\ ... & ... & ... & ... \\ \phi_1(\vec{r}_N) & \phi_2(\vec{r}_N) & ... & \phi_N(\vec{r}_N) \end{vmatrix}. \tag{2.29}$$

Here the $\frac{1}{\sqrt{N!}}$ was introduced for normalization. The above determinant is called "the Slater determinant" [102, 103]. Clearly, the RHF approximation [99, 104] is an improvement over

the Hartree approximation and captures quite a large number of physical properties of atoms. But, to actually find the ground state wave function and the ground state energy, one resorts to Rayleigh-Ritz [105] variational approximation, which works in the following way.

Let us suppose that our initial guess of the wave function Φ is given in terms of the orbitals in a Slater determinant form. Then the ground state energy reads

$$E_0 = \frac{\langle \Phi | \hat{H}_e | \Phi \rangle}{\langle \Phi | \Phi \rangle}, \tag{2.30}$$

whose content can also expressed in terms of the Lagrange multipliers (λ_{ij}) as

$$F = \langle \Phi | \hat{H}_e | \Phi \rangle - \sum_{i,j} \lambda_{ij} \left[\langle \phi_i | \phi_j \rangle - \delta_{ij} \right] \tag{2.31}$$

Under arbitrary variation of the orbitals

$$\phi_k(\vec{r}) \rightarrow \phi_k(\vec{r}) + \delta\phi_k(\vec{r}), \tag{2.32}$$

the ground state can be obtained for minimum F, that is when $\delta F = 0$.

Carrying out the functional derivatives in Eq. 2.31 explicitly and setting the right-hand side to zero, one ends up with

$$\hat{h}_1\phi_k(\vec{r}_1) + \sum_{i=1}^{N} \left\{ \int \phi_i^*(\vec{r}_2)\hat{h}_2[\phi_i(\vec{r}_2)\phi_k(\vec{r}_1)]d^3r_2 - \int \phi_i^*(\vec{r}_2)\hat{h}_2[\phi_i(\vec{r}_1)\phi_k(\vec{r}_2)]d^3r_2 \right\} = \sum_{i=1}^{N} \lambda_{ki}\phi_i(\vec{r}_1). \tag{2.33}$$

This equation is usually written in a more compact form, by defining the so called Fock operator

$$\hat{F} = \hat{h}_1 + \sum_{i=1}^{N}(\hat{J}_i - \hat{K}_i), \tag{2.34}$$

where

$$\hat{J}_i \phi_k(\vec{r}_1) = \int \frac{|\phi_i(\vec{r}_2)|^2}{|\vec{r}_2 - \vec{r}_1|} \phi_k(\vec{r}_1) d^3 r_2 \tag{2.35}$$

and

$$\hat{K}_i \phi_k(\vec{r}_1) = \int \frac{\phi_i^*(\vec{r}_2) \hat{P}_{12} [\phi_i(\vec{r}_2)\phi_k(\vec{r}_1)]}{|\vec{r}_2 - \vec{r}_1|} d^3 r_2 \tag{2.36}$$

where the operator $\hat{P}_{12}$ permutes its arguments (i.e., $\vec{r}_1$ and $\vec{r}_2$ of ϕ_k and ϕ_i). Here the $\hat{J}_i$ is the Coulomb operator and $\hat{K}_i$ is the exchange operator. With these definitions, we have

$$\hat{F}\phi_k = \sum_{i=1}^{N} \lambda_{ki} \phi_i. \tag{2.37}$$

The solutions we shall be interested in are those for which the lagrange multipliers are equal to the energies, namely

$$\lambda_{ki} = \delta_{ki} \epsilon_k \tag{2.38}$$

So finally, in terms of the Fock operator, we have an eigenvalue equation (HF equation). The eigenvalues (ϵ_k) are orbital energies as guaranteed by the Koopmans' theorem [106].

$$\hat{F}\phi_k = \epsilon_k \phi_k \tag{2.39}$$

But it is important to notice that this equation is a complicated integral equation which has to be solved in a self-consistent way.

2.2.1.3 Hartree-Fock-Roothan Method

Complexity of the HF equations led researchers to introduce further approximations and techniques. One such technique is that of Roothan [107] who suggested that instead of carrying out a functional variation of F as in Eq. 2.32, one assumes that a fixed basis set of functions $\chi_p(\vec{r})$ is given and the sought-after orbitals can be expressed as a finite sum over these basis set functions as

$$\phi_l(\vec{x}) = \sum_{p=1}^{N_b} \chi_p(\vec{r}) c_{pl}, \tag{2.40}$$

where c_{pl} are complex constants to be determined. Inserting Eq. 2.40 to 2.31, one obtains

$$
\begin{aligned}
F &= \sum_{i=1}^{N} \sum_{m,n=1}^{N_b} c_{mi}^* c_{ni} \langle \chi_m | \hat{h}_1 | \chi_n \rangle \\
&\quad + \frac{1}{2} \sum_{i,j=1}^{N} \sum_{m,n,q,r=1}^{N_b} c_{mi}^* c_{ni} c_{qj}^* c_{rj} \left[\langle \chi_m \chi_q | \hat{h}_2 | \chi_n \chi_r \rangle - \langle \chi_q \chi_m | \hat{h}_2 | \chi_n \chi_r \rangle \right] \\
&\quad - \sum_{i,j=1}^{N} \lambda_{ij} \left\{ \sum_{m,n=1}^{N_b} c_{mi}^* c_{nj} \langle \chi_m | \chi_n \rangle - \delta_{ij} \right\}.
\end{aligned}
\tag{2.41}
$$

Now one finds the minimum of F, from the requirement that

$$\frac{\partial F}{\partial c_{pl}} = \frac{\partial F}{\partial c_{pl}^*} = 0. \tag{2.42}$$

Lagrange multipliers are chosen as in the case of the HF method. Finally one ends up with

$$
\begin{aligned}
&\sum_{m=1}^{N_b} \left\{ \langle \chi_p | \hat{h}_1 | \chi_m \rangle + \sum_{i=1}^{N} \sum_{n,q=1}^{N_b} c_{ni} c_{qi}^* \left[\langle \chi_p \chi_q | \hat{h}_2 | \chi_m \chi_n \rangle - \langle \chi_q \chi_p | \hat{h}_2 | \chi_m \chi_n \rangle \right] \right\} c_{ml} \\
&= \epsilon_l \sum_{m=1}^{N_b} \langle \chi_p | \chi_m \rangle c_{ml}.
\end{aligned}
\tag{2.43}
$$

This equation is the Hartree-Fock-Roothan equation. It can be written as a matrix equation as

$$F \vec{c}_l = \epsilon_l O \vec{c}_l, \tag{2.44}$$

where F is the Fock matrix whose elements are given by

$$F_{pm} = \langle \chi_p | \hat{h}_1 | \chi_m \rangle + \sum_{i=1}^{N} \sum_{n,q=1}^{N_b} c_{ni} c_{qi}^* \left[\langle \chi_p \chi_q | \hat{h}_2 | \chi_m \chi_n \rangle - \langle \chi_q \chi_p | \hat{h}_2 | \chi_m \chi_n \rangle \right], \tag{2.45}$$

and O is the overlap matrix of the basis set functions

$$O_{pm} = \langle \chi_p | \chi_m \rangle. \tag{2.46}$$

As in the case of the HF approximation, one has to solve this matrix equation in a self-consistent way. But now, we have a finite dimensional equation which is significantly simpler than the HF integral equation. Of course the Hartree-Fock-Roothan method cannot give us a solution as good as the HF method. Furthermore, we have assumed that there is a suitable set of basis functions for our problem. This is not immediately clear: there are various choices in the literature and one has to decide which set works better for the problem at hand. One common basis set is that of the Slater type orbitals (STO) which are the atomic orbitals. So one can choose [95]

$$\chi_{nlm}(\vec{r}) = \frac{(2\zeta)^{n+\frac{1}{2}}}{(2n!)^{\frac{1}{2}}} r^{n-1} e^{-\zeta r} Y_{lm}(\theta, \varphi). \tag{2.47}$$

Here ζ is the screening constant and $Y_{lm}(\theta, \varphi)$ are the spherical harmonics. This type of basis set is physically quite relevant, but calculation of the Fock matrix elements in this basis is difficult. To remedy this problem, another set, called Gaussian-type orbitals [95], is introduced.

$$\chi_{nlm}(\vec{r}) = \frac{2^{n+1}}{(2\pi)^{1/4}} \frac{(\alpha)^{\frac{2n+1}{4}}}{((2n-1)!!)^{\frac{1}{2}}} r^{n-1} e^{-\alpha r^2} Y_{lm}(\theta, \varphi). \tag{2.48}$$

But since, near $r = 0$, the Gaussians do not decay as fast the more physical STO, there are some problems such as getting larger than expected contributions to the energy around the origin. To solve these problems STO can be expanded in terms of Gaussians. A specific notation is introduced for this case such as STO-3G which means STO orbitals are expanded in 3 Gaussians. Another commonly used basis set is the plane waves $\chi(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}$ which are more relavant in periodic structures than localized systems.

2.2.1.4 Semi-empirical Methods

As the name suggests, semi-empirical methods [108, 109, 110, 89] make use of the experimental data in solving the HF integrals. But, more properly, semi-empirical methods reduce the expensive and time-consuming task of solving the full HF equations. Recall that Molecular Dynamics was based on solving the Newton's equations for the nuclei, neglecting the

motion of electrons. On the opposite end, HF methods are designed to study the electronic properties taking into account all the electrons. The basic idea of semi-empirical methods is to concentrate on just the valance electrons (which makes sense from the point of chemistry) and take into account the core electrons by reducing the nuclear charge or more accurately by writing charge distribution functions that show the effect of repulsion from core electrons and attraction from the nuclei. Another important approximation adopted by the semi-empirical methods is about the 2-electron integrals of Eq. (2.43). In a given molecular orbit, there will be generally 2 electrons that repel each other. The integrals involving these 2 electrons are eliminated or replaced with experimental data that comes from spectroscopy analysis. In fact, this experimental data is parameterized, with the help of so called "Zero Differential Overlap" (ZDO) approximations, in such a way that the computer code accesses to it during the semi-empirical calculation. There are various semi-empirical methods; the original ones were introduced by Pople et. al, they are named as: Complete Neglect of Differential Overlap (CNDO) [111], Intermediate Neglect of Differential Overlap (INDO) [112] and Neglect of Differential Diatomic Overlap (NDDO) [113]. The crucial difference between these methods is their treatment of the 2 electon integrals. Later Dewar et. al modified these methods and developed Modified Neglect of Differential Overlap (MNDO) [114, 115], Austin Method 1 (AM1) [116], Parameterization Method 3 (PM3) [117, 118]. In this thesis AM1 and PM3 calculations have been performed using the HyperChem program package.

2.2.2 Density Based Method

2.2.2.1 Density Funcional Theory

In the HF and related methods, our main focus was to solve, one way another, the electronic many-body Schrödinger equation and obtain the wave function. This is a rather complicated problem when we have more than a few atoms. An important observation was made by Thomas [119] and Fermi [120] who realized that instead of the wave function, electron density can be used to predict the ground state energy of an atom or a molecule. (Later, Dirac [121] extended the model by adding exchange energy to it.) The wave function, itself, does not correspond to a physical, measurable reality, but its square does. This is the main point of this approach which will significantly lower the computing time since density is a single scalar function depending on the x, y, z coordinates. Our task now is to rewrite the electronic many-

body SE in terms of the density operator defined naturally as

$$\hat{n}(\vec{r}) = \sum_{i=1}^{N} \delta(\vec{r} - \vec{r}_i), \tag{2.49}$$

where $\vec{r}_i$ denote the locations of each electrons. Then, the expectation value of the density operator for the *supposedly known* wave function can be computed as

$$n(\vec{r}) = \langle \Psi | \hat{n}(\vec{r}) | \Psi \rangle = \sum_{i=1}^{N} \int \delta(\vec{r} - \vec{r}_i) |\Psi(\vec{r}_1, ..., \vec{r}_N)|^2 d^3 r_1 ... d^3 r_N$$

$$= N \int |\Psi(\vec{r}, ..., \vec{r}_N)|^2 d^3 r_2 ... d^3 r_N. \tag{2.50}$$

When the wave function is normalized to unity, one has $\int n(\vec{r}) d^3 r = N$, which is the total number of electrons as expected.

Let us recall the many-body electronic Hamiltonian (Eq. 2.24)

$$\hat{H}_e = \hat{T} + \hat{V}_{ne} + \hat{V}_{ee}$$

$$= -\frac{1}{2} \sum_{i}^{N} \nabla_i^2 - \sum_{i}^{N} \sum_{k}^{M} \frac{Z_k}{|\vec{r}_i - \vec{R}_k|} + \frac{1}{2} \sum_{i}^{N} \sum_{j \neq i}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|}. \tag{2.51}$$

We will discuss each term separately. Nuclei-electron interaction part gives

$$E_{ne} = < \Psi(\vec{r}_1, ..., \vec{r}_N) | \hat{V}_{ne} | \Psi(\vec{r}_1, ..., \vec{r}_N) >= -\sum_{i}^{N} \sum_{k}^{M} \int \frac{Z_k}{|\vec{r}_i - \vec{R}_k|} |\Psi(\vec{r}_1, ..., \vec{r}_N)|^2 d^3 r_1 ... d^3 r_N$$

$$= -\sum_{k}^{M} \int n(\vec{r}) \frac{Z_k}{|\vec{r} - \vec{R}_k|} d^3 r = \int n(\vec{r}) V_{ne}(\vec{r}) d^3 r. \tag{2.52}$$

The electron-electron interaction part is more complicated. Just as above, we can calculate the expectation value of the electron-electron potential (operator) to get

$$E_{ee} = \frac{1}{2} \int \int \frac{n^{(2)}(\vec{r}, \vec{r}')}{|\vec{r} - \vec{r}'|} d^3 r d^3 r', \tag{2.53}$$

24

where we have put $n^{(2)}(\vec{r}, \vec{r}')$ to denote the conditional probability that appears in the computation. What we mean by conditional probability is that it gives the probability to find an electron at $\vec{r}$ given that there is an electron at $\vec{r}'$. But, it is clear that we do not have the luxury of knowing this in the DFT approximation, where we only have the density $n(\vec{r})$. Therefore, we have to employ some kind of approximation here. At the first order we assume that any given two electrons are uncorrelated, therefore, we write

$$n^{(2)}(\vec{r}, \vec{r}') = n(\vec{r})n(\vec{r}') + \Delta E_{ee}, \qquad (2.54)$$

where the second part gives the corrections coming from the correlations. Then we have

$$E_{ee} = \frac{1}{2} \int \int \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r d^3r' + \Delta E_{ee}. \qquad (2.55)$$

Finally, we have to deal with the kinetic term:

$$T = -\frac{1}{2} \int \Psi^*(\vec{r}_1...\vec{r}_N) \nabla^2 \Psi(\vec{r}_1...\vec{r}_N) d^3r. \qquad (2.56)$$

It is clear that as it stands, we cannot write the kinetic term in terms of the density (the Laplacian operator does not allow this). Therefore, we make an assumption (not an approximation) here which says that the density can be written in terms of single-particle orbitals

$$n(\vec{r}) = \sum_{i=1}^{N} |\phi_i(\vec{r})|^2. \qquad (2.57)$$

Here $\phi_i(\vec{r})$ are called Kohn-Sham orbitals which are introduced at this stage. Then Eq. 2.56 yields

$$T = -\frac{1}{2} \sum_{i}^{N} \int \phi_i^*(\vec{r}) \nabla^2 \phi_i(\vec{r}) d^3r + \Delta T, \qquad (2.58)$$

where ΔT is the correction to the kinetic energy of the single particle orbitals. Altogether we have

$$E_e = -\frac{1}{2} \sum_i^N \int \phi_i^*(\vec{r}) \nabla^2 \phi_i(\vec{r}) d^3r + \int n(\vec{r}) V_{ne}(\vec{r}) d^3r + \frac{1}{2} \int \int \frac{n(\vec{r})n(\vec{r})}{|\vec{r} - \vec{r}'|} d^3r d^3r' + \Delta T + \Delta E_{ee},$$

(2.59)

where the first term is denoted as T_s, the second term denoted as E_{ext} and the third term as $E_{Hartree}$. Since, the orbitals are related to the density via Eq. 2.57, we can carry out differentiation of E with respect to the density. The last two terms play a crucial role in DFT calculations and together they are called exchange-correlation energy

$$E_{xc} = \Delta E_{ee} + \Delta T.$$

(2.60)

Before we discuss the correction terms (exchange-correlation effects), let us state two important theorems of Hohenberg and Kohn which rigorously show that for any ground state property of a quantum-mechanical system, one can use density instead of the wave function, not as an approximation but as an exact computation.

First Theorem of Hohenberg-Kohn:

Given that one measures $n(\vec{r})$ to be the ground state density of N electrons in some external potential (which could be the Coulomb potential, gravitational potential etc), that external potential is unique up to an additive constant. We skip the proof of this beautiful result since it can be found in many textbooks such as [95, 110] and in the Nobel lecture of Kohn [122], but the main implication of the theorem is that highly complicated wave function need not be computed, but it can be reconstructed from the density. Thus, the theorem proves that there exists a functional $E[n(\vec{r})]$. But, it does not prescribe a way to find or construct this functional. In fact, an exact functional has not yet been found but there are many approximations for such a functional.

Second Theorem of Hohenberg-Kohn:

Let $n(\vec{r})$ be the ground state density, then for any other density $\tilde{n}(\vec{r})$ one has

$$E[\tilde{n}(\vec{r})] \geq E[n(\vec{r})],$$

(2.61)

which is basically a restatement, in terms of the density, of the variational principle saying

26

that the ground state energy corresponds to the ground state wave function and any other wave function (satisfying the boundary conditions) will give a larger energy [123].

Kohn-Sham equations

In Eq. 2.57, we already introduced the Kohn-Sham orbitals that are assumed to exist and that describe single particle orbitals. Here we will elaborate more on this and introduce further approximations [124]. Minimizing Eq. 2.59 with respect to $\phi_i^\star(\vec{r})$ subject to the constraint that the integral of the density gives N, one arrives at the eigenvalue equation

$$\frac{\delta E_e}{\delta \phi_i^\star(\vec{r})} = \frac{\delta T_s}{\delta \phi_i^\star(\vec{r})} + \left\{ \frac{\delta E_{ext}}{\delta n(\vec{r})} + \frac{\delta E_{Hartree}}{\delta n(\vec{r})} + \frac{\delta E_{xc}}{\delta n(\vec{r})} \right\} \frac{\delta n(\vec{r})}{\delta \phi_i^\star(\vec{r})} = \epsilon_i \phi_i(\vec{r}). \tag{2.62}$$

Now we have to make an assumption about the exchange-correlation term E_{xc}. Within the so called Local Density Approximation (LDA) one assumes that it depends only on the local density and not in the derivatives of it: therefore, we have

$$E_{xc} = \int n(\vec{r}) \epsilon_{xc}(n(\vec{r})) d^3 r, \tag{2.63}$$

where $\epsilon_{xc}(n)$ has to be chosen properly to get the properties of a uniform electron gas. With this assumption one has

$$-\frac{1}{2} \nabla^2 \phi_i(\vec{r}) + \left\{ V_{ext}(\vec{r}) + \underbrace{\int \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3 r'}_{V_H} + \underbrace{\epsilon_{xc}[n] + n(\vec{r}) \frac{\delta \epsilon_{xc}[n]}{\delta n(\vec{r})}}_{V_{xc}} \right\} \phi_i(\vec{r}) = \epsilon_i \phi_i(\vec{r}) \tag{2.64}$$

Or, in short one has an eigenvalue problem of the form

$$\left[\hat{T} + V_{eff} \right] \phi_i(\vec{r}) = \epsilon_i \phi_i(\vec{r}), \tag{2.65}$$

which is called Kohn-Sham equation. Here V_{eff} represents the effective potential which is supposed to represent the potential that the single (non-interacting) particles (orbitals) move in. Of course, the effective potential itself depends on the orbitals and hence the equations

have to be solved in a self-consistent way, in the same way we described in the HF case. Finally, energy eigenvalues in the Kohn-Sham equation do not add up to the total energy of the system, therefore, the physical meaning of them is not that clear. In the above computations, we have not talked about the spin at all but implicitly assumed that the ground state is spin-unpolarized: which means that the density of spin-up and spin-down electrons are equal to each other. This of course need not be the case: if they are not equal then one can define a spin-polarization density

$$m(\vec{r}) \equiv n_{up}(\vec{r}) - n_{down}(\vec{r}). \tag{2.66}$$

Then, the energy is a functional of both $n(r)$ and $m(r)$. For the spin-polarized case, if one further assumes that this functional does not depend on the derivatives of $n(r)$ and $m(r)$, one has the so called local-spin-density approximation (LSDA). If on the other hand, one assumes further that this functional depends on the derivatives of the density (and the spin-polarization) in the form $\nabla n(\vec{r})$, $\nabla^2 n(\vec{r})$,... then one has the so called generalized gradient approximation (GGA). These non-local approximations can be used together with the local approximations (so called Hybrid methods) which work quite good in the systems where there are weak interactions (such as hydrogen bonds) [95].

CHAPTER 3

SYSTEMS STUDIED

3.1 S-Allyl Cysteine (SAC) and S-Allyl Mercaptocysteine (SAMC) Molecules

3.1.1 Introduction

The structural, dynamic and electronic properties of two allyl molecules (S-allyl cysteine and S-allyl mercaptocysteine) have been investigated theoretically by performing molecular mechanics (MM), semi-empirical molecular orbital (AM1 and PM3), ab initio (RHF) and density functional theory (DFT) calculations. The geometry of the molecules have been optimized, the vibrational spectra and the electronic properties of the molecules have been calculated in their ground states in gas phase. Geometry optimizations are carried out by using a conjugate gradient method (Polak-Ribiere algorithm) [125]. The calculations were performed using the HyperChem package.

3.1.2 Results and Discussion

3.1.2.1 Structural Information

For the structural calculations, firstly, molecular-mechanics (MM) method is used to get the pre-optimized structure using MM+ force field, in order to speed-up the structural optimization. Then, two levels of semi-empirical methods, AM1 and PM3, within the RHF approximation have been applied. Finally, the higher level of computation, ab initio (RHF) and DFT performed. We used the STO-3G basis with an exchange and correlation potentials B3LYP

29

[126, 127] for DFT and both STO-3G and 3-21G basis sets for ab initio (RHF). In all cases (AM1, PM3, RHF and DFT), we found a convergence criterion of 0.001 kcal/mol/Å to be sufficient for structural optimization.

S-allyl cysteine (SAC) molecule ($C_6H_{11}SO_2N$) contains 6 carbon atoms, 1 sulfur atom, 1 nitrogen atom, 2 oxygen atoms and 11 hydrogen atoms with a molecular point group $C1$. In Fig. 3.1, we present the optimum structure of the SAC resulting from a geometry optimization using the DFT (with STO-3G basis set) level of accuracy. In Fig. 3.2, we show the bond length information between neighboring atoms for every step of our calculations (four different levels of computation). We note that semi-emprical, ab intio (RHF) and DFT calculations give molecular configurations that are close to each other, this behaviour is seen in our previous calculations of other garlic molecules including allicin, methyl propyl disulfide and allyl methyl sulfide [19]. The differences are of the order of a tenth of an Angstrom. The carbon-carbon bond length is found around 1.5 Å which is found similar to the cases containing the other molecules [19]. The C-C double bond, C-S bond and C-H bond length values are found respectively 1.3 Å, 1.8-1.9 Å and 1.1 Å which are found to be same as in the other molecules allicin, MPD and AMS. The bond lengths between C-N and C-O and C-O double bond lengths are calculated respectively as, 1.5 Å, 1.4 Å and 1.2 Å. Other structural information we present here is the angle between atoms in different parts of the molecule. We note that the angles between the 3 carbon atoms on the left and on the right side of the molecule (see Figure 3.2) are (4-5-6) 124° and (7-2-8) 109°. The angles between 2 carbon atoms and the sulfur atom on the left (5-6-3) and right side (3-7-2) are 109° and 112° respectively. The angles involving the nitrogene atom are found to be (7-2-1) 110° and (8-2-1) 108°. For the oxygen atom which sits almost perpendicular to the plane containing the two carbon atoms and the extra oxygen atom labeled 10. The angles including the oxygen atom are (2-8-10) 127° and (9-8-10) 122°. For completeness, we present the actual coordinates of all atoms of the molecule in Table A.1.

Figure 3.1: Optimized structure of SAC (C$_6$H$_{11}$SO$_2$N) from DFT calculation. Bottom right panel shows the charge on the atoms of the molecule.

Figure 3.2: SAC molecule: Bond length (Å) information using semi-empirical (AM1, (PM3)), ab initio {RHF}and [DFT/B3LYP/STO-3G] calculations.

S-allyl mercaptocysteine (C$_6$H$_{11}$S$_2$O$_2$N) molecule contains 6 carbon atoms, 2 sulfur, 2 oxygen, 1 nitrogen and 11 hydrogen atoms with the $C1$ molecular group. We present the optimized structure of SAMC from a geometrical optimization using DFT in Fig. 3.3. In Fig. 3.4, we show the bond length information between the atoms using different level of methods

31

(semi-empirical, ab initio (RHF) and DFT). The first thing we can infer from the figure is that semi-empirical, RHF and DFT calculations give similar molecular configuration with what has been found in the case of the first molecule (SAC) which is studied here and also the molecules from our previous calculations [19]. The difference in the bond lengths in cases of the methods is again in the order of 0.1 Å. The C-C double bond length is found 1.3 Å, C-C bond length is 1.5 Å, C-S bond length 1.8-19 Å, C-N 1.5 Å, C-O double bond length is 1.2 and C-O single bond length is calculated as 1.3 Å. These specific bond lengths are similar to the ones calculated from the other molecules (allicin, MPD, SAC). The bond length between S-S is calculated as 2.1-2.2 Å for this molecule and it is found to be the same as in the cases of allicin and MPD molecules. We get the information about the angles between the atoms from the optimized geometry. From Figure 3.4, the angles between three carbon atoms on the left and right side of the molecule are found to be respectively 124º and 109º which are found as same as for the ones in SAC molecule. The angles (2-3-4), (5-6-7) between the two carbon and one sulfur atoms are calculated as 107.3º and 112.3º, respectively. The angle (4-5-6) between the two sulfur and one carbon is found to be 99º which is found to be same as in allicin molecule [19]. The angles involving the nitrogen atom are found to be (6-7-11) 116º and (8-7-11) 108º which are exactly the same as for SAC molecule. The angles including oxygen atom which makes the double bond with a carbon atom (labeled 10) is calculated as (7-8-10) 123º. The angle between the one carbon and two oxygen atoms is found to be (9-8-10) 124º. For completeness, we present the actual coordinates of all atoms of the molecule in Table A.2.
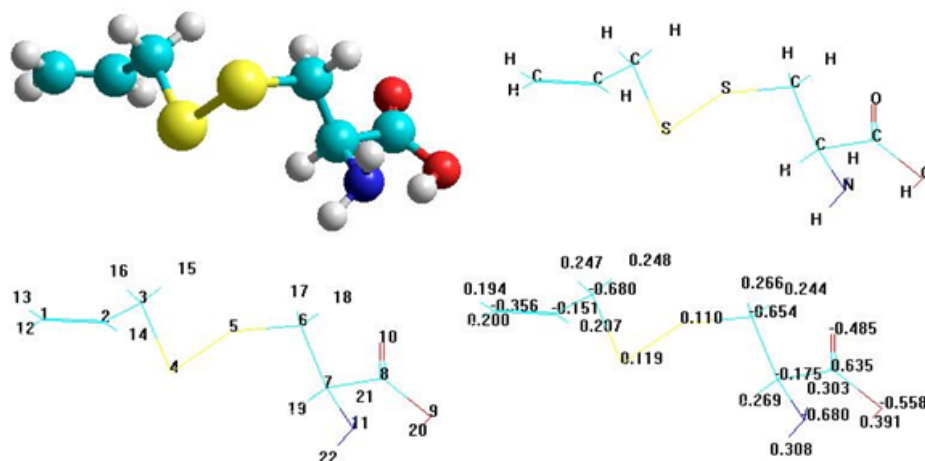
Figure 3.3: Optimized structure of SAMC (C$_6$H$_{11}$S$_2$O$_2$N) from DFT calculation. Bottom right panel shows the charge on the atoms of the molecule.
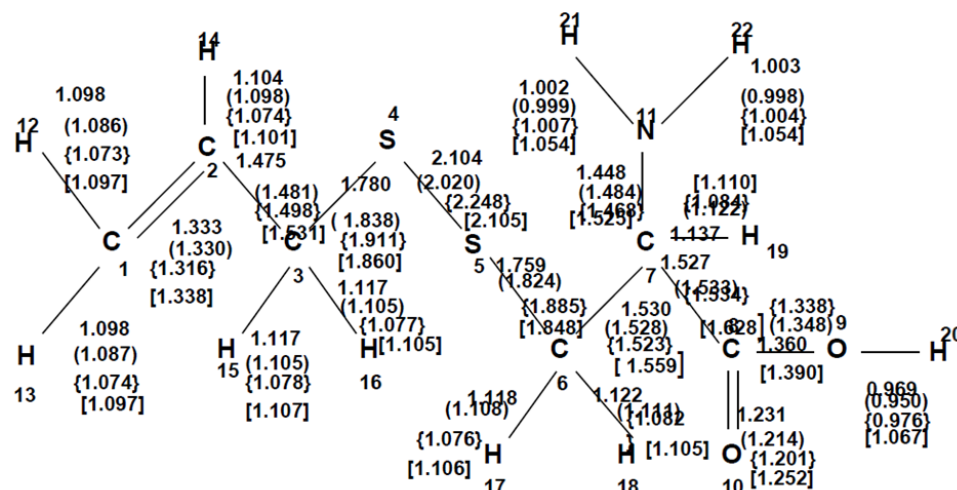
Figure 3.4: SAMC molecule: Bond length (Å) information using semi-empirical (AM1, (PM3)), ab initio {RHF} and [DFT/B3LYP/STO-3G] calculations.

3.1.2.2 Vibrational Properties

The vibrational dynamics of the molecules are examined using two semi-empirical methods (AM1 and PM3) and by the RHF approximation using the 3-21G small basis set. Note that, for the RHF level, we found that the STO-3G basis set give a poor vibrational spectrum which is improved when the 3-21G basis set is used. For the molecules studied here, we found that the DFT with the STO-3G basis set to be unsuitable for the vibrational calculations. Cal-

culations of the vibrational spectra using DFT and higher level basis sets were found to be prohibitive in computational time and were not attempted. And also, we found that a convergence criterion of 0.0001 kcal/mol/Å is needed to reduce the residual forces to less than 0.001 kcal/mol/Å in order to avoid modes with imaginary frequencies.

The frequencies of the normal modes in the harmonic approximation for SAC molecule are given in Table A.3. We note that here, in general, when comparing the calculated frequencies with those of measured by infrared (IR) spectroscopy, one rescales the calculated frequencies. The rescaling of the calculated frequencies is due to the fact that the measured frequencies are anharmonic, on the one hand, and the computational methods used here are implemented with finite basis sets. Scott and Radom [128] have investigated the difference between the calculated and the measured frequencies for a variety of molecules and computational methods. They concluded that with a single scaling factor, one may achieve a relatively good agreement with the experimental observation. However, the scaling factor varies from one method to the other. The scaling factors for the methods used here are 0.9532 for AM1, 0.9761 for PM3 and 0.9085 for RHF/3-21G. Here after, we will present values of our calculations and will add between parentheses the scaled ones. The SAC molecule contains 21 atoms and hence 57 normal modes. We examined again that the vibrational spectrum can be divided in to two bands: a low frequency band containing 46 modes and a high frequency band with 11 modes. The low frequency band for the AM1 method runs from 20.27 (19.32) cm^{-1} to 2087.01 (1989.34) $^{-1}$, while the high frequency band starts at 2928.33 (2791.28) cm^{-1} and ends at 3453.28 (3291.66) cm^{-1}. For the PM3 method, these boundaries for the low band are 12.30 (12.01) cm^{-1} to 1980.35 (1933.02) cm^{-1} and for the high band are 2808.97 (2741.83) cm^{-1} to 3851.70 (3759.64) cm^{-1}. For the RHF/3-21G, we found that these limits for the low band starts at 39.21 (35.62) cm^{-1} to 1967.67 (1787.62) cm^{-1} and for high band from 3164.15 (2874.63) cm^{-1} to 3872.67 (3518.32) cm^{-1}. Although there is no available experimental IR spectra in the literature for the molecules considered, the calculated frequencies qualitatively agree with available frequency values for particular vibrations in various molecular systems [129].

The displacements of some particular modes of the SAC molecule are illustrated in Fig. 3.5. From the Fig. 3.5(a), the mode between the N-H bond (correspond to 3446.22 cm^{-1} frequency) shows asymmetric stretching mode. The asymmetric mode here means displacement

of H and N atoms are not equal for both N-H bonds. The mode between the H-C-H bond (3086.81 cm^{-1}) in Fig. 3.5(b) shows the asymmetric stretching mode and again in the same figure, the C-H bond gives the stretching. In Fig. 3.5(c), the mode (1787.61 cm^{-1}) between the C-O double bond shows stretching mode as it is seen in the SAMC molecule modes also. The mode between the C-S bond in Fig. 3.5(d) gives the stretching mode also.

56-3793.31(3446.22)cm^{-1}

54-3397.70(3086.81)cm^{-1}

(a)

(b)

46-1967.65(1787.61)cm^{-1}

18-1854.78(1685.07)cm^{-1}

(c)

(d)
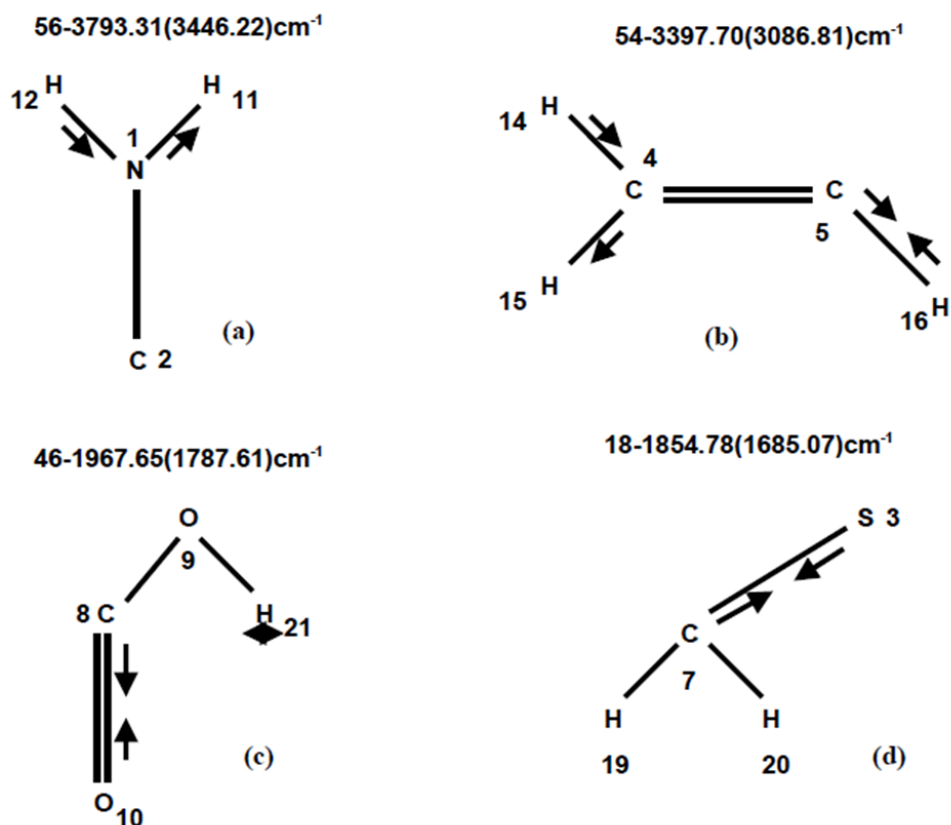
Figure 3.5: Some selected vibrational modes with the corresponding frequencies for the SAC molecule.

The S-allyl mercaptocysteine (SAMC) molecule contains 22 atoms and has 60 normal modes (see Table A.4). For this molecule again the vibrational spectrum can be divided in to two parts as low frequency and high frequency bands. The low frequency band contains 49 normal modes, and the high frequeny band 11 modes. The low frequency band for the AM1 method starts at 16.00 (15.25) cm^{-1} to 2070.03 (1973.15) cm^{-1}, for the high frequency band runs from 2949.35 (2811.32) cm^{-1} to 3455.97 (3294.23) cm^{-1}. For the PM3 method the low frequency band starting from 11.01 (10.74) cm^{-1} to 1977.87 (1930.59) cm^{-1}, while high fre-

quency band from 2819.17 (2751.79) cm^{-1} to 3870.14 (3777.64) cm^{-1}. The low frequency for the RHF/3-21G, we found to be starting from 24.65 (22.39) cm^{-1} to 1986.90 (1805.09) cm^{-1}, for high frequency band starts at 3245.42 (2948.46) cm^{-1} to 3768.23 (3423.44) cm^{-1}. The vibrational modes of some specific modes of the SAMC molecule are given in Fig. 3.6. In Fig. 3.6(a), the mode between the C-O double bond (corresponding the frequency of 1805.01 cm^{-1}) shows the stretching type mode. From the Fig. 3.6(b), the mode between C-H (corresponding the frequency of 3087.12 cm^{-1}) is so called C-H asymmetric stretching mode (type of stretching motion). While the N-H mode (see Fig. 3.6(c)) is again is an example of asymmetric stretching mode, in Fig. 3.6(d) the double bond between the carbon atoms shows the stretching mode explicitly. The modes are examined for these specific bonds between the atoms are similar to those which are seen for the other small molecules [19].
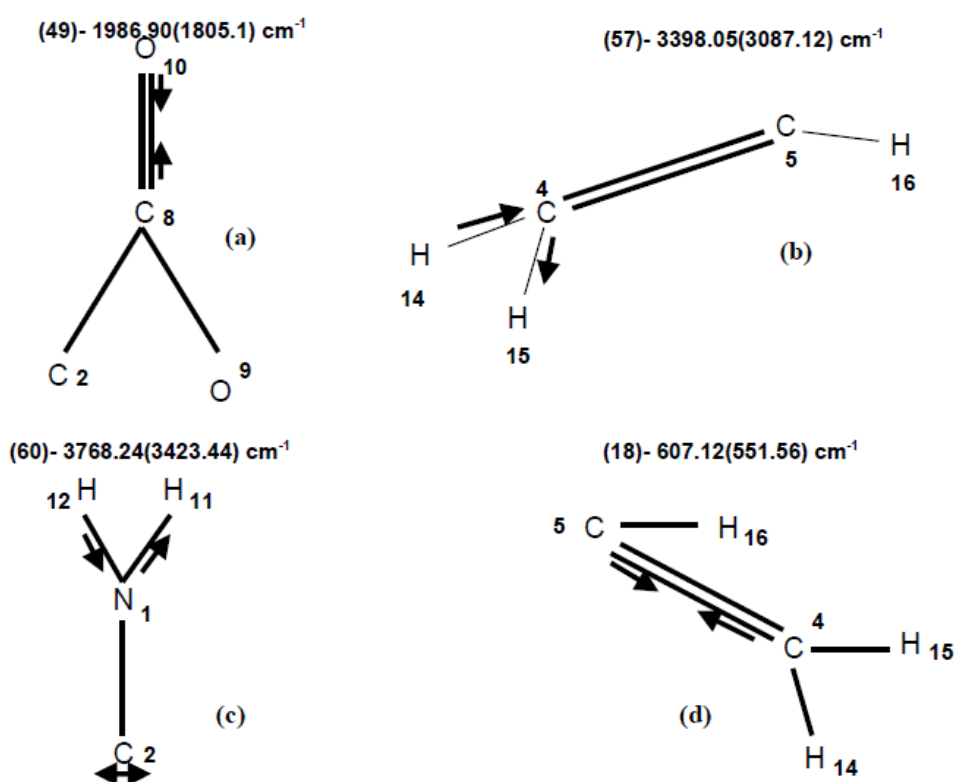
Figure 3.6: Some selected vibrational modes with the corresponding frequencies for the SAMC molecule.

3.1.2.3 Electronic Structure

The electronic structures of the molecules studied were calculated with DFT using the 3-21G basis set with an exchange and correlation potentials B3LYP. We have performed a single point calculation on the previously optimized structure using a DFT calculation with an STO-3G basis set.

In Figures 3.1 and 3.3, we present the excess charge on the atoms of the molecules. We note that the total charge of the molecules is zero (neutral). In Fig. 3.1, for SAC molecule, the charges on the hydrogen atoms are all positive and the value is about 0.2 electron except the one (0.4 electron) which is neighbor of oxygen atom (9). While the carbon atoms (6-7) next to the sulfur atom having a large negative charge value of -0.6 electron, the carbon atom (8) which is neighbor of the two oxygen atoms has a large positive charge value (0.694 electron). The charges on the oxygen atoms are found to be around -0.5, -0.6 electrons, and the sulfur atom has the charge of 0.3 electron. The charges on the atoms of the SAMC molecule are given in Fig. 3.3. From the figure, we can infer that, the hydrogen atoms have positive charge of 0.2-0.4 electrons as it is seen before for the other molecule studied (SAC). The only positively charged carbon atom (8) is found to be 0.6 electron. The other carbons which have the bond with sulfur atoms have negatif values of -0.7 electron. The charge distribution of the atoms is seen similar to the SAC molecule. The difference only becomes significant for the sulfur atoms charge distribution is shared between two sulfur atoms (4-5).

The other information we get from our calculations is the dipole moments, the quadrupole moments and the HOMO and LUMO energy levels and gaps for the molecules. The dipole moments for the molecules are calculated respectively, for SAC is (in Debye) 2.7611 (X: 1.3526, Y: -2.3588, Z: -0.4798), for SAMC molecule the dipole moment is (in Debye) 4.677 (X: -2.0252, Y: -3.9942, Z: -1.2424). The different components of the quadrupole moments of the molecules are given in Table 3.1.

The HOMO and LUMO energy levels of the molecules are calculated. In Fig. 3.7, we show the 3D plots of HOMO and LUMO for SAC molecule. The HOMO energy level is -5.759 eV and the LUMO energy level is 0.216 eV giving rise to a gap of 5.543 eV. The HOMO energy level for the SAMC molecule is found to be -6.386 eV and the LUMO energy level is -1.424

Table 3.1: The different components of the quadrupole moments of the molecules SAC and SAMC (in Debye·Å)

Molecule	XX	YY	ZZ	XY	XZ	YZ
SAC	-60.19	-81.22	-68.50	3.07	-2.72	-1.98
SAMC	-66.47	-94.55	-79.52	-4.47	3.53	-6.44

eV (shown in Fig. 3.8) yielding rises to a gap of 4.962 eV. HOMO of SAC is localized on sulfur and nitrogen atoms, whereas LUMO is localized on double bonded oxygen and carbon atoms and on carbon atoms on the left branch of the molecule (see Fig. 3.7). On the other hand, HOMO of SAMC is localized on oxygen, nitrogen and carbon atoms bonded to them. However, LUMO of SAMC is localized mainly on sulfur atoms (see Fig. 3.8).
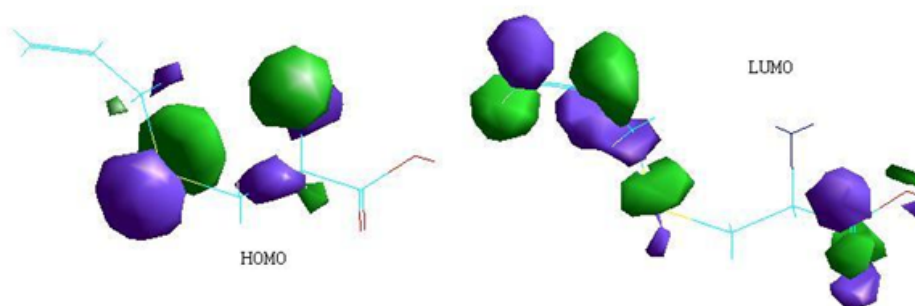
Figure 3.7: 3D HOMO and LUMO plots on the SAC molecule, DFT/B3LYP/MP2/3-21G results.
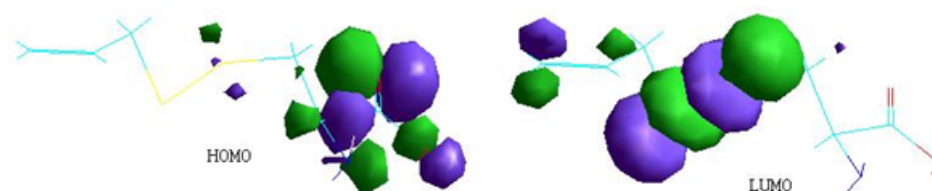
Figure 3.8: 3D HOMO and LUMO plots on the SAMC molecule, DFT/B3LYP/MP2/3-21G results.

3.2   Allicin, Methyl Propyl Disulfide (MPD) and Allyl Methyl Sulfide (AMS) Molecules

3.2.1   Introduction

The structural, dynamic and electronic properties of three ally molecules (allicin, methyl propyl disulfide, allyl methyl sulfide) have been investigated theoretically by performing calculations at three levels of accuracy: semi-empirical molecular orbital (PM3 and AM1), ab-initio (RHF) and density functional theory (DFT). (Just as above, MM method with MM+ force field is used to get pre-optimized structure.) The geometry of the molecules have been optimized, their vibrational spectra and the electronic properties have been calculated in their ground state in gas phase. Geometry optimizations are carried out by using a conjugate gradient method (Polak-Ribiere algorithm). In the calculations, the HyperChem package program was used.

3.2.2   Results and Discussion

3.2.2.1   Structural Information

For the structural calculations, all the procedure and the settings are same as in section 3.1.2.1.

Allicin (diallyl thiosulphinate) ($C_6H_{10}S_2O$) contains 6 carbon atoms, 2 sulfur atoms, one oxygen and 10 hydrogen atoms with a molecular point group $C1$. In Fig. 3.9 we show the optimized structure of allicin resulting from a geometrical optimization using the DFT/STO-3G level of accuracy. We note that the configuration of the molecule is complex due to the presence of an oxygen atom that breaks the symmetry of the molecule. The 3 carbon atoms from each end of the molecule arrange themselves in such a way that they form incomplete hexagons with an angle close to 120° resembling that of a benzene molecule. In Fig. 3.10 we show the bond length between neighboring atoms for different levels of computation. We note from this figure that semi-empirical, ab-initio and DFT calculations give molecular configurations that are close to each other. The differences are of the order of a tenth of an Angstrom with the exception of the bonds involving the sulfur atom attached to the oxygen atom where

the difference in the bonds between the DFT results and the AM1 results is of the order of 0.2
Å, reflecting the complexity of the local electronic structure in the region where substantial
charge transfer occur between the oxygen and the sulfur atoms. Other structural information
worth mentioning is the angle between atoms in different parts of the molecule. We note that
the angles between the 3 carbon atoms on the left and on the right side of the molecule (see
Figure 2) are (1-2-3) 124° and (6-7-8) 122° close to the 120° angle for a benzene molecule.
The angles between 2 carbon atoms and the sulfur atom on the left and right side in Figure 2
are (2-3-4) 107° and (5-6-7) 112° respectively, while the angles between the two sulfur atoms
and the carbon atoms on the left and right side are (3-4-5) 99° and (4-5-6) 94°, respectively.
The oxygen atom sits almost perpendicular to the plane containing the two sulfur atoms and
the carbon atom labeled 3. The angles involving the oxygen atom are (3-4-9) 102° and (9-4-5)
104°. Though presence of the oxygen atom on one of the sulfur atoms breaks the symmetry
of the molecule, the resulting effects on the bond lengths and angles are minor. For complete-
ness, we present in Table A.5 the actual coordinates of all the atoms of the molecule, which
can be used by the reader to reconstruct the molecule.
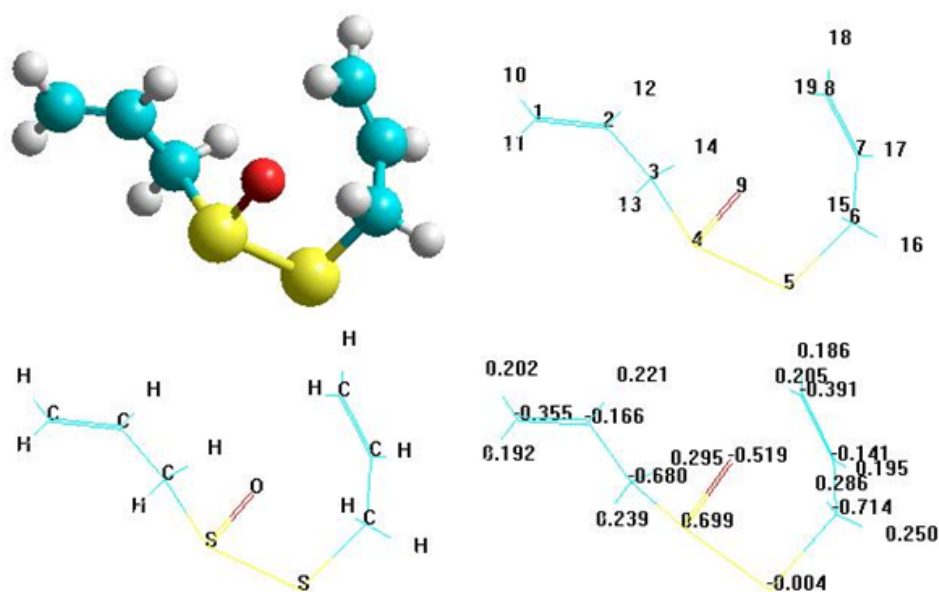
Figure 3.9: Optimized structure of allicin ($C_6H_{10}S_2O$) from DFT calculation. Top left panel
shows the ball and stick model, top right panel shows the number labels of the atoms, bottom
left panel shows the atom symbols, and bottom right panel shows the charge on the atoms of
the molecule.
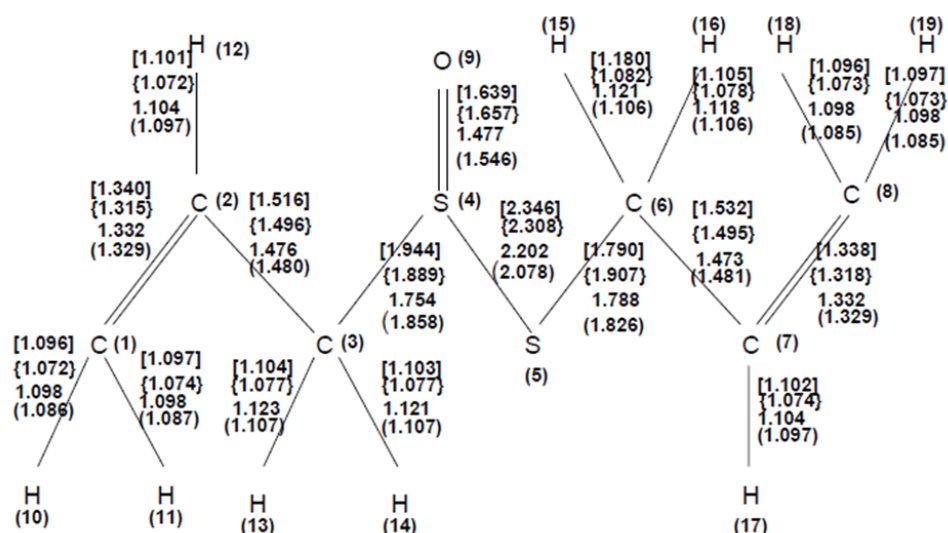
Figure 3.10: Allicin molecule: Bond (Å) information using semi-empirical MO-SCF ((PM3), AM1), ab-initio {RHF} and [DFT/B3LYP/STO-3G] calculations.

Methyl propyl disulfide ($C_4H_{10}S_2$) (MPD) contains 4 carbon atoms, 2 sulfur and 10 hydrogen atoms with the $C1$ molecular group. In Fig. 3.11, we present the optimized structure of MPD from a geometrical optimization using DFT. The coordinates of the optimized MPD molecule are given in Table A.6. In Fig. 3.12, we show bond length values between the neighboring atoms using different methods (semi-empirical, ab-initio and DFT). We can infer from this figure that semi-empirical, ab-initio and DFT calculations yield similar molecular configurations in line with what has been also found in case of allicin molecule. The difference in the bond lengths are again of the order of a tenth of an Angstrom. In general, for all the bond lengths, the difference in the bonds between the DFT and the others is in the order of 0.1 Å. The carbon-sulfur bond length is about 1.8-1.9 Å which is same as in allicin. The other interesting bond length is sulfur-sulfur which is found to be between 2.1-2.2 Å (2.1-2.3 Å for allicin, here sulfur has a bond with oxygen). The carbon-carbon bond is found between 1.5-1.6 Å (same as for the allicin molecule for the case of single bond). From the optimized geometry, we can get information about the angles between the atoms from different parts of the molecule. From Figure 3.12, we note that, the angles between the carbon atom and the two sulfur atoms (on the left side of the molecule) is (1-2-3) 100°, again the angles between the two sulfur and carbon atoms (2-3-4) is found to be 100° (for allicin molecule, same angle (3-4-5) is 99°). The angle between one sulfur and two carbon atoms (3-4-5) is calculated as 110° which is similar to the values for the case of allicin molecule (2-3-4), (6-7-8). On the

right side of the molecule, the angle between 3 carbon atoms (4-5-6) is found to be 112°.
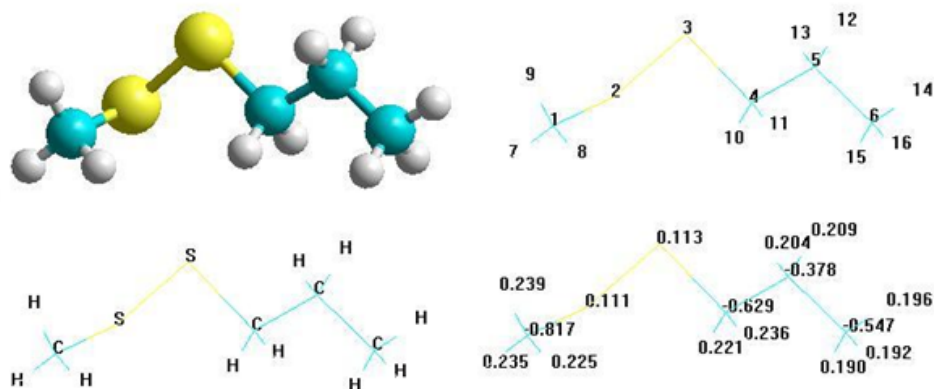
Figure 3.11: Optimized structure of methyl propyl disulfide (MPD) ($C_4H_{10}S_2$) from DFT calculation. Top left panel shows the ball and stick model, top right panel shows the number labels of the atoms, bottom left panel shows the atom symbols, and bottom right panel shows the charge on the atoms of the molecule.
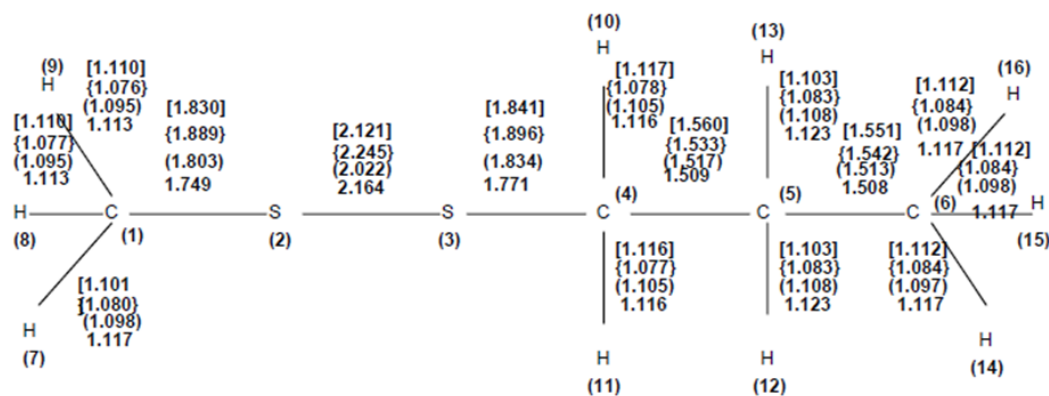
Figure 3.12: Methyl propyl disulfide (MPD) molecule: Bond (Å) information using semi-empirical MO-SCF (AM1, (PM3)), ab-initio {RHF} and [DFT/B3LYP/STO-3G] calculations.

The last molecule in this study (allyl methyl sulfide (AMS), (C_4H_8S) contains 4 carbon atoms, 1 sulfur and 8 hydrogen atoms with $C1$ molecular point group. In Fig. 3.13, we present the optimized structure of the molecule from the results of DFT calculation. The coordinates of the optimized AMS molecule are given in Table A.7. We have calculated the bond lengths between the neighboring atoms using the actual positions which are calculated using all methods considered and discussed in the paper. We show all bond length information in Fig. 3.14.

We draw general remarks about the bonds by analyzing specific parts of the molecule. The general tendency about the bond lengths is found to be similar for all calculation methods (semi-empirical, ab-initio and DFT). This has also been seen for allicin and MPD. In general, the difference (0.02-0.05 Å) in the bond lengths between DFT and other methods is less pronounced than what is seen for the other molecules (0.1 Å). The double bond between the carbon atoms is found to be 1.34 Å which is the same as for the allicin molecule (1.34 Å). The single carbon-carbon bond length is found to be 1.51 Å which is about the same for the other molecules. The carbon-sulfur and carbon-hydrogen bond lengths are calculated respectively 1.85 Å and 1.11 Å (similar with the values for the other molecules 1.83-1.88 Å, 1.11 Å). The angles between the atoms from the optimized structure has been calculated and presented in Fig. 3.14. From Fig. 3.14, angles among 3 carbon atoms (on the left side of the molecule) is (2-3-4) 124°, the corresponding angles of the same environment ((1-2-3), (6-7-8) in Fig. 3.10) for the allicin molecule are found as 124° and 122°. The angle among the two carbon and one sulfur atoms (2-4-1) is found to be 110° similar to that for the allicin molecule (2-3-4 in Fig. 3.10). At the right side of the molecule, the angle among the atoms (1-4-5) is calculated 97° which is close to the values for the MPD molecule (3-4-5 in Fig. 3.12) and allicin (5-6-7 in Fig. 3.10).
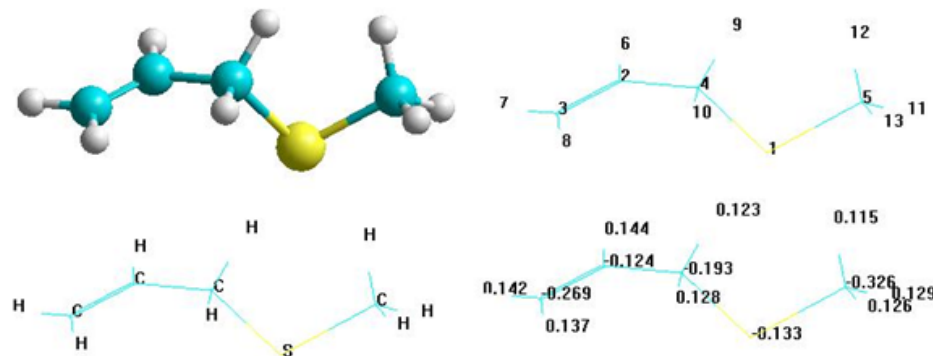
Figure 3.13: Optimized structure of allyl methyl sulfide (AMS) (C_4H_8S) from DFT calculation. Top left panel shows the ball and stick model, top right panel shows the number labels of the atoms, bottom left panel shows the atom symbols, and bottom right panel shows the charge on the atoms of the molecule.
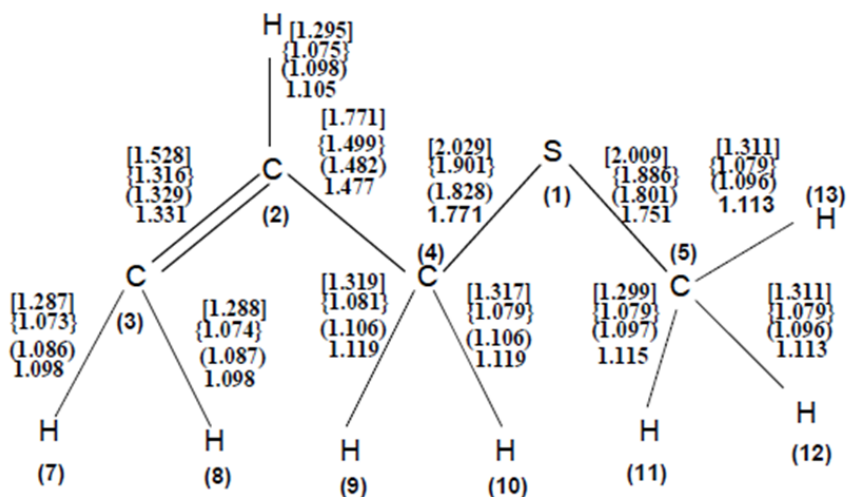
Figure 3.14: Allyl methyl sulfide (AMS) molecule: Bond (Å) information using semi-empirical MO-SCF (AM1, (PM3)), {RHF /3-21G}and [DFT/B3LYP/STO-3G] calculations.

3.2.2.2 Vibrational Properties

The vibrational dynamics of the three molecules studied here were calculated using two semi-empirical MO-SCF methods (AM1 and PM3) and by the Restricted-Hartree-Fock (RHF) approximation using the 3-21G basis set. (Similar criteria are also valid for here as in section 3.1.2.2.)

The frequencies of the normal modes in the harmonic approximation for allicin are presented in Table A.8. As discussed in section 3.1.2.2, the frequencies are rescaled. The scaling factors for the methods used here are 0.9532 for AM1, 0.9761 for PM3 and 0.9085 for RHF/3-21G. Here after, we will present values as they come out of our calculations and will be added between parenthesis the scaled ones. Unfortunately, the measured IR spectrum for allicin is not available in the literature to compare our findings with. However, we have found IR spectra for the two other molecules and comparisons will be made. The allicin molecule contains 19 atoms and hence has 51 normal modes. We found that the vibrational spectrum can be divided in two bands: a low frequency band containing 42 modes and a high frequency band with 9 modes. The low frequency band for the AM1 method runs from 21.8 (20.8) cm^{-1} to 1854.47 (1804.7) cm^{-1}, while the high frequency band starts at 3011.88 (2870.9) cm^{-1}

and ends at 3212.91 (3062.5) cm^{-1}. This range is typical for molecules containing carbon and hydrogen atoms [130]. For the PM3 method, these boundaries are 21.46 (20.9) cm^{-1} to 1848.87 (1804.7) cm^{-1} and 2968.35 (2897.4) cm^{-1} to 3144.90 (3069.7) cm^{-1}. And for RHF/3-21G we found these two bands to be between 51.09 (46.4) cm^{-1} and 1856.21 (1686.4) cm^{-1}; and 3225.6 (2930.5) to 3408.7 (3096.8) cm^{-1}. We note that when rescaled, these values are relatively close to each other.

Now we turn our attention to the displacements of some particular modes that we illustrate in Fig. 3.15. From Fig. 3.15(a), the mode between C-H (corresponding to frequency of 3096.83 cm^{-1}) is the so called C-H asymmetric stretching mode (type of stretching motion). The stretching motion takes place in the plane defined by the carbon and two hydrogen atoms, the displacement of the C atom is less pronounced when it is compared with the displacement of hydrogen atoms (C atom is much more massive than H atoms). The stretching of the two C-H bonds is asymmetric means that the displacements of C and H atoms are not equal for both C-H bonds. In Fig. 3.15(b), the mode (corresponding the frequency of 1686.36 cm^{-1}) between the C-C double bond shows the stretching, and the H-C-H (18-8-19) gives H-C-H out-of-plane wagging motion. This occurs out of the plane defined by the C atom and the two H atoms. The out of plane wagging is a type of bending motion. The mode (corresponding the frequency of 1028.50 cm^{-1}) seen in Fig. 3.15(c) shows rocking motion, and the mode (660.98 cm^{-1}) between the O-S double bond shows stretching motion (see Fig. 3.15(d)).

Methyl propyl disulfide molecule contains 16 atoms and has 42 normal modes. For this molecule the vibrational spectrum can also be divided in to two parts: a low frequency band containing 32 modes and high frequency band with 10 modes. The low frequency band for AM1 method starts from 23.45 (22.35) cm^{-1} to 1434.58 (1367.44) cm^{-1}, while the high frequency band runs from 3010.43 (2869.54) cm^{-1} to 3161.67 (3013.70) cm^{-1}. For the PM3 method, these are 12.97 (12.66) cm^{-1} to 1420 (1386.06) cm^{-1} and 2952.86 (2882.28) cm^{-1} to 3197.85 (3121.42) cm^{-1}. For RHF/3-21G we found that these boundaries are 51.12(46.44) cm^{-1} to 1673.46 (1520.33) cm^{-1} for low frequency band and 3197.61 (2905.02) cm^{-1} to 3357.45(3050.24) cm^{-1} for high frequency band. We note that, values between parentheses are rescaled. The scaled ranges are consistent with the ranges of experimental spectra [131]. The calculated vibrational frequencies of MPD at different levels of calculations are given in Table A.9.

In Fig. 3.16, we present some of the vibrational modes of the methyl proplyl disulfide

(51)- 3408.73 (3096.83) cm⁻¹

(a)

(41)- 1856.21 (1686.36) cm⁻¹

(b)

(24)- 1132.09 (1028.50) cm⁻¹

(c)

(17)- 727.56 (660.98) cm⁻¹

(d)
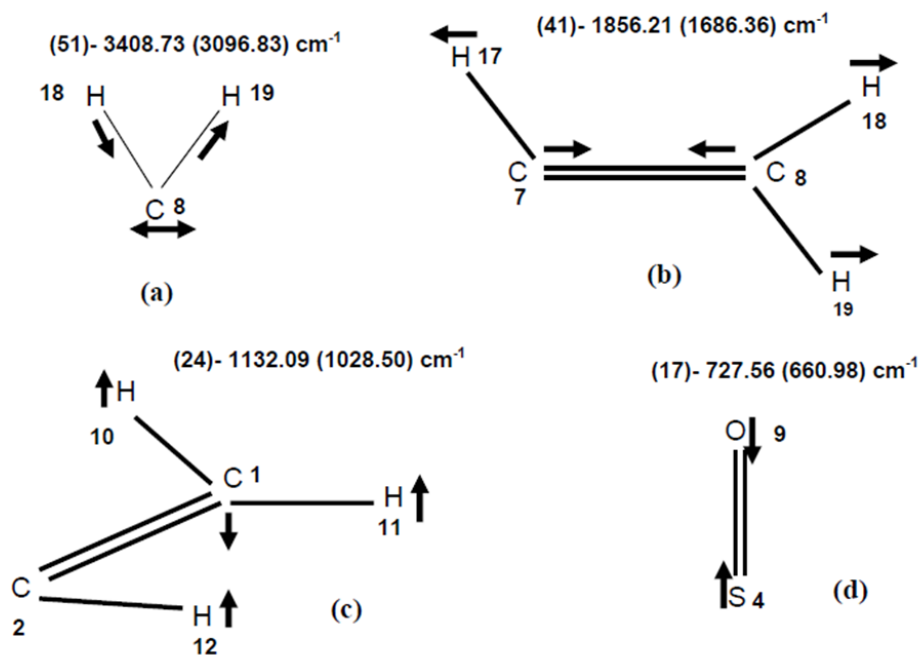
Figure 3.15: Some selected vibrational modes with the corresponding frequencies for allicin molecule.

molecule. From the Fig. 3.16(a) and Fig. 3.16(d), while the modes between C-S and S-S show the stretching modes explicitly, the mode in Fig. 3.16(a) between the H atoms (9-7) gives the out of plane wagging mode. For the modes between H-C-H atoms in Fig. 3.16(b) is the example of symmetric stretching motion. The stretching of the two C-H bonds is symmetric meaning that displacements of the C and H atoms are equal for both C-H bonds. The mode in Fig. 3.16(c) between the H-C-H atoms is an example of asymmetric stretching motion which has been discussed before.

The allyl methyl sulfide molecule contains 13 atoms and has 33 normal modes. The vibrational spectrum is divided in to two parts: a low frequency band containing 25 modes and high frequency band with 8 modes. The low frequency band for AM1 method starts from 33.82 (32.23) cm⁻¹ to 1857.30 (1770.38) cm⁻¹, while the high frequency band runs from 3025.53 (2883.93) cm⁻¹ to 3214.36 (3063.92) cm⁻¹. For the PM3 method, these are 23.40 (22.84) cm⁻¹ to 1852.84 (1808.55) cm⁻¹ and 2972.76 (2901.71) cm⁻¹ to 3200.03 (3123.55) cm⁻¹. For RHF/3-21G we found that these boundaries are 60.61 (55.06) cm⁻¹ to 1856.5 (1686.6) cm⁻¹ for low frequency band and 3234.13 (2938.2) cm⁻¹ to 3395.99 (3085.25) cm⁻¹ for high
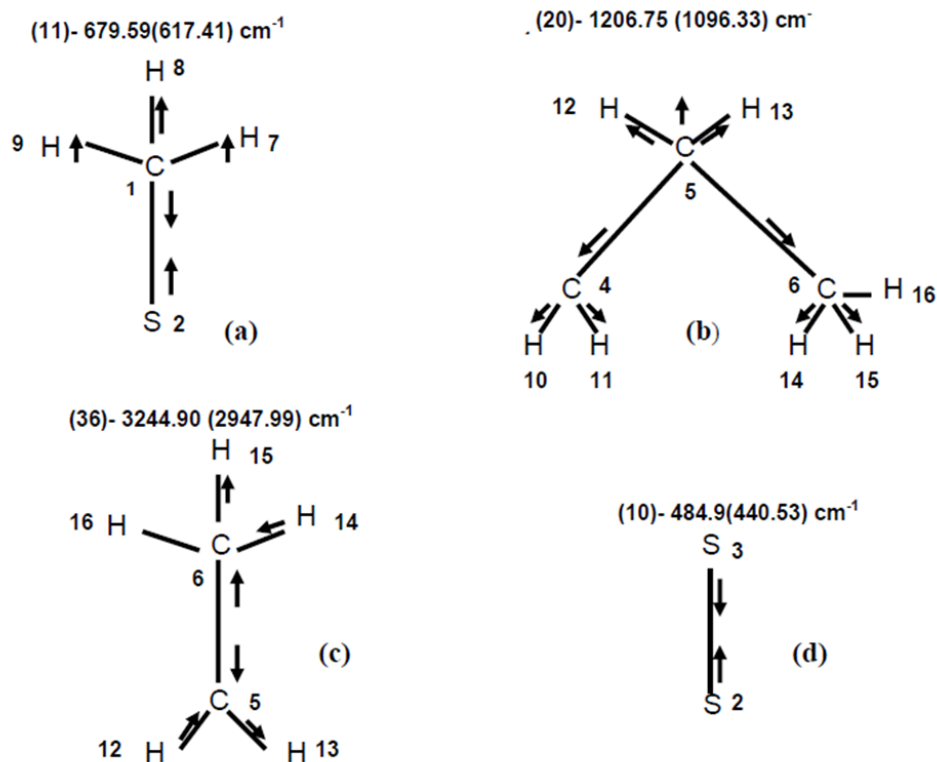
Figure 3.16: Some selected vibrational modes with the corresponding frequencies for methyl propyl disulfide molecule.

frequency band. We note that, the values between the parentheses are rescaled. The scaled ranges are consistent with the ranges of the experimental spectra [131]. The calculated vibrational frequencies of AMS at different levels of calculations are given in Table A.10.

For the vibrational modes of AMS molecule, we present the two selected modes in Fig. 3.17. This molecule has less number of atoms (13 atoms) than the others examined in this study. The vibrational modes for this molecule show much more complicated (mixed) characteristics which make it difficult to sort out the local modes between specific bonds. Here we present two modes which are examined explicitly, in Fig.3.17(a) the mode between the H-C-H shows the symmetric stretching as it is seen for the other molecules. Again, C-C double bond shows the stretching mode which is also examined in the modes of allicin molecule. The mode in Fig. 3.17(b), in case of the C-S bond gives the stretching mode like in MPD molecule and the H-C-H shows the symmetric stretching motion.

Figure 3.17: Some selected vibrational modes with the corresponding frequencies for the ally methyl sulfide molecule.

3.2.2.3 Electronic Structure

The electronic structures of the molecules studied were calculated with DFT using the 3-21G basis set with an exchange-correlation functional B3LYP. We have performed a single point calculation on the previously optimized structure using a DFT calculation with an STO-3G basis set.

In Fig. 3.9, 3.11, and Fig. 3.13 we show the excess charge on the atoms of the molecules. Note that the total charge of the molecules is zero (neutral). In Fig. 3.9 (for allicin molecule), as one would expect that the charge on the hydrogen atoms is always positive and is about 0.2 electron. The carbon atoms neighboring the sulfur atoms have a large negative charge of about -0.7 electrons while the sulfur atom that bonds to the oxygen atom has a positive charge of 0.71 electron with the other sulfur atom almost neutral. Finally, the oxygen atom is negatively charged with -0.528 electron. Note that the left side of the neutral sulfur atom has a net charge of about +0.1 electron while the right side has about -0.1 electron. We present the charge distribution on the atoms of the MPD molecule in Fig. 3.11. For the MPD molecule, we see the same charge values on hydrogen atoms (0.2 electrons). The sulfur atoms that bonds to carbon atom has positive charges 0.11 electron. The carbon atoms neighboring the sulfur atoms have highly large negative charges of about -0.8 and -0.6 electron. In the case for AMS

molecule, the charge distribution on the atoms given in Fig. 3.13, the hydrogen atoms have the charges always positive and 0.1 electron. The sulfur atom neighboring the carbon atoms has a negative charge of about -0.1 electrons. Carbon atoms on the left and right sides of the sulfur atom respectively, have negative charge values of -0.2 and -0.3 electrons.

Other information we obtain from our calculations is the dipole moments, the quadruple moments and the HOMO and LUMO energy levels and gaps for the molecules. The dipole moment values (in Debye) for the molecules are found to be respectively, for allicin is 2.4527 (X: -1.3245; Y: -1.9554; Z: 0.6616), for MPD is 2.5805 (X: 0.0638, Y: 1.0378, Z: 2.3618), and for AMS is 2.1937 (X: 1.0305, Y: 1.2344, Z: -1.4921). Among these three molecules MPD has largest dipole moment. Calculated dipole moment values of the molecules studied are not small, comparable with water value. Same level of calculation gives the dipole moment value of water as 2.289 Debyes (experimental value of dipole moment for water is about 1.85 Debyes). The difference in calculated and experimental dipole moment values for the molecules studied may be same order as that of water. We may conclude that DFT results for dipole moments, probably it is true also for the higher order moments, overestimates the experimental values. The calculated values for the different components of the quadruple moments of the molecules are given in Table 3.2.

Table 3.2: The different components of the quadrupole moments of the molecules Allicin, MPD and AMS (in Debye·Å)

Molecule	XX	YY	ZZ	XY	XZ	YZ
Allicin	-72.72	-62.25	-73.25	-3.95	-0.32	0.34
MPD	-54.14	-50.61	-58.14	-4.18	-0.28	-3.90
AMS	-34.81	-35.62	-43.43	-5.87	1.48	3.49

The HOMO and LUMO energy levels of all molecules are calculated. In Figures 3.18-3.20, we show the 3D plots of the HOMO and LUMO for all three molecules. The HOMO energy level of allicin is -9.556 eV and the LUMO energy level is 1.183 eV giving rise to a gap of 10.739 eV. HOMO of allicin is localized mainly on sulfur, oxygen atoms and C-C double bond at the right branch (see Fig. 3.18), however LUMO is localized only on sulfur and oxygen atoms. The HOMO energy level for the MPD molecule is found to be -6.473 eV and the LUMO energy level is -0.778 eV yielding rise to a gap of 5.696 eV. HOMO of MPD is

localized on carbon atom bonded to sulfur at the left branch (see Fig. 3.19), on the other hand, LUMO is localized on sulfur atom and carbon atom bonded to sulfur at the left branch. For the last molecule, AMS, the HOMO and LUMO energy levels are calculated to be respectively -5.917 and -2.997 eV giving rise to a gap of 2.92 eV. HOMO of AMS is localized only on sulfur atom, whereas LUMO is localized on sulfur atom and carbon atom bonded to sulfur at the left branch (see Fig.3.20). It is interesting to point out that the HOMO-LUMO gap of allicin is considerably larger than that of MPD and AMS molecules. This large difference in gap is an indication that the electronic spectra of these molecules may not be similar.
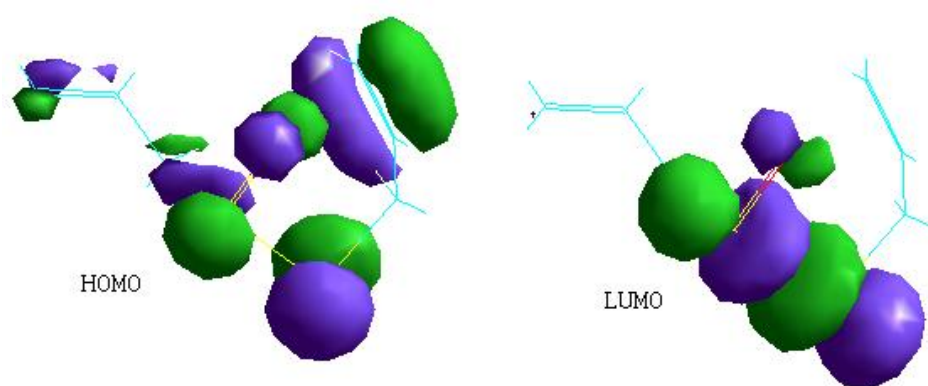
Figure 3.18: 3D HOMO and LUMO plots on allicin molecule, DFT/B3LYP/MP2/3-21G results.
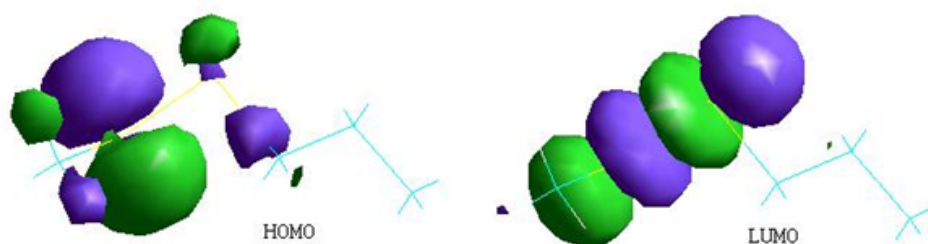
Figure 3.19: 3D HOMO and LUMO plots on methyl propyl disulfide molecule, DFT/B3LYP/MP2/3-21G results.
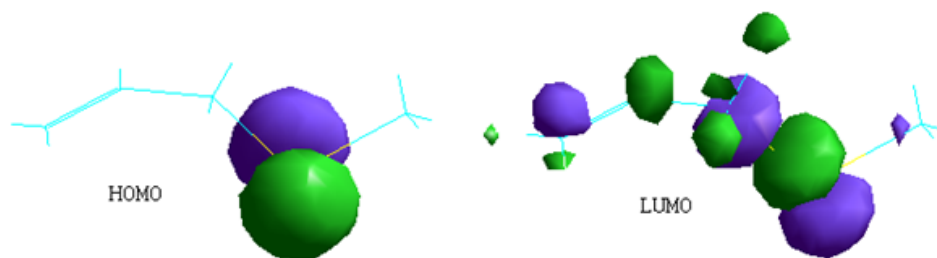
Figure 3.20: 3D HOMO and LUMO plots on allyl methyl sulfide molecule, DFT/B3LYP/MP2/3-21G results.

3.3 Dipropyl Sulfide (DPS) Molecule

3.3.1 Introduction

The structural, vibrational, electronic and QSAR properties of the dipropyl sulfide (DPS) molecule in gas phase have been investigated theoretically by performing molecular mechanics (MM), semi-empirical molecular orbital (AM1 and PM3), ab initio (RHF) and density functional theory (DFT) calculations. The geometry of the molecule has been optimized, infrared spectrum (vibrational modes and intensities) and the electronic properties of the molecule have been calculated in its ground state. Geometry optimizations are carried out by using a conjugate gradient method (Polak-Ribiere algorithm). We have performed all the calculations by using the HyperChem 7.5 packet program.

3.3.2 Results and Discussion

3.3.2.1 Structural Information

DPS molecule ($C_6H_{14}S$) contains 6 carbon atoms, 1 sulfur atom and 14 hydrogen atoms with a molecular point group C_{2v}. The geometry optimization of all methods considered (MM, PM3, AM1, ab-initio/3-21G and DFT/B3LYP/STO-3G). Preoptimization has been performed by applying the MM method using MM+ force field; this makes easier to perform full optimization by extended methods. The SCF convergency is set to 0.001 kcal/mol in the calculations to get sufficient structural optimization.

The optimized structures of DPS molecule with respect to all methods are very similar to each other. Only some of the bond lengths and the bond angles are observed to change slightly. In Fig. 3.21 we show the optimized structure of DPS as resulting from a geometrical optimization using the DFT level of accuracy. The bond length between neighboring atoms of the DPS molecule is shown in Fig. 3.22. Some of the molecular properties of the system considered are given in Table 3.3. Table A.11 depicts the actual coordinates of the atoms from which one can reconstruct the molecule.
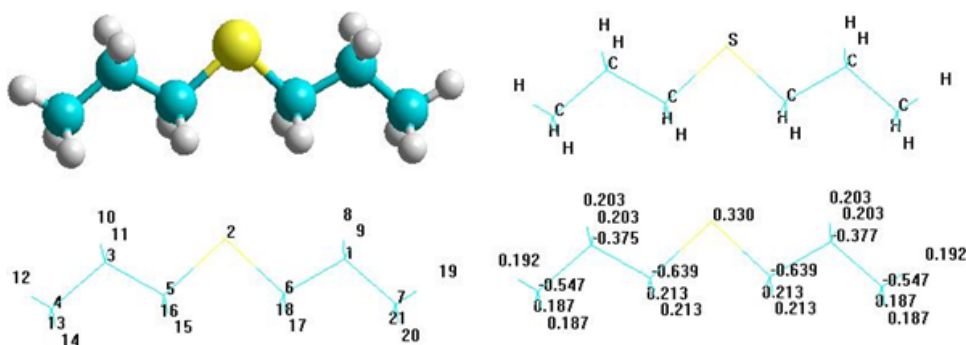
Figure 3.21: Optimized structure of DPS ($C_6H_{14}S$) from DFT calculation. Top left panel shows the ball and stick model, top right panel shows the atom symbols, bottom left panel shows the number labels of the atoms, and bottom right panel shows the charge on the atoms of the molecule.
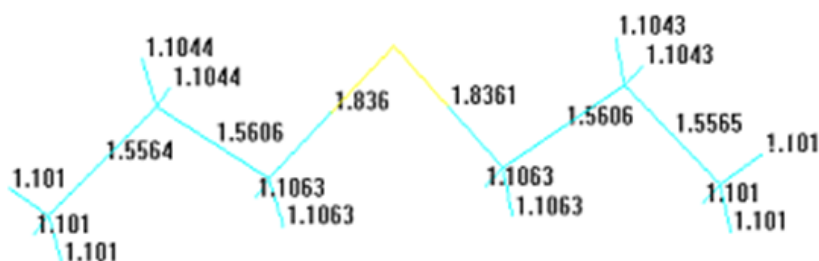
Figure 3.22: DPS Molecule: Bond (Å) Information using DFT/B3LYP/STO-3G calculations.

Table 3.3: Some of the molecular properties of the DPS molecule in its ground state (singlet state) with zero total charge and C_{2v} symmetry according to AM1, PM3, RHF/3-21G, and DFT/3-21G methods.

Quantity	AM1	PM3	RHF	DFT
Number of electrons	44	44	66	66
Number of doubly occupied levels	22	22	33	33
Number of Virtual orbitals	20	20	62	20
Number of total orbitals	42	42	95	53
Number of Primitive Gaussians	-	-	159	159

3.3.2.2 Vibrational Properties

The vibrational dynamics of the DPS molecule studied here were calculated using two semi-empirical MO-SCF methods (AM1 and PM3), the Restricted-Hartree-Fock (RHF) approximation using the 3-21G basis set and DFT at 3-21G basis set.

The calculated frequencies of the normal modes with first five maximum intensities for each method of calculation in the harmonic approximation for DPS are presented in Table A.12 and the IR spectra of DPS are shown in Figures 3.23-3.26. Note that as we discussed in section 3.1.2.2, the frequencies have to be scaled. The scaling factors for the methods used here are 0.9532 for AM1, 0.9761 for PM3 and 0.9085 for RHF/3-21G. Unfortunately, the measured IR spectrum for DPS is not available in the literature to compare our findings with. The DPS molecule contains 21 atoms and hence has 57 normal modes. We found that the vibrational spectrum can be divided in two bands: a low frequency band containing 43 modes and a high frequency band with 14 modes. Here we will present some of the calculated frequency values with their scaled values in parenthesis. The low frequency band for the AM1 method runs from 23.53 (22.43) cm^{-1} to 1435.90 (1368.70) cm^{-1}, while the high frequency band starts at 3010.12 (2869.25) cm^{-1} and ends at 3157.44 (3009.67) cm^{-1}. This range is typical for molecules containing carbon and hydrogen atoms [128]. For the PM3 method, these boundaries are 12.98 (12.67) cm^{-1} to 1422.52 (1388.52) cm^{-1} and 2954.03 (2883.43) cm^{-1} to 3182.15 (3106.10) cm^{-1}. For RHF/3-21G the low frequency band runs from 31.45 (28.57) cm^{-1} and 1674.44 (1521.23) cm^{-1}, while the high frequency band runs from 3196.12 (2903.68) to 3303.59 (3001.31) cm^{-1}. On the other hand, for DFT/3-21G calculations scale factor was not determined. Thus here we just give only the calculated values of these two

bands: low frequency range varies from 104.93 cm^{-1} to 1574.76 cm^{-1}, whereas high frequency range varies from 2958.43 cm^{-1} to 3069.56 cm^{-1}.

The displacements of some particular modes are illustrated in Fig. 3.27. In Fig. 3.27(a), there are two dominant modes, the leftmost mode between H(13)- C(4)-H(14) is so called asymmetric stretching mode (type of stretching motion). The rightmost mode between H(20)-C(7)-H(21) is again asymmetric stretching mode. But the leftmost mode is more dominant than rightmost mode. The stretching motion takes place in the plane defined by the carbon and two hydrogen atoms, the displacement of the C atom is less pronounced when it is compared with the displacement of hydrogen atoms (C atom is much more massive than H atoms). The stretching of the two C-H bonds is asymmetric means that the displacements of C and H atoms are not equal for both C-H bonds. In Fig. 3.27(b), the mode between the C(4)-H(12) bond shows the asymmetric stretching, and the H(13)-C(4)-H(14) gives symmetric stretching, in which two H atoms take equal displacement. The mode seen in Fig. 3.27(c), shows asymmetric stretching between the C(7)-H(19) and symmetric stretching between the H(20)-C(7)-H(21). From Fig. 3.27(d), it is seen that the mode between the C(4)-H(12)-H(13)-H(14) is symmetric stretching, and also seen that there is asymmetric stretching between the H(10)-C(3)-H(11) and between the H(8)-C(1)-H(9). Finally, it is said that DPS molecule generally shows stretching motion.
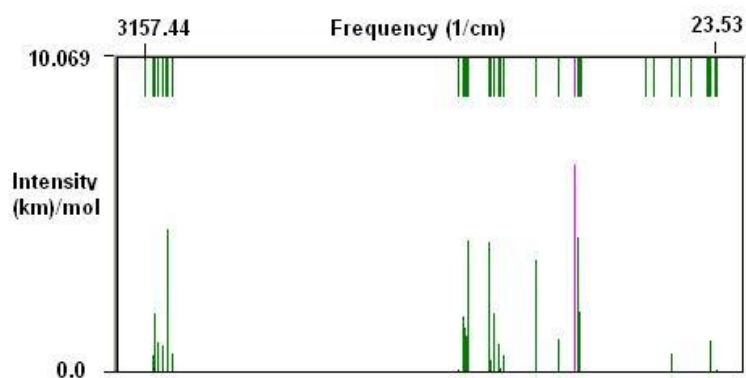
Figure 3.23: IR spectrum of DPS molecule at AM1 calculation.
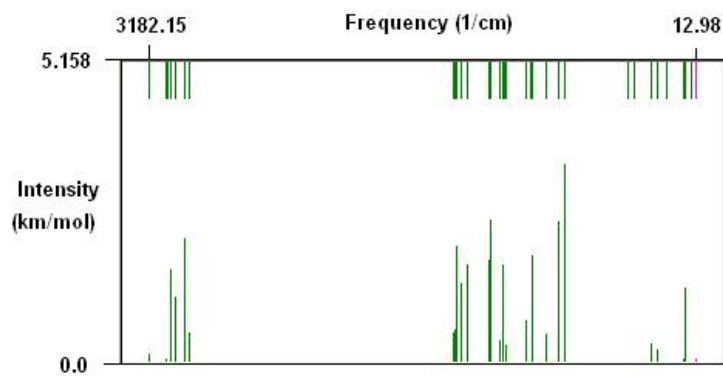
Figure 3.24: IR spectrum of DPS molecule at PM3 calculation.
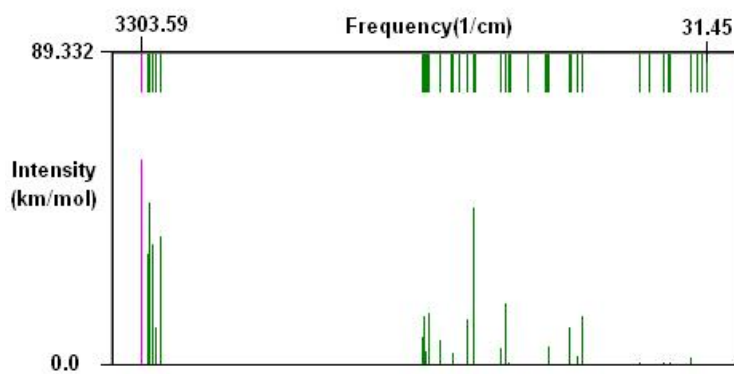
Figure 3.25: IR spectrum of DPS molecule at ab-initio (RHF) calculation.
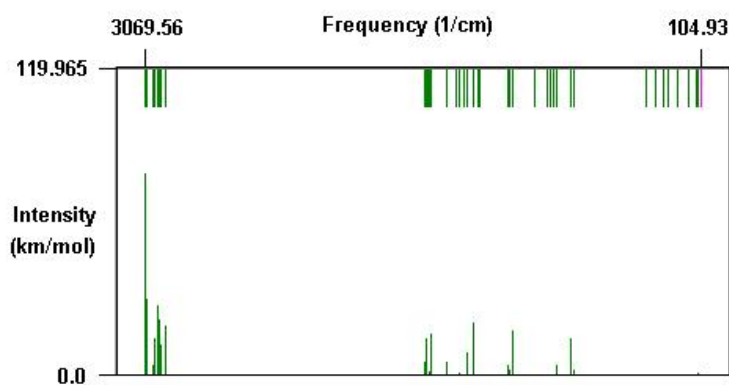
Figure 3.26: IR spectrum of DPS molecule at DFT calculation.
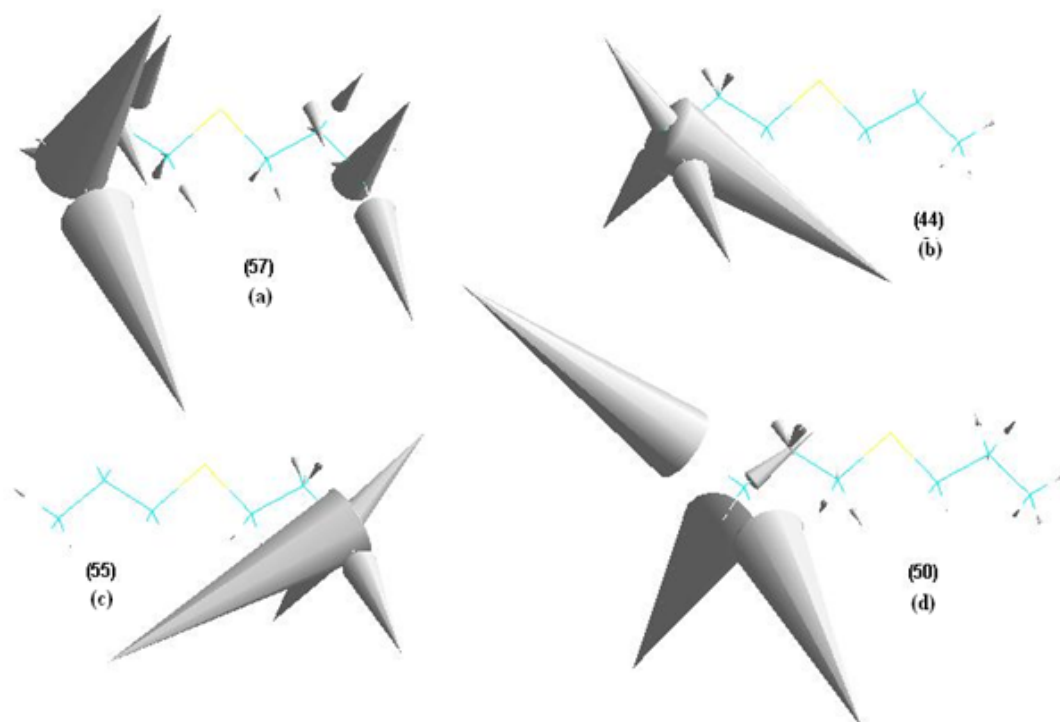
Figure 3.27: Some chosen vibrational modes with the corresponding fequencies for the DPS molecule (DFT calculation results)

.

3.3.2.3 Electronic Structure

The electronic structures of the DPS molecule studied were calculated with DFT using the 3-21G basis set with an exchange-correlation functional B3LYP. We have performed a single point calculation on the previously optimized structure using a DFT calculation with an STO-3G basis set.

In Fig. 3.21, we show the excess charge on the atoms of the molecule. Note that the total charge of the molecules is zero (neutral). As one would expect that the charge on the hydrogen atoms is always positive and is about 0.2 electron. The sulfur atom has a positive excess charge of about 0.3 electron. The carbon atoms which are respectively on the left side and right side of the sulfur atom has a negative charge accumulation of about -0.6. Other information we get from our calculations is the calculated energy values obtained by different methods (MM, PM3, AM1, DFT) of the system studied are given in Tables 3.4 -3.6, the

highest occupied and the lowest unoccupied molecular orbital energies (HOMO and LUMO, respectively), molecular orbital energy gap (HOMO-LUMO energy difference) with the calculated dipole moment value of the system considered are also given in Table 3.6 and the MO eigenvalue spectrum is shown in Fig. 3.28. 3D plots of HOMO and LUMO are presented in Fig. 3.29. According to MM method (with MM+ force field) the Van der Waals interactions have the largest contribution (2,922 kcal/mol) and dihedral interactions have the smallest contribution (0.008 kcal/mol). According to AM1 and PM3 calculations binding energy of the DPS molecule is about -1857 and -1849 kcal/mol, respectively. AM1 method gives about 8 kcal/mol more binding energy with respect to PM3 method. This difference is even more pronounced in heat of formation; AM1 method gives about -36 kcal/mol heat of formation, whereas PM3 method gives about -28 kcal/mol heat of formation. Both semi-empirical methods (PM3 and AM1) give similar stable structures and exothermic heat of formation. On the other hand, according to DFT calculation frontier molecular orbital energy gap, namely the HOMO-LUMO gap of the DPS molecule is about 7.2 eV.

The DPS molecule has a dipole moment value of about 1.7 Debyes, which is comparable with the dipole moment value of water (1.85 Debyes). Thus DPS can be considered as a polar molecule. This gives us some information about intermolecular forces and contribution to the ability of a substance to act as a solvent for ionic solids.

3D plots of charge density and electrostatic potentials obtained from DFT calculations are displayed in Fig. 3.30. In this figure charge distribution shows an interesting feature; the maximum positive excess charge accumulation is on the sulfur atom, on the other hand, the maximum negative excess charge accumulation appears on the carbon atoms bonded to sulfur atom. All carbon atoms have negative excess charge, varying from -0.4 to -0.6 electron charge. In electrostatic potential, the green regions show reaction sites of DPS molecule. Again according to DFT calculations, another interesting feature appears in the localization of HOMO and LUMO. Both HOMO and LUMO are localized mainly on sulfur atom and carbon atoms bonded to sulfur atom. This HOMO-LUMO localization on DPS molecule may show interesting spectroscopic properties, especially in electronic spectra.

Table 3.4: Some of the calculated energy values (in kcal/mol) of DPS molecule according to molecular mechanics method with MM+ force field.

Energy component	value	Energy component	value
Bond	0.283	Stretch-bend	0.095
Angle	0.595	Electrostatic	0.0
Dihedral	0.008	Total Energy	3.902
Vdw	2.922		

Table 3.5: Some of the calculated energy values (in kcal/mol) of DPS molecule according to AM1 and PM3 methods.

Quantity	AM1	PM3
Total Energy	-26674.353	-25691.041
Binding Energy	-1856.751	-1849.288
Isolated Atomic Energy	-24817.602	-23841.753
Electronic Energy	-121751.704	-118772.021
Core-Core Interaction	95077.351	93080.980
Heat of Formation	-35.583	-28.120

Table 3.6: Some of the calculated energy values (in kcal/mol) and dipole moment of DPS molecule according to DFT/B3LYP/MP2/3-21G method.

Quantity	value	Quantity	value
Total Energy	-396449.532	Lowest MO	-2400.496
Electronic Kinetic Energy	394352.066	HOMO (eV)	-5.763
eK, ee and eN Energy	-625843.809	LUMO (eV)	1.430
Nuclear Repulsion Energy	229394.277	Highest MO	60.520
MP2 Correlation Energy	-582.906	HOMO-LUMO gap (eV)	7.193
Total Energy (with MP2)	-397032.438	Dipole Moment (Debye)	1.696
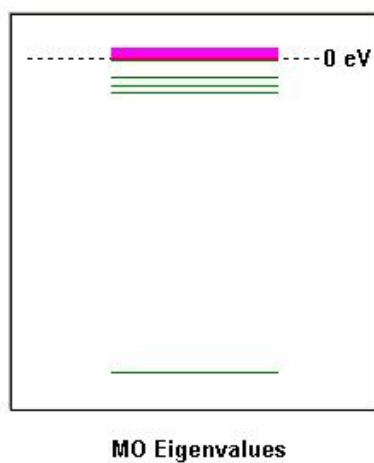
MO Eigenvalues

Figure 3.28: MO eigenvalues of the DPS molecule, DFT/B3LYP/MP2/3-21G results.

.
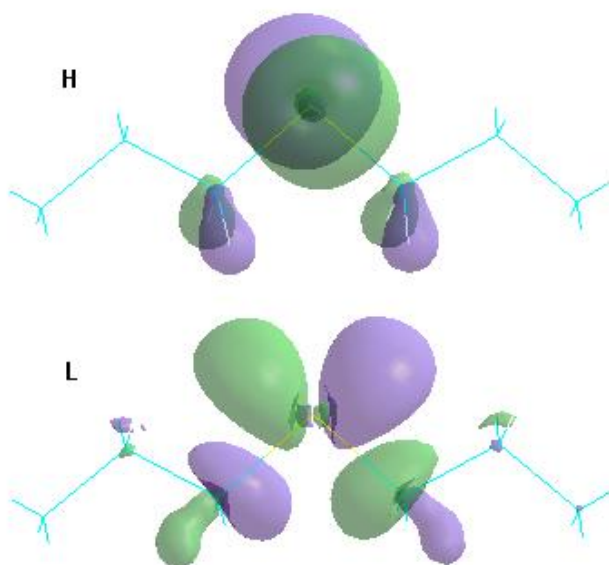
Figure 3.29: 3D HOMO and LUMO plots on the DPS molecule, DFT/B3LYP/MP2/3-21G results.

.
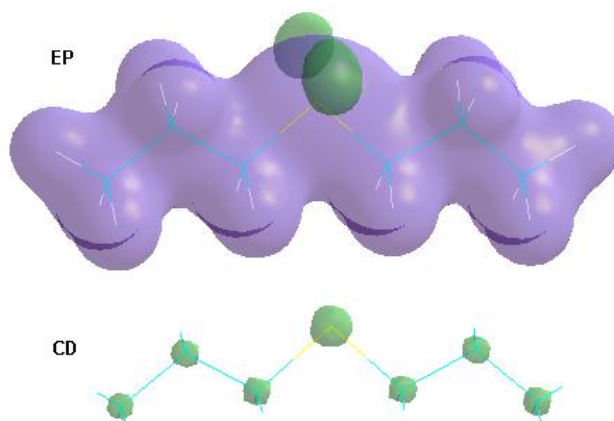
Figure 3.30: 3D charge distribution and electrostatic potential plots on the DPS molecule, DFT/B3LYP/MP2/3-21G results.

.

3.3.2.4 QSAR Properties

The other information is obtained from Quantitative Structure Activity Relationship (QSAR) values of the DPS molecule (see Table 3.7). QSAR tries to find viable relationship between biological activity and molecular properties, so that these features are used to evaluate the activity of new compounds [132]. The Log P value of DPS has been calculated as 2.08 at different level of computation. This value indicates moderate hydrophobicity for this molecule. (VdW: Van der Waals Surface; SAS: Solvent-Accessible Surface)

Table 3.7: QSAR values of DPS molecule at various levels of calculations.

Quantity	AM1	PM3	RHF	DFT
Log P	2.08	2.08	2.08	2.08
Surf. Area (App.)(VdW)	180.68	182.28	184.64	184.40
Surf. Area (App.)(SAS)	383.35	387.73	393.59	391.67
Surf. Area (Grid)(VdW)	172.32	173.00	175.03	177.23
Surf. Area (Grid)(SAS)	335.29	334.58	338.36	340.11
Volume (VdW)	131.53	131.83	132.63	133.63
Volume (SAS)	491.48	493.07	496.49	498.17
Hydration Energy (kcal/mol)	2.45	2.46	2.49	2.49
Refractivity	37.73	37.73	37.73	37.73
Polarizability	14.78	14.78	14.78	14.78
Mass (amu)	118.24	118.24	118.24	118.24

3.4 Cyfluthrin Molecule

3.4.1 Introduction

The structural, vibrational, electronic and QSAR properties of the cyfluthrin molecule in gas phase have been investigated theoretically by performing molecular mechanics method by using MM+ force field, and semi-empirical molecular orbital AM1 and PM3 calculations. The geometry of the molecule has been optimized, infrared spectrum (vibrational modes and intensities) and the electronic properties of the molecule have been calculated in its ground state. Geometry optimizations are carried out by using a conjugate gradient method (Polak-Ribiere algorithm). The SCF convergency is set to 0.001 kcal/mol and the RMS gradient is set to 0.001 kcal/(Å mol) in the calculations. We have performed all the calculations by using the HyperChem 7.5 packet program.

3.4.2 Results and Discussion

3.4.2.1 Structural Information

The closed formula of the cyfluthrin molecule is in the form $C_{22}H_{18}Cl_2FNO_3$ with a molecular point group C1. The geometry optimization has been performed successively by MM, AM1, and PM3 methods.

The optimized structures of cyfluthrin molecule with respect to both AM1 and PM3 methods are the same but MM method are not the same, there is torsion and bending on the leftmost ring. Also, most of the bond lengths and the bond angles are observed to change slightly. In Fig. 3.31, we present the optimized structure of the cyfluthrin molecule as resulting from a geometry optimization using the PM3 level of accuracy. Some of the molecular properties of the system considered are given in Table 3.8. The bond length between neighboring atoms of the cyfluthrin molecule is shown in Fig. 3.32. The actual coordinates of all the atoms of the molecule are given in Table A.13.
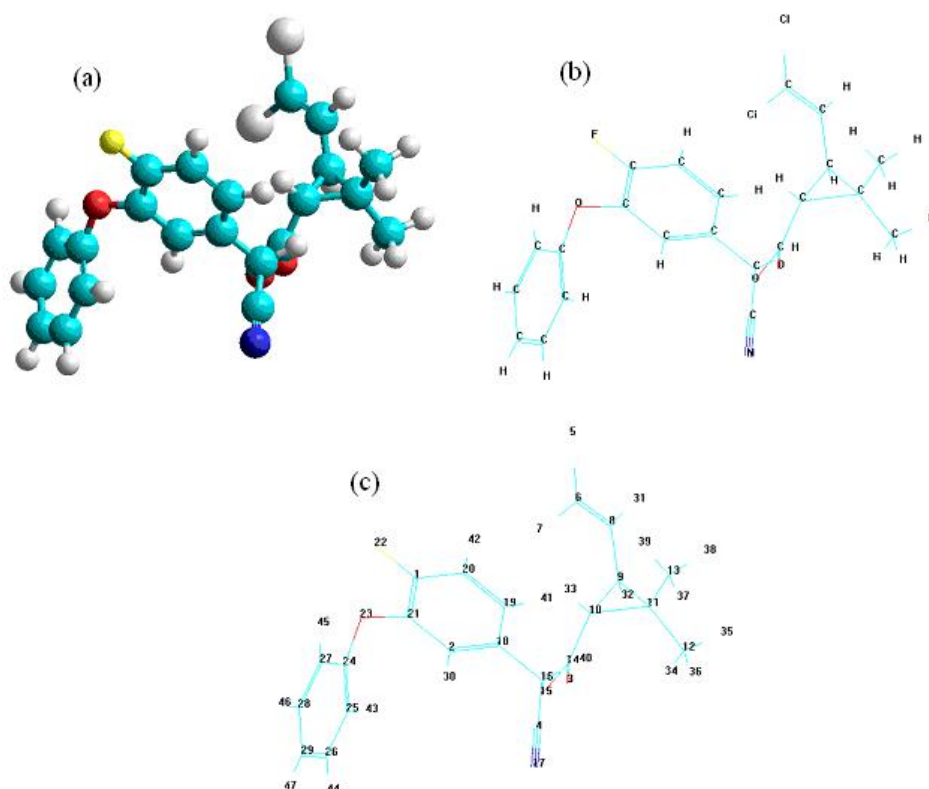
Figure 3.31: Optimized structure of cyfluthrin ($C_{22}H_{18}Cl_2FNO_3$) from PM3 calculation. Panel (a) shows the ball and stick model. Panel (b) shows atomic labels, and panel (c) shows the number labels of the atoms of the molecule.

Table 3.8: Some of the molecular properties of the cyfluthrin molecule in its ground state (singlet state) with zero total charge and C_1 symmetry according to AM1 and PM3 methods.

Quantity	AM1	PM3
Number of electrons	150	150
Number of doubly occupied levels	75	75
Number of Virtual orbitals	59	59
Number of Primitive Gaussians	134	134

62
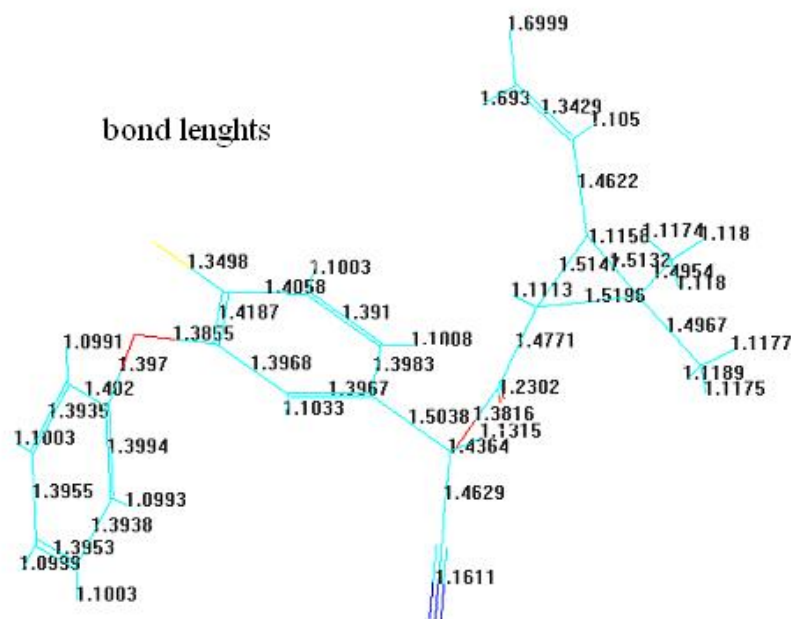
Figure 3.32: Bond lengths (Å) of the cyfluthrin molecule using PM3 method.

3.4.2.2 Vibrational Properties

The vibrational dynamics of the cyfluthrin molecule studied here were calculated using two semi-empirical MO-SCF methods (AM1 and PM3).

The calculated frequencies of the normal modes with first ten maximum intensities for each method of calculation in the harmonic approximation for cyfluthrin are presented in Table A.14 and the IR spectra of cyfluthrin are shown in Fig. 3.33 and Fig. 3.34. The scaling factors of the frequencies for the methods used here are 0.9532 for AM1 and 0.9761 for PM3. Unfortunately, the measured IR spectrum for cyfluthrin molecule is not available in the literature to compare our findings with. The cyfluthrin molecule contains 47 atoms and hence has 135 normal modes. We found that the vibrational spectrum can be divided in two bands: a low frequency band containing 116 modes and a high frequency band with 19 modes. Here we will present some of the calculated frequency values with their scaled values in parenthesis. The low frequency band for the AM1 method runs from 5.55 (5.29) cm^{-1} to 2072.36

(1975.37) cm^{-1}, while the high frequency band starts at 2557.55 (2437.86) cm^{-1} and ends at 3203.08 (3053.18) cm^{-1}. For the PM3 method, these boundaries are 3.39 (3.31) cm^{-1} to 2001.15 (1953.32) cm^{-1} and 2467.06 (2408.10) cm^{-1} to 3170.57 (3094.79) cm^{-1}.

Fig. 3.35 shows the displacements of some particular modes. In Fig. 3.35(a), the mode (corresponding the frequency of 974.77 cm^{-1}) represents the angle bending motion of C-C-C (12-11-13). In Fig. 3.35(b), the mode (corresponding the frequency of 2001.15 cm^{-1}) between the C-O (14-3) double bond shows the bond-stretching. In Fig. 3.35(c), the mode (corresponding the frequency of 3067.13 cm^{-1}) between C-H (20-42) and C-H (19-41) again show bond-stretching motion, but first one is more dominant than other. The mode (corresponding the frequency of 3157.08 cm^{-1}) seen in Fig. 3.35(d) represents the symmetric stretching motion of the C-H bonds on C13. Here we note that the stretching modes occur at higher frequency than bending. This behavior is consistent with a well established trend: stretching bonds require more energy than angle bending [133].
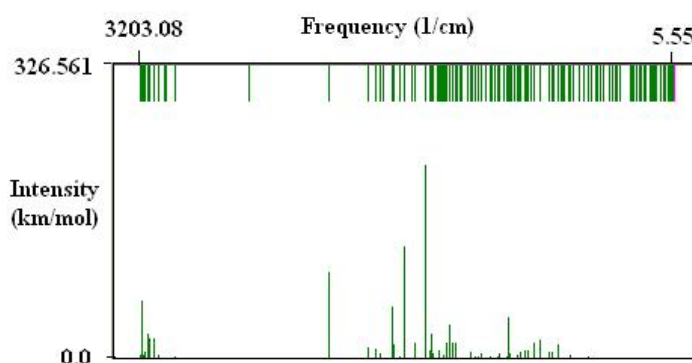
Figure 3.33: IR spectrum of cyfluthrin molecule at AM1 calculation.
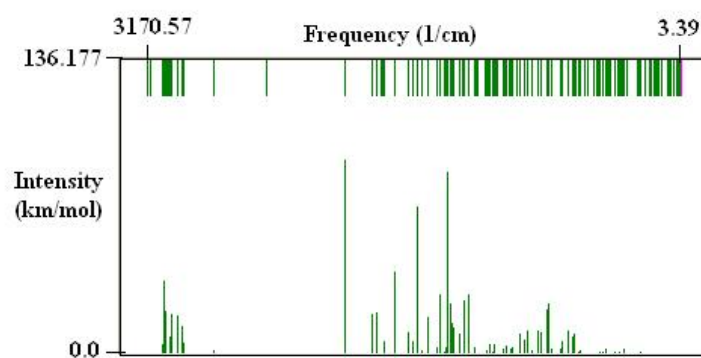
Figure 3.34: IR spectrum of cyfluthrin molecule at PM3 calculation.
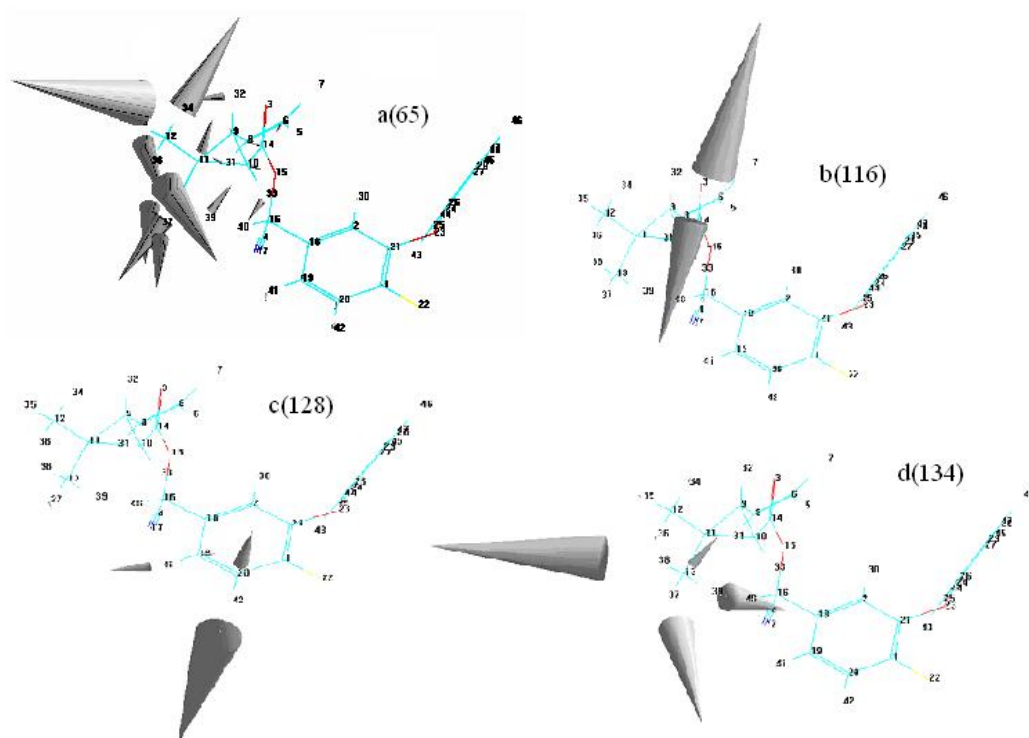
Figure 3.35: Some chosen vibrational modes (a-d) with the corresponding mode number for the cyfluthrin molecule (PM3 calculation results).

3.4.2.3 Electronic Structure

The excess charge on the atoms of the cyfluthrin molecule is given in Fig. 3.36. The total charge of the molecule is zero (neutral). In the molecule some of the carbon atoms have positive excess charge, some of them have negative excess charge; the magnitude of positive charges vary from 0.025 to 0.393, whereas the magnitude of negative charges vary from -0.060 to -0.225. As one would expect that the charge on the hydrogen atoms is always positive and is found to be 0.1, 0.2 electron. The chlorine atoms have positive charge of 0.097 and 0.121 electron, while fluorine and nitrogen atoms have negative charge of -0.076 and -0.028 electron, respectively. All the oxygen atoms have negative excess charge, their magnitude vary from -0.116 to -0.314. The large charge accumulation takes place on the oxygen atoms. Other information we get from our calculations is the calculated energy values obtained by different methods (MM, PM3, AM1) of the system studied are given in Tables 3.9 and 3.10, the highest occupied and the lowest unoccupied molecular orbital energies (HOMO and LUMO, respectively), molecular orbital energy gap (HOMO-LUMO energy difference) with the calculated dipole moment value of the system considered are also given in Table 3.10.

According to MM method (with MM+ force field) the Van der Waals interactions have the largest contribution (8.795 kcal/mol) and dihedral interactions have the smallest contribution (-4.221 kcal/mol). According to AM1 and PM3 calculations binding energy of the cyfluthrin molecule is about -5107 and -5114 kcal/mol, respectively. PM3 method gives about 7 kcal/mol more binding energy with respect to AM1 method. This difference is even more pronounced in heat of formation; AM1 method gives about -41 kcal/mol heat of formation, whereas PM3 method gives about -48 kcal/mol heat of formation. Both semi-empirical methods (PM3 and AM1) give similar stable structures and exothermic heat of formation. According to PM3 calculation frontier molecular orbital energy gap, namely the HOMO-LUMO gap of the cyfluthrin molecule is about 9 eV. This relatively large value indicates that cyfuthrin molecule could be kinetically stable, and the exothermic nature of heat of formation makes this molecule thermodynamically stable [134, 135].

The dipole moment value for cyfluthrin is found to be about 2.8 Debyes. Calculated dipole moment value of the molecule studied is not small, comparing with water value. Same level of calculation gives the dipole moment value of water as 1.739 Debyes (experimental value of dipole moment for water is about 1.85 Debyes). The difference in calculated and experimental dipole moment value for this molecule studied is larger than order of water. Thus cyfluthrin

molecule can be considered as a polar molecule.

The MO eigenvalue spectrum is shown in Fig. 3.37. The 3D plots of the HOMO and LUMO obtained from PM3 calculations are displayed in Fig. 3.38. The HOMO mainly localized on top right part of the molecule including chlorine atoms, while the LUMO mainly localized on middle ring bonded to fluorine atom. This HOMO-LUMO localization on cyfluthrin molecule may show interesting spectroscopic properties, especially in electronic spectra. Again according to PM3 calculation, 3D plots of charge density (CD) and electrostatic potentials (ESP) are displayed in Fig. 3.39. CD plot shows a uniform distribution. However, ESP in the immediate vicinity of N(17) and O(3) show a changing gradient, the rest of the molecule do not exhibit a significant alteration.
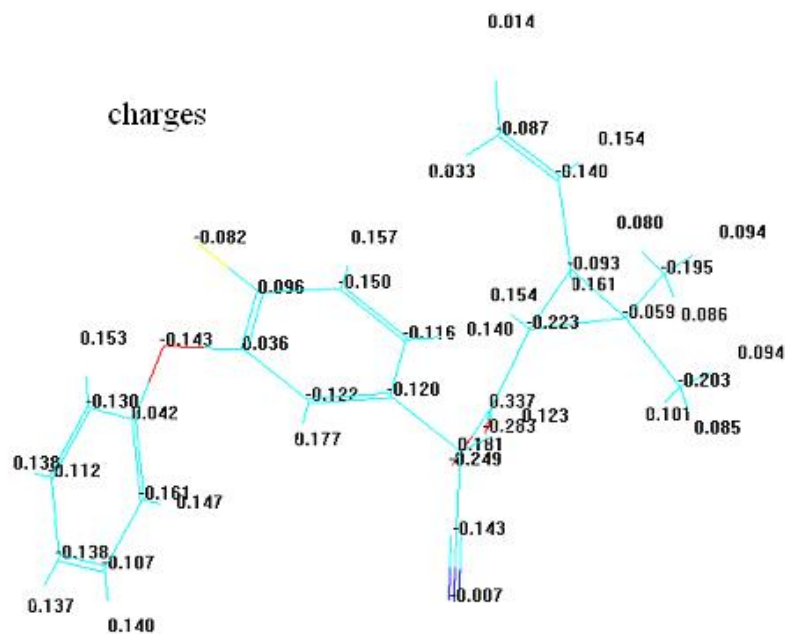
Figure 3.36: Excess charge (in units of electron charge e) on atoms of the cyfluthrin molecule using PM3 method.

Table 3.9: Some of the calculated energy values (in kcal/mol) of cyfluthrin molecule according to molecular mechanics method with MM+ force field.

Energy component	value	Energy component	value
Bond	1.292	Stretch-bend	-0.057
Angle	7.390	Electrostatic	4.252
Dihedral	-4.221	Total Energy	17.451
Vdw	8.795		

Table 3.10: Some of the calculated energy values (in kcal/mol) and dipole moment of cyfluthrin molecule according to AM1 and PM3 methods.

Quantity	AM1	PM3
Total Energy	-125958.076	-115252.061
Binding Energy	-5107.482	-5114.546
Isolated Atomic Energy	-120850.594	-110137.514
Electronic Energy	-926755.469	-902410.382
Core-Core Interaction	800797.394	787158.321
Heat of Formation	-41.519	-48.583
Lowest MO	-49.948	-57.171
HOMO (eV)	-9.272	-9.701
LUMO (eV)	-0.587	-0.660
Highest MO	6.272	6.638
HOMO-LUMO gap (eV)	8.685	9.041
Dipole moment (Debye)	3.661	2.756
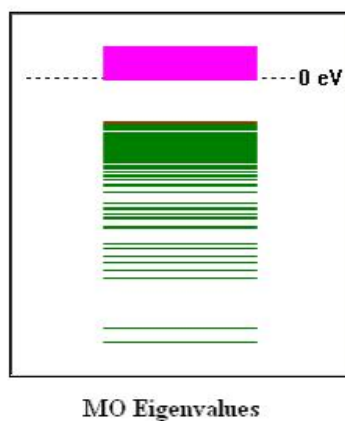
MO Eigenvalues

Figure 3.37: MO eigenvalues of the cyfluthrin molecule, PM3 result. The values of the lowest and the highest MO eigenvalues are given in Table 3.10 for the scale.
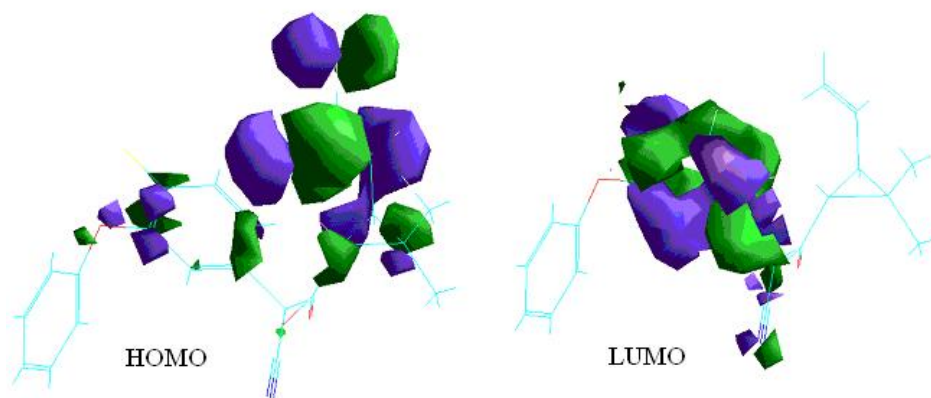
Figure 3.38: 3D HOMO and LUMO plots on the cyfluthrin molecule, PM3 results.
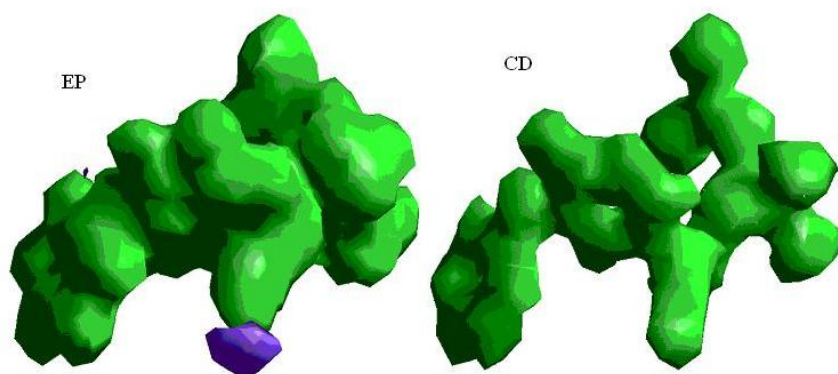
Figure 3.39: 3D electrostatic potential (EP) and charge density (CD) plots on the cyfluthrin molecule, PM3 results.

3.4.2.4 QSAR Properties

More details about the cyfluthrin molecule is obtained from Quantitative Structure Activity Relationship (QSAR) values (see Table 3.11). A QSAR help to find a mathematical relationship between biological activity or chemical reactivity of the molecular system and its molecular properties, so that these features are used to evaluate the activity of new compounds [132]. The Log P value of cyfluthrin has been calculated as 2.63 at AM1 and PM3 level of computation. This value indicates moderate hydrophobicity for this molecule. Here

we should point out that in the calculation of hydration energy the package program used did not take into account C4 atom.

Table 3.11: QSAR values of cyfluthrin molecule at two levels of calculations.

Quantity	AM1/PM3
Log P	2.63
Surf. Area (App.)(VdW) (Å^2)	434.54
Surf. Area (App.)(SAS) (Å^2)	628.13
Surf. Area (Grid)(VdW) (Å^2)	425.04
Surf. Area (Grid)(SAS) (Å^2)	672.56
Volume (VdW) (Å^3)	365.96
Volume (SAS) (Å^3)	1150.05
Hydration Energy (kcal/mol)	-8.80
Refractivity (Å^3)	118.24
Polarizability (Å^3)	42.62
Mass (amu)	434.29

3.5 Short Beta Strand Peptide (DDATKTFT)

3.5.1 Introduction

Unfolding of the 8-residue peptide (DDATKTFT) of Immunoglobulin G-Binding Protein G was studied at five different temperatures by Molecular Dynamics: 290, 300, 325, 430, and 500 K. We also performed simulations at 300 K with the 6 and 7 residue versions DDATKT and DDATKTF. The initial conformation of the peptide was taken from the protein data bank (PDB code: 2OED, residues 46 to 51, 52 or 53 respectively) [56]. Unfolding simulations were performed using GROMACS code [136] with the united atom GROMOS96 43a1 force field [137]. The peptide was confined into a cubic box of SPC (simple point charge) water molecules [138]. Simulations use periodic boundary conditions in all three directions, a 2 fs timestep. The energy of the system was initially minimized using steepest descent [125]. After the minimization, the resulting system was heated to five different temperatures independently, keeping pressure and temperature fixed (NPT ensemble) using a Berendsen thermostat [90] with coupling time constant of 0.1 ps and an anisotropic Berendsen barostat [90] with a coupling constant of 0.2 ps with reference pressure 1 bar in all directions and compressibility 1.12×10^{-6} bar^{-1}. Bond lengths were constrained by the linear constraint solver (LINCS) al-

gorithm [92]. The simulations used the particle-mesh-ewald (PME) technique [139, 140] for the long-range electrostatics with a 0.12 nm Fourier grid spacing and a 1.2 nm cut-off radius. The shift scheme was used to compute the short range van der Waals interactions with a 1nm cut-off radius. The neighborlist was updated at every step. The Newton's equations of motion were numerically integrated using leap-frog algorithm [141]. The Visual Molecular Dynamics (VMD) software program [142] was used for snapshots. The simulations were carried out 9 ns for 290 K and 20 ns for the others.

3.5.2   Results and Discussion

3.5.2.1   General Characterization

We first analyze the immunoglobulin peptide with eight residues. The RMSD of the backbone atoms was calculated with the native PDB structure (2OED) as reference. The RMSD plots of the peptide as a function of time at different temperatures are shown in Fig. A.1 in Appendix A. For 290 K and 300 K, the peptide remained native-like for several hundred picoseconds and then unfolded and continued to fluctuate. For 325 K and above the peptide unfolded in less than a hundred ps. Maximal conformational changes are found for 430 and 500 K where RMSD values exceed 0.5 nm. During the simulations, the peptide unfolded and refolded several times. A backbone RMSD of 0.2 nm or more, over a significant time is considered unfolded. Our RMSD values are on the average above 0.2 nm, but they fluctuate. Our structure is best classified as unstable. A high RMSD is expected [143] for such peptides due to their flexibility and their tendency to deviate from the crystal structure in water, hence they are unstable. Our result confirms this expectation. Tsai and Levitt [144] ran 18 simulations (for 10 ns at 298 K) of two C-terminal beta-hairpins from B1 domain of protein-G (GEWTY-DDATKTFTVTE and GEWTYDDATKTFTVT) for 16 and 15 residues. They classified their results in 3 groups (stable, unstable, unfolded) for 16 residues and in two groups (unstable, unfolded) for 15 residues. We find unstable structures at all temperatures.

It has been recognized in the literature that hydrogen bonding plays an essential role in the stability of the secondary structure of proteins [145, 146, 147]. Also the effect of the number of hydrogen bonds (H-bonds) on the stability of the protein was explored [148, 149] as well as

the length of H-bonds [150] and H-bond energies [151]. We find that the locations of the H-bonds are also crucial for stability. Munekata et al. [61] found that the 6-mer variant DDATKT forms a rigid loop (stable structure). But the 8mer, DDATKTFT, turns out to be unstable (i.e. does not form a specific conformation). To understand the reason as to why, we investigated at 300 K for 6 ns each DDATKT, DDATKTF and DDATKTFT. From the RMSD versus time Fig. 3.40, one observes that DDATKT has a stable structure in agreement with Munekata et al. The other two cases are not stable, that is they unfold and refold several times during the simulation. We analyzed the H-bonds using VMD during the entire simulation (Fig. 3.41, where we also show the H-bonds in the native states). An H-bond is considered formed when the donor and acceptor distance is below 0.4 nm and the donor-acceptor-hydrogen angle is below 40 degrees. Most of the time, an H-bond exists between the amide hydrogen of D1 and carbonyl oxygen of T4 or the carbonyl oxygen of D1 and of amide hydrogen T6. But for DDATKTF and DDATKTFT either no H-bond is formed or an H-bond is formed between carbonyl oxygen of T4 and amide hydrogen of T6 and between carbonyl oxygen of D2 and amide hydrogen of K5. We show the timeseries of the relevant hydrogen bonds in Supplementary Information (Fig. A.2).

From these results, we conclude that the H-bonds between D1-T4 and D1-T6 contribute to the stability of the peptide. This result agrees with Munekata el al. who found that mutations on the residues of D1, T4 and K5 where there is an H-bond network destabilize the peptide. In the native state for all cases, there are three hydrogen bonds between amide hydrogens and carbonyl oxygens, D1-T6 (both ways) and the amide hydrogen of K5 with the carbonyl of D2. Comparing we find for DDATKT that during the simulation even though the amide hydrogen of D1 and carbonyl oxygen of T6 and the amide hydrogen of K5 and carbonyl oxygen of D2 bonds are broken, we still have D1-T4 or D1-T6 bonds and the peptide is stable. For DDATKTF and DDATKTFT the only remaining bond is the amide hydrogen of K5 to the carbonyl oxygen of D2 and the peptides are not stable. This again shows the role played by the D1-T6 bond in the stability. It also appears that K5-D2 plays at best a minor role. Some of the native H-bonds appear to be essential and are indeed maintained during simulation, but the DSSP (definition of secondary structure) [152] data do not reflect this finding and imply a less well defined conformational behavior. Thus looking additionally at the modal behavior could help to resolve this discrepancy.

The distance between C of the first and the last residue of the 8-mer peptide at three different temperatures are shown in Fig. 3.42. This distance changes with time for all temperatures

of the peptide. However, at 430 and 500 K, distance fluctuations are larger than at the other temperatures consistent with the RMSDs. Visual comparisons between the native structure and snapshots at different temperatures can be found in Supplementary Information (Figures A.3 and A.4).

The secondary structure of the peptide was studied using DSSP analysis. In Fig. 3.43, the secondary structure analysis of DDATKTFT peptide is shown for all temperatures. At each temperature, residues mainly take coil, bend or turn structures. Besides, at 300K, 325 K, and 430K alpha-helix structure appears; moreover at 325K and 500K 3-helix structure appears. For 325 K, at 6 ns residues 3-7 and at 13 ns residues 1-4 adopt an alpha-helix structure. At 500 K at 11 ns residues 5-7 adopt a 3-helix structure. One would expect that at lower temperatures (290 or 300 K) the peptide should retain its turn-like structure, as taken from the PDB, and at higher temperatures move away from it. But, these DSSP results provide us with another proof that the peptide does not maintain a specific conformation. For this protein we also studied a salt bridge (see Fig. A.5) between D2 and K5. Comparing with the stability we find that the weaker the salt bridge is the more stable is our protein. This analysis agrees with Hendsch et al [153] but disagrees with Tsai and Levitt [144].
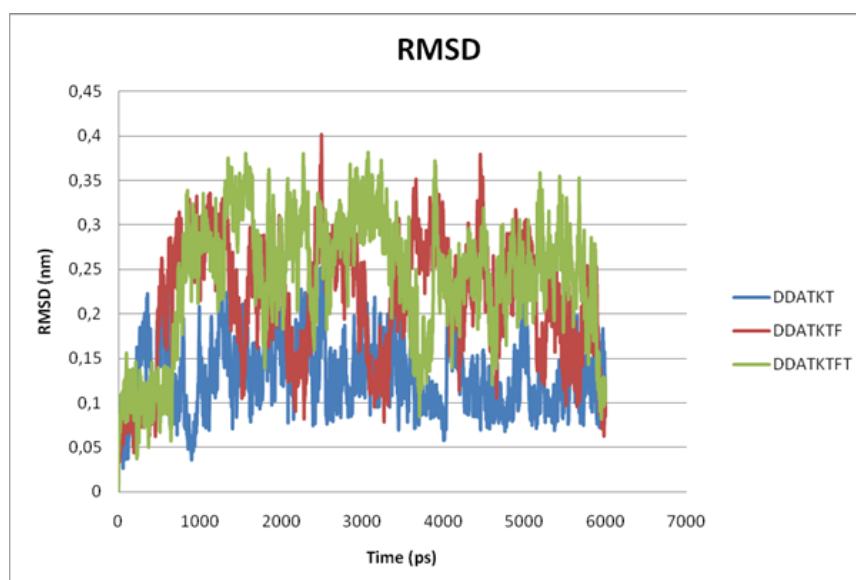
Figure 3.40: RMSDs at 300K for the different Protein G sequences
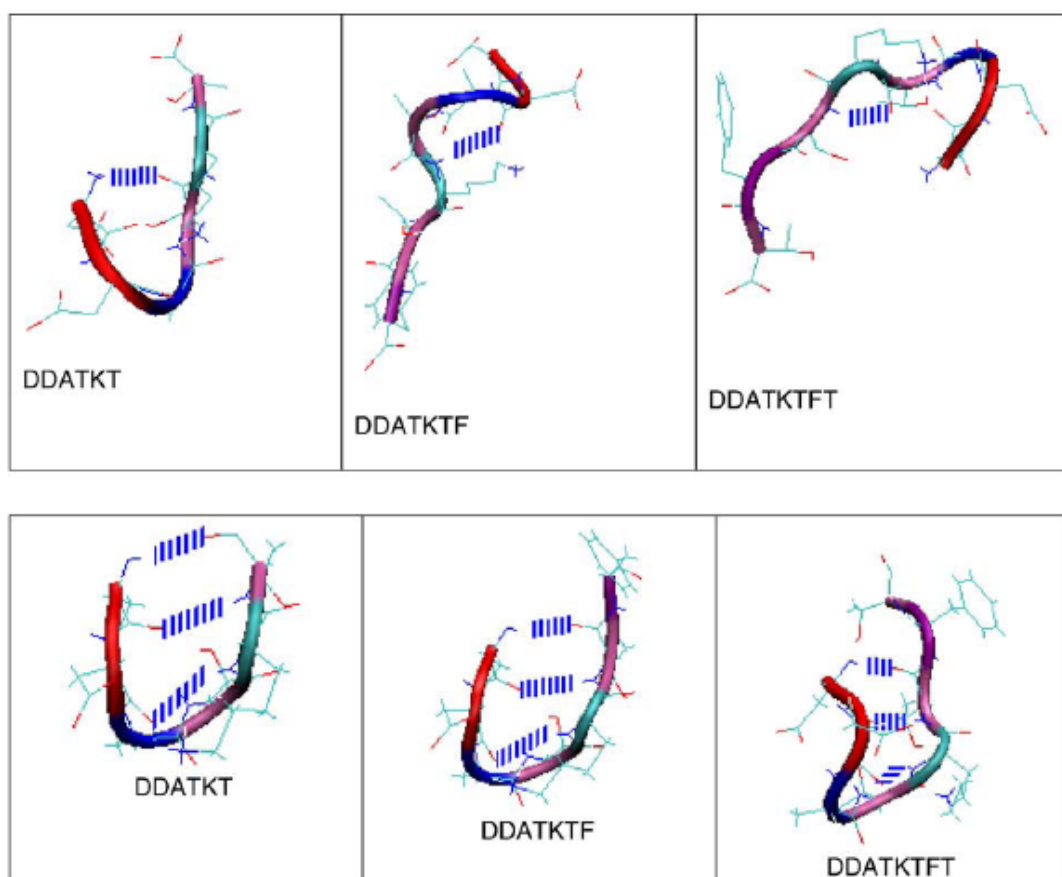
Figure 3.41: Hydrogen bonding in DDATKTFT and its variants. Top line our simulation, bottom: native states
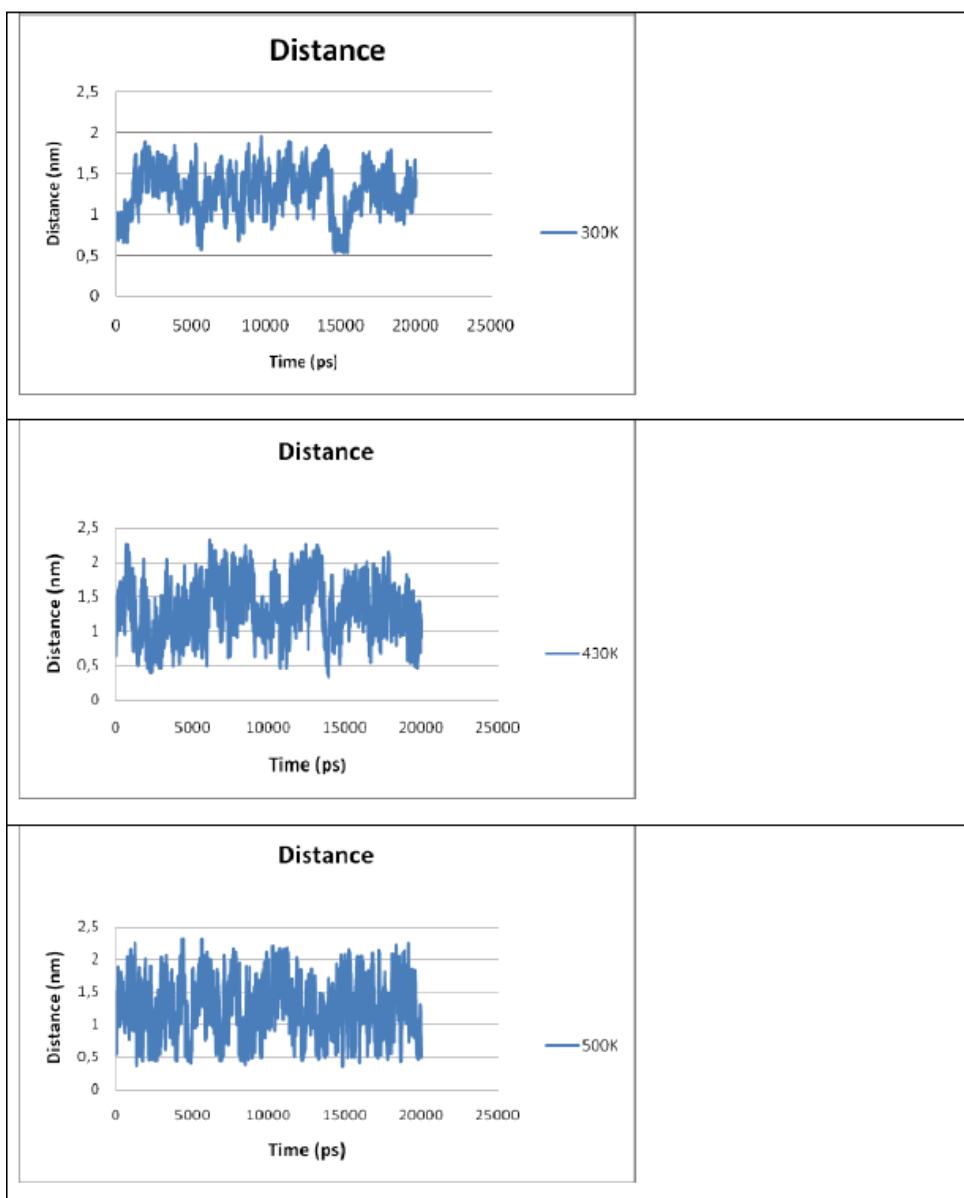
Figure 3.42: Distance between C_α atoms of the first and the last residues of the immunoglob-ulin peptide at different temperatures.
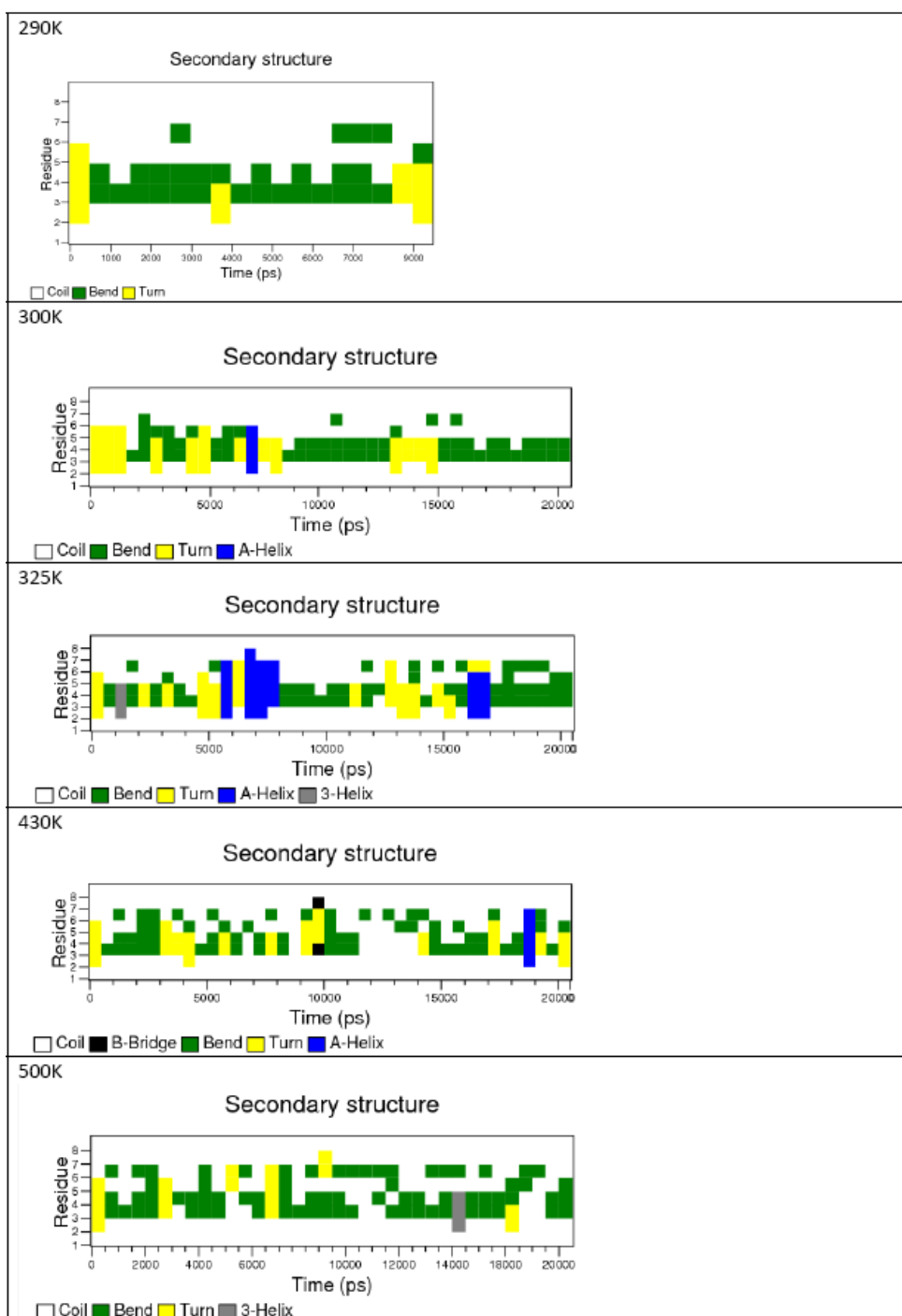
Figure 3.43: Time evolution of the secondary structure of the DDATKTFT peptide at different temperatures.

3.5.2.2 Principal Component Analysis

We used the Karhunen-Loeve (KL) version of Principal Component Analysis (PCA) to analyze the unfolding of the DDATKTFT peptide. The details of PCA are given in the Appendix B.

In Fig. 3.44 the scaled eigenvalues of the covariance matrix and the percent variance are shown for five different temperatures. The descending curves refer to the former and the ascending ones to the latter. The first 2 out of a total of 8 modes constitute more than 85% of the variance for 290 K and 300 K and decreasing at higher temperatures. These first 2 modes dominate the unfolding dynamics.

Figures 3.45 and 3.46 depict the first three spatial eigenvalues and their time-dependence of the modal amplitudes for the lowest and highest temperatures, while the others are shown in Appendix A (see Figures A.6, A.7 and A.8). We see that the first modal amplitude is intermittently active whereas the higher modes are more constantly active throughout the simulation, although their influence is diminished. Considering the spatial eigenvectors, we can discern which residues contribute to the fundamental motion in each mode. We observe that the first mode is essentially active along the molecular axis and represents an unraveling mode which can be depicted schematically as the rightmost diagram in Fig. 3.47. Similarly, the second mode represents a butterfly motion in the plane of the molecule as the activity shifts to the y-plane (middle diagram in Fig. 3.47). The third mode (at least at the lowest temperature) appears to become two dimensional with the most activity along the center of mass of the hairpin whereas the turn and the ends remain relatively stationary (leftmost diagram in Fig. 3.47).

If we compare the 8-mer to its shorter neighbors at 300 K we find that the 6- and 7-mer are quite close to each other with all the directions separated (Fig. 3.48). For all three, the first mode is the same but for the 8-mer the second and third modes combine the two short directions. The first mode is always unraveling; the second for the shorter ones is a shearing motion of the open ends against the turn and the third is a weak correction out of plane. The folding propensity decreases with length for these three peptides which correlates with increasing importance of modes 2 and 3.
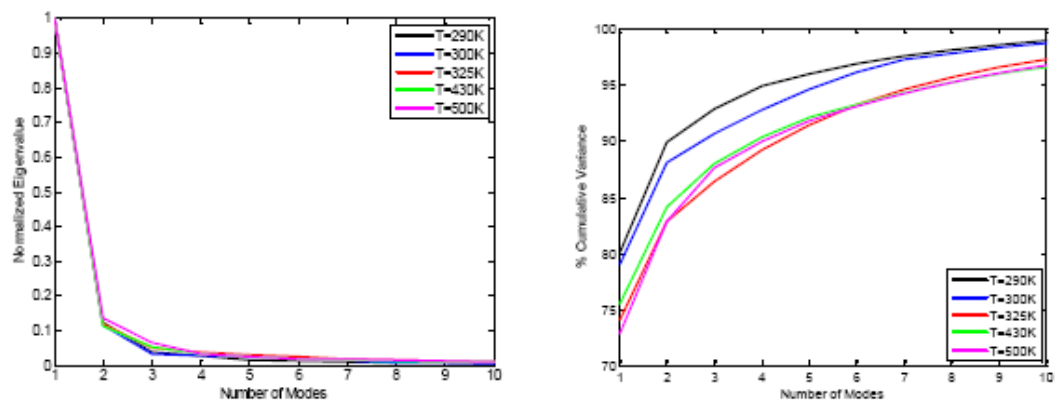
Figure 3.44: Scaled eigenvalues (left) for DDATKTFT and percent variance (right) for five different temperatures.
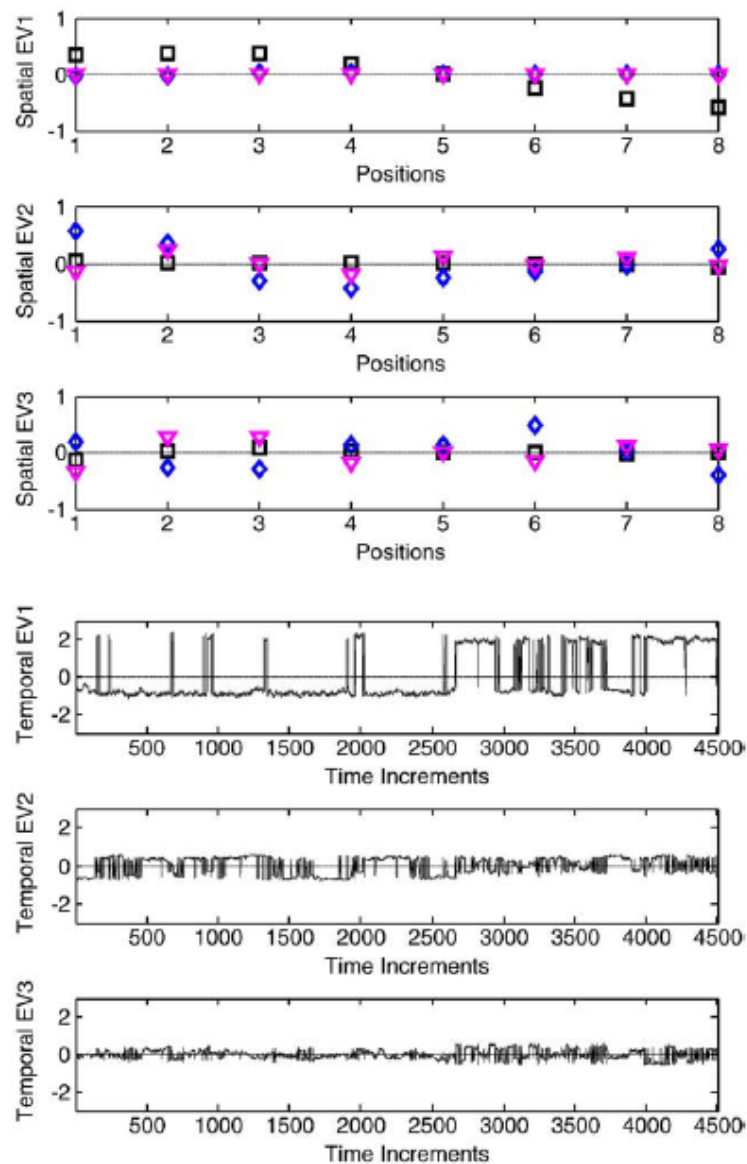
Figure 3.45: First three spatial eigenvalues based on the C-alpha atoms at 290K together with modal amplitudes (protein G unfolding). The x direction is indicated by squares, y by diamonds and z by triangles
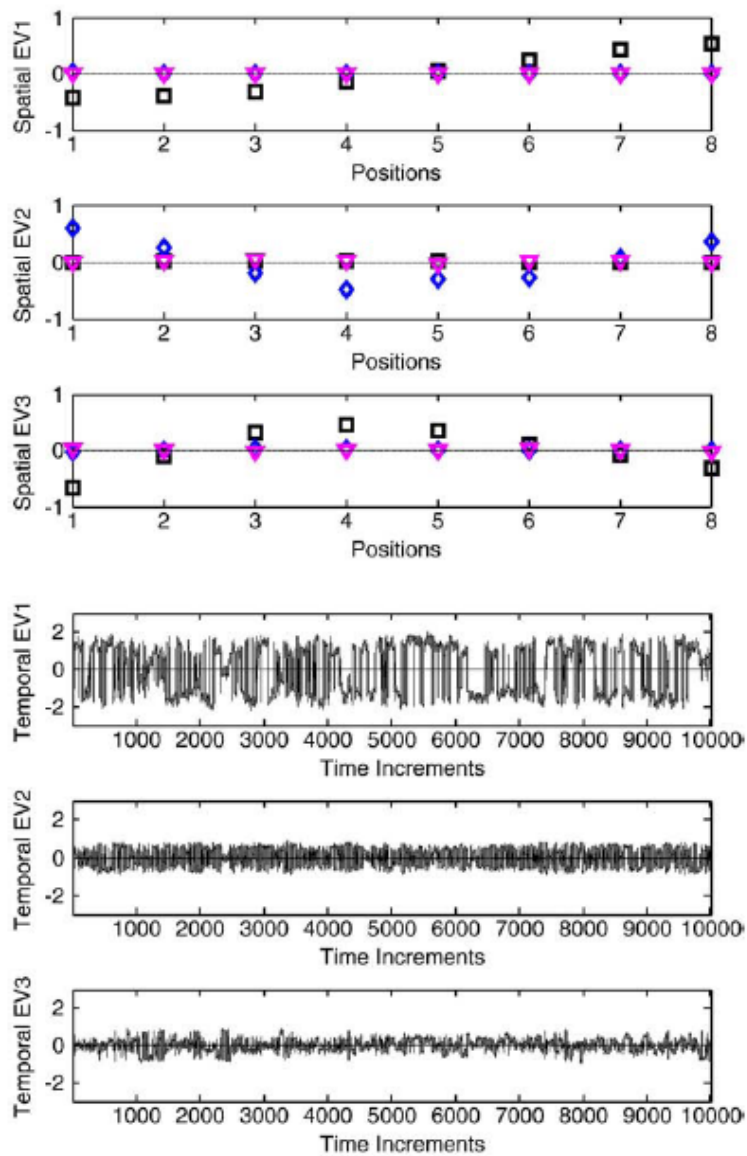
Figure 3.46: First three spatial eigenvalues based on the C-alpha atoms at 500 K together with modal amplitudes (protein G unfolding).
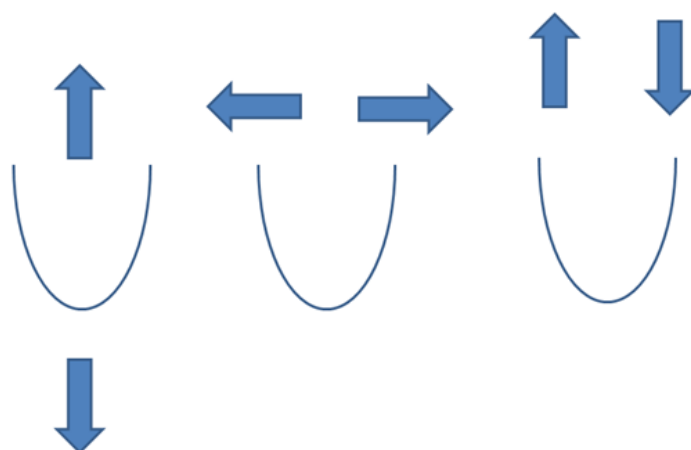
Figure 3.47: Sketch of important eigenmodes of protein G
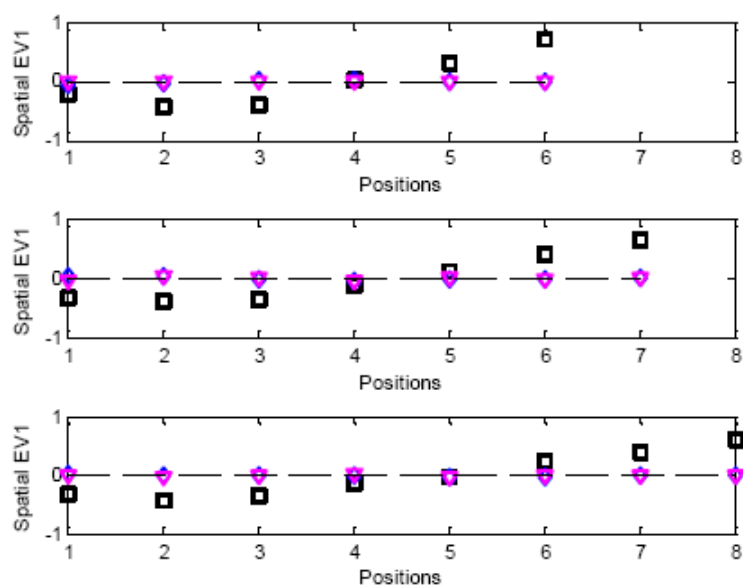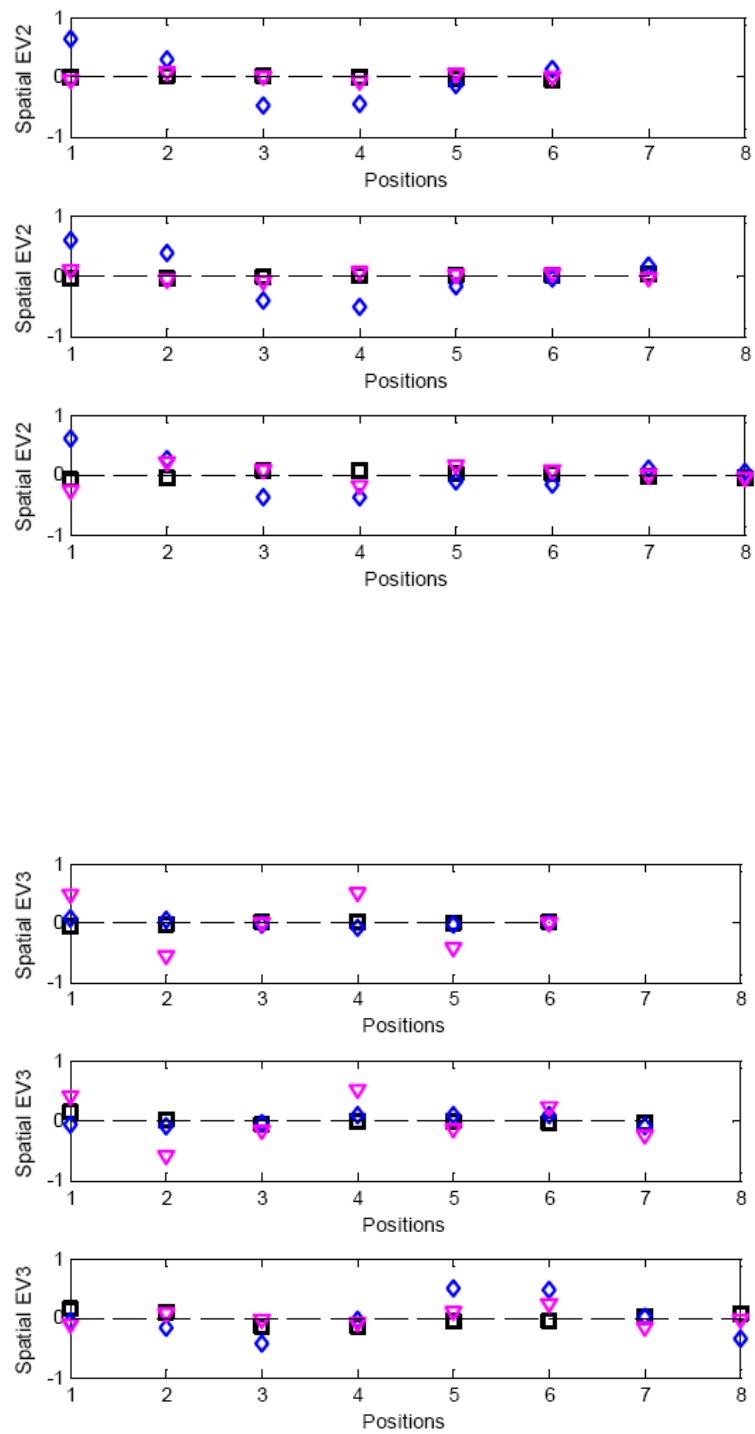
Figure 3.48: Comparison of the unfolding modes of proteins G peptides of different lengths

CHAPTER 4

CONCLUSION

In this thesis, using molecular mechanics, semi-empirical methods, ab-initio RHF and density functional theory, we have studied the SAC, SAMC, allicin, MPD and AMS molecules which are found in onion and garlic and are beneficial molecules for human health. We have also studied a molecule called cyfluthrin which is an insecticide, with molecular mechanics and semi-empirical methods. We have also studied, as part of the protein folding and unfolding problem, the unfolding of the small peptide DDATKTFT and its variants where the last two residues are changed. Below is a brief summary of our findings.

Theoretical investigations will aid in the elucidation of antiplatelet aggregation and antioxidant activity and clarification of the uncertainty about the conditions under which biologically active sulfur-containing compound such as SAC , SAMC, allicin, MPD, AMS and DPS or other potential bioactive agents before they can be safely evaluated and commercially developed as beneficial therapeutic agents. DPS is soluble in lipid; however other five molecules are soluble in water. This important chemical property certainly originates from conformational and electronic structures of the molecules and substituent groups.The optimized structures of the SAC , SAMC, allicin, MPD and AMS molecules in gas phase are 3D; the carbon and sulfur containing skeleton of the molecules are not linear, they are bended (curved). Sulfur atom is playing an important role in this feature; bending usually takes place at sulfur positions. But in the case of DPS, the carbon and sulfur containing skeleton of the molecule is nearly planar and symmetric with respect to the sulfur atom. Calculated IR spectra of MPD and AMS qualitatively agree with experimental spectra [131], there is no experimental IR data for the rest of the molecules. However, calculated IR spectra and vibrational modes of these

molecules are consistent qualitatively between each other. It is worth to mention that when comparing the calculated vibrational frequency values obtained from three different methods (four different methods for DPS), there is no systematic correlation among the methods for a particular mode. This situation could be due to the parameterization of the semi-empirical methods. From electronic properties point of view one of the remarkable features is that the frontier molecular orbital energy separations, namely the HOMO-LUMO separations, of the SAC and SAMC molecules are close to each other, SAC has a gap of about 5.5 eV, whereas SAMC has a gap of about 5.0 eV. This (relatively) small difference in gaps should show similar features in the electronic spectra of these molecules. But, this difference between allicin, MPD and AMS is quite different. Allicin has a gap of about 11 eV, whereas MPD has a gap of about 6 eV, and AMS has the value of about 3 eV for the gap. HOMO-LUMO separation for DPS is 7.2 eV. The HOMO-LUMO difference can be used as a rough indicator of kinetic stability [134, 135, 154, 155]. A kinetically stable molecule resists to chemical reactions. Small or no HOMO-LUMO gap means, a molecule is chemically reactive. Therefore, we can conclude that while the AMS molecule is the most chemically reactive molecule, allicin molecule is the most kinetically stable one. Dipole moment of SAMC (4.68 D) is relatively larger than that of SAC (2.76 D). The dipole moment values for the other molecules are found to be: 2.45 D for allicin, 2.58 D for MPD and 2.19 D for AMS. And the dipole moment of the DPS molecule is 1.7 D. Thus SAMC, having the largest dipole moment and so being the most polar molecule among the ones we studied, is expected to be more reactive in the aqueous environment of the cell than the others.

Theoretical work, such us the one we presented here, will help clarify the uncertainty about cyfluthrin molecule before it can be safely used as a synthetic cyano-containing pyrethroid insecticide. The striking features of the molecules that we have found may be summarized as follows: In the optimized structure of the molecule, the skeleton of the molecule is not planar. Cyfluthrin molecule generally shows a stretching motion. Studying the electronic properties revealed that a large charge accumulation takes place on the oxygen atoms. We calculated that cyluthrin is a polar molecule with a dipole moment of 2.8 D. The HOMO-LUMO difference has a gap value of about 9 eV, indicating that this molecule is kinetically stable. Furthermore the reactivity of molecules can be controlled by this (relatively) large difference in gap. On the other hand, the heat of formation of cyfluthrin molecule is exothermic, which shows that

this molecule is thermodynamically stable. Although charge density distribution of molecule looks uniform, electrostatic potential shows different characteristic in the molecule. QSAR properties has been also used extensively in health sciences and toxicology.

In the final part of the thesis, we have studied the short peptide (DDATKTFT) under a variety of conditions. Both the RMSD and especially the principal components can point out some significant differences. We find a destabilization of the peptide with increasing length which is associated with a change in the normal modes. Although we must not infer an immediate predictive connection between the sequence of a peptide and its normal modes, PCA may certainly help in identifying critical events during the folding or unfolding process and their exact location within a peptide sequence. We also find a correlation between hydrogen bonds and the principal components. This had been pointed out earlier by Suzuki et al. [55]. In conclusion, a combined view of structural as well as modal information yields more reliable information about the (dis)appearance of well defined secondary structure elements during the simulation. Thus, a divide-and-conquer strategy as presented by Ho and Dill [57] (i.e. splitting the entire protein into segments small enough to fold into their favored conformation quickly and using well the folding peptides as originators for propagating the folding process when regrowing the whole protein sequence again) might benefit from the support PCA may give in reliably identifying key folding events. In our work, unfolding dynamics of DDATK-TFT was investigated using MD simulations at various temperatures. We have also performed three more simulations for DDATKT, DDATKTF and DDATKTFT to understand the effect of how various parts play role on the stability (or instability) of the peptide. We have found that the location of the H-bonds is important for the stability of the peptide. Our findings are consistent with some experimental results in certain cases. Our results reveal some information about structure-stability relationship which may help for the de novo peptide design.

REFERENCES

[1] Watve M.G., Tickoo R., Jog M.M., and B.D. Bhole. How many antibiotics are produced by the genus streptomyces? *Arch. Microbiol.*, 176:386–390, 2001.

[2] Baltz R.H. Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News*, 55:186–196, 2005.

[3] Wang J., Soisson S.M, Young K., Shoop W., Kodali S., Galgoci A., Painter R., Parthasarathy G., Tang Y.S, Cummings R., Ha S., Dorso K., Motyl M., Jayasuriya H., Ondeyka J., Herath K., Zhang C., Hernandez L., Allocco J., Basilio A., Tormo J.R, Genilloud O., Vicente F., Pelaez F., Colwell J., Lee S.H., Michael B., Felcetto T., Gill C., Silver L.L, Hermes J.D, Bartizal K., Barrett J., Schmatz D., Becker J.W, Cully D., and Singh S.B. Platensimycin is a selective fabf inhibitor with potent antibiotic properties. *Nature*, 441:358–361, 2006.

[4] Freiberg C., Brunner N.A, Schiffer G., Lampe T., Pohlmann J., Brands M., Raabe M., Häbich D., and Ziegelbauer K. Identification and characterization of the first class of potent bacterial acetyl-coa carboxylase inhibitors with antibacterial activity. *J. Biol. Chem.*, 279:26066–26073, 2004.

[5] Holtzel A., Schmid D. G., Nicholson G. J., Stevanovic S., Schimana J., Gebhardt K., Fiedler H. P., and Jung G. Arylomycins a and b, new biaryl-bridged lipopeptide antibiotics produced by streptomyces sp. tu 6075. ii. structure elucidation. *J. Antibiot. (Tokyo)*, 55:571–577, 2002.

[6] Johnson J. J and Mukhtar H. Curcumin for chemoprevention of colon cancer. *Cancer Lett*, 255:170–181, 2007.

[7] Shishodia S., Chaturvedi M. M., and Aggarwal B. B. Role of curcumin in cancer therapy. *Curr. Prob. Cancer*, 31:243–305, 2007.

[8] Subrahmanyam D., Renukaa B., Raoa C.V.L., Sagarb S.P., Deevib D.S., Babuc J.M., and Vyasc K. Novel d-ring analogues of podophyllotoxin as potent anti-cancer agents. *Bioorg. Med. Chem. Lett.*, 8:1391–1396, 1998.

[9] Utsugi T., Shibata J., Sugimoto Y., Aoyagi K., Wierzba K., Kobunai T., Terada T., Oh hara T., Tsuruo T., and Yamada Y. Antitumor activity of a novel podophyllotoxin derivative (top-53) against lung cancer and lung metastatic cancer. *Cancer Res.*, 56:2809–2814, 1996.

[10] Jordan M.A. and Wilson L. Microtubules as target for anticancer drugs. *Nat. Rev. Cancer*, 4:253–265, 2004.

[11] You Y. Podophyllotoxin derivatives: current synthetic approaches for new anticancer agents. *Curr. Pharm. Design*, 11:1695–1717, 2005.

[12] Kuznetsova L., Chen J., Sun L., Wu X., Pepe A., Veith J.M., Pera P., Bernacki R.J., and Ojima I. Syntheses and evaluation of novel fatty acid second generation taxoid conjugates as promising anticancer agents. *Bioorg. Med. Chem. Lett.*, 16:974–977, 2006.

[13] Fito M., Cladellas M., de la Torre R., Martí J., Alcántara M., Pujadas-Bastardes M., Marrugat J., Bruguera J., López-Sabater M.C., Vila J., Covas M.I., and members of the SOLOS Investigators. Antioxidant effect of virgin olive oil in patient with stable coronary heart disease: A randomized, crossover controlled, clinical trial. *Atherosclerosis*, 181:149–158, 2005.

[14] Hollman P.C., Hertog M.G., and Katan M.B. Role of dietary flavonoids in protection against cancer and coronary heart disease. *Biochemical Society Transactions*, 24:785–789, 1996.

[15] Nagata C. Ecological study of the association between soy product intake and mortality from cancer and heart disease in japan. *International Journal of Epidemiology*, 29:832–836, 2000.

[16] Von Schacky C. The role of omega-3 fatty acids in cardiovascular disease. *Current Atherosclerosis Reports*, 5:139–145, 2003.

[17] Willcox J.K., Catignani G.L., and Lazarus S. Tomatoes and cardiovascular health. *Critical Reviews in Food Science and Nutrition*, 43:1–18, 2003.

[18] Calisir E.D, Erkoc S., Yildirim H., Kara A., Rahman T.S., Selvi M., and Erkoc F. Theoretical comparative study of the structure, dynamics and electronic properties of two allyl molecules: S-allyl cysteine (sac) and s-allyl mercaptocysteine (samc). *Int. J. Pure Appl. Chem.*, 1(45–55), 2006.

[19] Yildirim H., Kara A., Rahman T. S., Calisir E. D., Erkoc S., Selvi M., and Erkoc F. Theoretical comparative study of the structure, dynamics and electronic properties of three allyl molecules: Allicin, methyl propyl disulfide (mpd) and allyl methyl sulfide (ams). *Int. J. Pure Appl. Chem.*, 1:171–182, 2006.

[20] Calisir E.D. and Erkoc S. Structural and electronic properties of dipropyl sulfide: A theoretical investigation. *Int. J. Mod. Phys. C*, 17:1179–1190, 2006.

[21] Calisir E.D. and Erkoc S. Structural, electronic and qsar properties of the cyfluthrin molecule: A theoretical am1 and pm3 treatment. *Int. J. Mod. Phys. C*, 17:1391–1402, 2006.

[22] Maaß A., Tekin E.D., Schüllera A., Palazoglu A., Reitha D., and Faller R. Folding and unfolding characteristics of short beta strand peptides under different environmental conditions and starting configurations. submitted to BBA-Proteins and Proteomics, 2010.

[23] Borek C. Antioxidant health effects of aged garlic extract. *J. Nutr*, 131:1010S–1015S, 2001.

[24] Tapiero H., Townsend D.M., and Tew K.D. Organosulfur compounds from alliaceae in the prevention of human pathologies. *Biomedicine and Pharmacotherapy*, 58:183–193, 2004.

[25] Banerjee S.K., Mukherjee P.K., and Maulik S.K. Garlic as an antioxidant: The good, the bad and the ugly. *Phytother. Res.*, 17:97–106, 2003.

[26] Maldonado P.D., Barrera D., Rivero I., Mata R., Medina-Campos O.N., Hernandez-Pando R., and Pedraza-Chaverri J. Antioxidant s-allylcysteine prevents gentamicin-induced oxidative stress and real damage. *Free Rad. Biol. Med*, 35:317–324, 2003.

[27] Xiao D., Pinto J.T., Soh J.W., Deguchi A., Gundersen G.G., Palazzo A.F., Yoon J.T., Shirin H., and Weinstein I.B. Induction of apoptosis by the garlic-derived compound s-allylmercaptocycteine (samc) is associated with microtubule depolimerization and c-jun nh(2)-terminal kinase 1 activation. *Cancer Res*, 63:6825–6837, 2003.

[28] Kuettner E.B., Hilgenfeld R., and Weiss M.S. Purification, characterization, and crystallization of alliinase from garlic. *Arch. Biochem. Biophys.*, 402:192–200, 2002.

[29] Block E. The chemistry of garlic and onions. *Scientific American*, 252:94–99, 1985.

[30] MacDonald J.A. and Langler R.F. Structure-activity relationships for selected sulfur-rich antithrombotic comounds. *Biochem. Biophys. Res. Commun.*, 273:421–424, 2002.

[31] Silagy C.A. and Neil A. Garlic as a lipid-lowering agent-a metaanalysis. *J. Royal Coll. Physicians London*, 28:39–45, 1994.

[32] Ankri S. and Mirelman D. Antimicrobial properties of allicin from garlic. *Microbes Infect.*, 1:125–129, 1999.

[33] Belman S., Solomon J., Segal A., Block E., and Barany G. Inhibition of soybean lipoxygenase and mous skin tumor promotion by onion and garlic components. *J. Biochem. Toxicol.*, 4:151–160, 1989.

[34] Dirsch V.M., Alexandra K.K., Wagner H., and Vollmar A.M. Effect of allicin and ajoene, two compounds of garlic, on inducible nitric oxide synthase. *Atherosclerosis*, 139:333–339, 1998.

[35] Sela U., Ganor S., Hecht I., Brill A., Miron T., Rabinkov A., Wilchek M., Mirelman D., Lider O., and Hershkoviz R. Allicin inhibits sdf-1α-induced t cell interactions with fibronectin and endothelial cells by down-regulating cytoskeleton rearrangement, pyk-2 phosphorylation and vla-4 expression. *Immunology*, 111:391–399, 2004.

[36] Shadkchan Y., Shemesh E., Mirelman D., Miron T., Rabinkov A., Wilchek M., and Osherov N. Efficacy of allicin, the reactive molecule of garlic, in inhibiting aspergillus spp. in vitro, and in a murine model of disseminated aspegillosis. *J. Antimicrob. Chemoth.*, 53:832–836, 2004.

[37] http://www.healingpeople.com/ht/EN/articles/2000/3/15/593.tmpl. Online; accessed 15-November-2005.

[38] http://www.epochem.com/products/flavor/2179-60-4.htm. Online; accessed 15-November-2005.

[39] Matsuda T., Takada N., Yano Y., Wanibuchi H., Otani S., and Fukushima S. Dose-dependent inhibition of glutathione s-transferase placental form-positive hepatocellular foci induction in the rat by methyl propyl disulfide and propylene sulfide from garlic and onions. *Cancer Letters*, 86:229–234, 1994.

[40] Takada N., Matsuda T., Otoshi T., Yano Y., Otani S., Hasegawa T., Nakae D., Konishi Y., and Fukushima S. Enhancement by organosulfur compounds from garlic and onions of diethylnitrosamine-induced glutathione-s-transferase positive foci in the rat-liver. *Cancer Res.*, 54:2895–2899, 1994.

[41] Fukushima S., Takada N., Hori T., and Wanibuchi H. Cancer prevention by organosulfur compounds from garlic and onion. *Journal of Cellular Supplement*, 27:100–105, 1997.

[42] Tamaki T. and Sonoki S. Volatile sulfur compounds in human expiration after eating raw or heat-treated garlic. *J. Nutr. Sci. Vitaminol.*, 45:213–222, 1999.

[43] Lee S.N., Kim N.S., and Lee D.S. Comparative study of extraction techniques for determination of garlic flavor components by gas chromatography-mass spectrometry. *Anal. Bioanal. Chem.*, 377:749–756, 2003.

[44] Lawson L.D. and Wang Z.J. Allicin and allicin-derived garlic compounds increase breath acetone through allyl methyl sulfide: Use in measuring allicin bioavailability. *Agric. Food Chem.*, 53:1974–1983, 2005.

[45] IPCS International Programme on Chemical Safety. www.inchem.org/documents/jecfa/jecmono/v44jec09.htm, 2000. Online; accessed 15-July-2006.

[46] Snow C.D., Sorin E.J., Rhee Y.M., and Pande V.S. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.*, 34:43–69, 2005.

[47] Caflisch A. and Karplus M. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl Acad. Sci. USA*, 91:1746–1750, 1994.

[48] Wong K.B., Clarke J., Bond C.J., Neira J.L., Freund S.M., Fersht A.R., and Daggett V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.*, 296:1257–1282, 2000.

[49] Tirado-Rives J., Orozco M., and Jorgensen W.L. Molecular dynamics simulations of the unfolding of barnase in water and 8 m aqueous urea. *Biochemistry*, 36:7313–7329, 1997.

[50] Daggett V. and Levitt M. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.*, 232:600–618, 1993.

[51] Lazaridis T. and Karplus M. New view of protein folding reconciled with the old through multiple unfolding simulations. *Science*, 278:1928–1931, 1997.

[52] Finkelstein A.V. Can protein unfolding simulate protein folding? *Protein Eng.*, 10:843–845, 1997.

[53] Dastidar S.G. and Mukhopadhyay C. Unfolding dynamics of the protein ubiquitin: Insight from simulation. *Phys Rev E*, 71:051928/1–051928/10, 2005.

[54] Day R., Bennion B.J., Ham S., and Daggett V. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.*, 322:189–203, 2002.

[55] Suzuki S., Galzitskaya O.V., Mitomo D., and Higo J. General dynamic properties of abeta12-36 amyloid peptide involved in alzheimer's disease from unfolding simulation. *J. Biochem.*, 136:583–594, 2004.

[56] Rcsb protein data bank. http://www.pdb.org/pdb/home/home.do. Online; accessed 15-September-2007.

[57] Ho B.K. and Dill K.A. Folding very short peptides using molecular dynamics. *PLoS Comput Biol*, 2(4):228–237, 2006.

[58] Munoz V., Thompson P.A., Hofrichter J.A., and Eaton W.A. Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 390:196–199, 1997.

[59] Gnanakaran S., Nymeyer H., Portman J., Sanbonmatsu K.Y., and Garcia A.E. Peptide folding simulations. *Current Opinion in Structural Biology*, 13:168–174, 2003.

[60] Pande V.S. and Rokhsar D.S. Molecular dynamics simulations of unfolding and re-folding of a beta-hairpin fragment of protein g. *Proc. Nat. Acad. Sci.*, 96:9062–9067, 1999.

[61] Kobayashi N., Honda S., Yoshii H., and Munekata E. Role of side-chains in the co-operative β-hairpin folding of the short c-terminal fragment derived from streptococcal protein g. *Biochemistry*, 39:6564–6571, 2009.

[62] Karhunen K. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn. Ser. A. I.*, 37:3–79, 1947.

[63] Loeve M.M. Fonctions aleatories de seconde ordre. In Levy P., editor, *Process Stochastiques et Movement Brownien*. Hermann, Paris, 1948.

[64] Loeve M.M. *Probability Theory*. van Nostrand, Princeton(NJ), 3 edition, 1963.

[65] Palazoglu A., Gursoy A., Arkun Y., and Erman B. Characterization of dynamics of protein folding from denatured to native state. *J. Computational Biology*, 11:1149–1168, 2004.

[66] Das A. and Mukhopadhyay C. Application of principal component analysis in pro-tein unfolding: An all-atom molecular dynamics simulation study. *J. Chem. Phys.*, 127:165103–165108, 2007.

[67] Lou H. and Cukier R.I. Molecular dynamics of apo-adenylate kinase: A principal component analysis. *J. Phys. Chem. B.*, 110:12796–12808, 2006.

[68] Maisuradze G.G., Liwo A., and Scheraga H.A. Principal component analysis for pro-tein folding dynamics. *J. Mol. Biol.*, 385:312–329, 2009.

[69] Burkert U. and Allinger N.L. *Molecular Mechanics*. ACS Monograph 177, Washing-ton, D.C., 1982.

[70] HyperChem(TM) Professional 7.51. *Hypercube Inc.* Wiley, Gainesville, FL, USA., hyperchem 7.5 edition, October 2003.

[71] Allinger N.L. Conformational analysis 130. mm2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *J. Am. Chem. Soc.*, 99:8127–8134, 1977.

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

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

[74] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M. Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., and Kollman P.A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[75] Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S., and Karplus M. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

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

[77] Pranata J., Wierschke S.G., and Jorgensen W.L. Opls potential functions for nucleotide bases. relative association constants of hydrogen-bonded base pairs in chloroform. *J. Am. Chem. Soc.*, 113:2810–2819, 1991.

[78] Allinger N.L., Yuh Y.H., and Lii J.H. Molecular mechanics. the mm3 force field for hydrocarbons.1. *J. Am. Chem. Soc.*, 111:8551–8566, 1989.

[79] London F. On the theory and system of molecular forces. *Z. Phys.*, 63:245–275, 1930.

[80] Polder D. Casimir H.B.G. The influence of retardation on the london-van der waals forces. *Phys. Rev.*, 73:360–372, 1948.

[81] Casimir H.B.G. On the attraction between two perfectly conducting plates. *Proc. Kon. Ned. Akad. Wet.*, 51:793–795, 1948.

[82] Alder B.J and Wainwright T.E. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.

[83] McCammon J.A, Gelin B.R, and Karplus M. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[84] Rapaport D.C. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, New York, 1995.

[85] Haile J.M. *Molecular Dynamics Simulation: Elementary Methods*. John Wiley & Sons, Inc., New York, 1992.

[86] Allen M.P. and Tildesley D.J. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1989.

[87] Verlet L. Computer 'experiments' on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[88] Verlet L. Computer 'experiments' on classical fluids. ii. equilibrium correlation functions. *Phys. Rev*, 165:201–214, 1968.

[89] Leach A. R. *Molecular modelling principles and applications*. Pearson, Harlow, England, second edition, 2001.

[90] Berendsen H. J. C, Postma J. P. M, van Gunsteren W. F, DiNola A., and Haak J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(3684–3690), 1984.

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

[92] Hess B., Bekker H., Berendsen H. J. C, and Fraaije J. G. E. M. Lincs: A linear constraint solver for molecular simulations,. *J. Comput. Chem*, 18:1463–1472, 1997.

[93] Frenkel D. and Smit B. *Understanding Molecular Simulation from algorithms to applications*. Academic Press, San Diego, second edition, 2002.

[94] Born M. and Oppenheimer R. Zur quantentheorie der moleküle (on the quantum theory of molecules). *ann. Phys.*, 84:457–484, 1927.

[95] Springborg M. *Methods of Electronic-Structure Calculations From Molecules to Solids*. John Wiley&Sons Ltd., England, 2000.

[96] Erkoc S. and Uzer T. *Lecture notes on Atomic and Moleculer Physics*. World Scientific, Singapore, 1996.

[97] Hartree D. R. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. *Proc. Cam. Phil. Soc.*, 24(01):89–110, 1928.

[98] Hartree D. R. The wave mechanics of an atom with a non-coulomb central field. part ii. some results and discussion. *Proc. Cam. Phil. Soc.*, 24(01):111–132, 1928.

[99] Hartree D. R. The wave mechanics of an atom with a non-coulomb central field. part iii. term values and intensities in series in optical spectra. *Proc. Cam. Phil. Soc.*, 24(03):426–437, 1928.

[100] Hartree D. R. The distribution of charge and current in an atom consisting of many electrons obeying dirac's equations. *Proc. Cam. Phil. Soc.*, 25(02):225–236, 1929.

[101] Hartree D. R. The wave mechanics of an atom with a non-coulomb central field. part iv. further results relating to terms of the optical spectrum. *Proc. Cam. Phil. Soc.*, 25(03):310–314, 1929.

[102] Slater J. C. Wave functions in a periodic potential. *Phys. Rev.*, 51:846–851, 1937.

[103] Slater J. C. A simplification of the hartree-fock method. *Phys. Rev.*, 81:385–390, 1951.

[104] Fock V. First publication on hartree-fock method. *Z. Phys.*, 61:126, 1930.

[105] Arfken G. B. and Weber H. J. *Mathematical Methods for Physicists*. Academic Press, Boston, sixth edition, June 2005.

[106] Koopmans T. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica(Elsevier)*, 1:104–113, 1934.

[107] Roothan C. C. J. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, 1951.

[108] Sadlej J. *Semi-empirical methods of quantum chemistry*. John Wiley, New York, 1985.

[109] Zerner M. C. Semiempirical molecular orbital methods. *Rev. Comput. Chem*, 2:313–365, 1991.

[110] Jensen F. *Introduction to Computational Chemistry*. John Wiley Sons, Chichester, England, 1999.

[111] Pople J. A. and Segal G. A. Approximate self-consistent molecular orbital theory. ii. calculations with complete neglect of differential overlap. *Journal of Chemical Physics*, 43:S136–S152, 1965.

[112] Pople J., Beveridge D. L., and Dobosh P. A. Approximate self-consistent molecular orbital theory. v. intermediate neglect of differential overlap. *Journal of Chemical Physics*, 47:2026–2033, 1967.

[113] Pople J. and Beveridge D. *Approximate Molecular Orbital Theory*. McGraw-Hill, New York, 1970.

[114] Dewar M. J. S. and Thiel W. Ground states of molecules. 38. the mndo method. approximations and parameters. *Journal of the American Chemical Society*, 99:4899–4907, 1977.

[115] M. J. S. Dewar and Thiel W. Ground states of molecules. 39. mndo results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *Journal of the American Chemical Society*, 99:4907–4917, 1977.

[116] Dewar M. J. S., Zoebisch E. G., Healy E. F., and J. J. P. Stewart. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107:3902–3909, 1985.

[117] J. J. P. Stewart. Optimization of parameters for semiempirical methods i. method. *J. Comput. Chem.*, 10:209–220, 1989.

[118] J. J. P. Stewart. Optimization of parameters for semiempirical methods ii. applications. *J. Comput. Chem.*, 10:221–264, 1989.

[119] Thomas L. H. The calculation of atomic fields. *Proc. Camb. Phil. Soc.*, 23:542–548, 1927.

[120] Fermi E. Un metodo statistico per la determinazione di alcune prioprietà dell'atomo. *Rend. Accad. Naz. Lincei*, 6:602–607, 1927.

[121] Dirac P.A.M. Note on the exchange phenomena in the thomas-fermi atom. *Proc. Cambridge Phil. Roy. Soc.*, 26:376–385, 1930.

[122] Kohn W. Nobel lecture: Electronic structure of matter, wave functions and density functionals. *Reviews of Modern Physics*, 71(5):1253–1266, 1999.

[123] Hohenberg P. and Kohn W. Inhomogeneous electron gas. *Phys. Rev. B*, 136:864–871, 1964.

[124] Kohn W. and Sham L.J. Self-consistent equations including exchange and corelation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[125] Fletcher P. *Practical Methods of Optimization*. Wiley, New York, 1990.

[126] Becke A.D. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[127] Lee C., Yang W., and Parr R.G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, 37:785–789, 1988.

[128] Scott A. P. and Radom L. Harmonic vibrational frequencies: An evaluation of hartree-fock, moller-plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.*, 100:16502–16513, 1996.

[129] Characteristic infrared absorption frequencies. http://wwwchem.csustan.edu/Tutorials/INFRARED.HTM, 1998. Online; accessed 1-November-2005.

[130] McClain B.L., Clark S.M., Gabriel R.L., and Ben-Amotz D. Educational applications of ir and raman spectroscopy: A comparison of experiment and theory. *J. Chem. Education*, 77:654–660, 2000.

[131] http://www.sigmaaldrich.com/spectra/ftir/. Online; accessed 17-January-2006.

[132] Kubinyi H. *QSAR : Hansch analysis and related approaches*. WCH, Weinheim, 1993.

[133] Henderson G. and Liberatore C. Animated vibrational modes of triatomic molecules. http://jchemed.chem.wisc.edu/JCEWWW/Articles/WWW0001/index.html. Online; accessed 17-July-2006.

[134] Awad M. K., Khairou K. S., and Diab M. A. Theoretical investigations of the stability of degradation products of polystyrene and poly(4-vinylpyridine). *Polymer Degrad. Stability*, 46:165–170, 1994.

[135] Watanabe M., Ishimaru D., Mizorogi N., Kiuchi M., and Aihara J. Thermodynamically and kinetically stable isomers of the c_{88} and c_{90} fullerenes. *J. Mol. Struct. (Theochem)*, 726:11–16, 2005.

[136] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., and Berendsen H. J. C. Gromacs: Fast, flexible, and free. *J. Comp. Chem.*, 26:1701–1718, 2005.

[137] van Gunsteren W. F., Billeter S. R., Eising A. A, Hünenberger P. H, Krüger P., Mark A. E., Scott W. R. P., and Tironi I. G. *Biomolecular simulation: The GROMOS manual and user guide*. Vdf-KONTROL ET, Zürich, 1996.

[138] Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., and Hermans J. Interaction models for water in relation to protein hydration. In Pullman B., editor, *Intermolecular Forces*, pages 331–342. Reidel, Dordrecht, 1981.

[139] Darden T., York D., and Pedersen L. Particle mesh ewald: an n-log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[140] Essman U., Perela L., Berkowitz M.L., Darden H.L.T, and Pedersen L.G. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8592, 1995.

[141] Hockney R.W. The potential calculation and some applications. In Rotenberg M., editor, *Methods in computational physics. IX. Plasma physics*, pages 135–211. Academic Press, London, 1970.

[142] Humphrey W., Dalke A., and Schulten K. Vmd-visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[143] Gutte B. *Peptides: Synthesis, Structures and Applications*. Academic Press, San Diego, 1995.

[144] Tsai J. and Levitt M. Evidence of turn and salt bridge contributions to beta-hairpin stability: Md simulations of c-terminal fragment from the b1 domain of protein g. *Biophys. Chem*, 101–102:187–201, 2002.

[145] Jeffrey G. A. Hydrogen bonding in biological molecules. In *An Introduction to Hydrogen Bonding*, chapter 10. Oxford University Press, 1997.

[146] Kobko N. and J.J. Dannenberg. Cooperativity in amide hydrogen bonding chains. relation between energy, position, and h-bond chain length in peptide and protein folding models. *J. Phys. Chem. A*, 107:10389–10395, 2003.

[147] Dinner A. R., Lazaridis T., and Karplus M. Understanding b-hairpin formation. *Proc. Nat. Acad. Sci. USA*, 96:9068–9073, 1999.

[148] Hamill S. J., Meekhof A. E., and Clarke J. The effect of boundary selection on the stability and folding of the third fibronectin type iii domain from human tenascin. *Biochemistry*, 37:8071–8079, 1998.

[149] Kumar S., Tsai C. J., and Nussinov R. Factors enhancing protein thermostability. *Protein Eng.*, 13:179–191, 2000.

[150] Viswanathan R., Asensio A., and Dannenberg J. J. Cooperative hydrogen-bonding in models of antiparallel β-sheets. *J. Phys. Chem. A.*, 108:9205–9212, 2004.

[151] Wang Z. X., Wu C., Lei H., and Duan Y. Accurate ab initio study on the hydrogen-bond pairs in protein secondary structures. *J. Comp. Theor. Comput.*, 3:1527–1537, 2007.

[152] Kabsch W. and Sander C. Dictionary of protein secondary structures: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[153] Hendsch Z. S. and Tidor B. Do salt bridges stabilize proteins? a continuum electrostatic analysis. *Protein Science*, 3:211–226, 1994.

[154] Aihara J. Correlation found between the homo.lumo energy separation and the chemical reactivity at the most reactive site for isolated-pentagon isomers of fullerenes. *Phys. Chem. Chem. Phys.*, 2:3121–3125, 2000.

[155] Fujine K., Ishidab T., and Aihara J. Localization energies for graphite and fullerenes. *Phys. Chem. Chem. Phys.*, 3:3917–3919, 2001.

APPENDIX A

SUPPLEMENTARY INFORMATION

Table A.1: Atomic positions of SAC molecule (in Fig. 3.1) from the optimized structure (DFT/B3LYP/STO-3G). Atom labels are as shown in Fig. 3.2

Atom Label	x	y	z
1	-0.49393333	1.16841134	0.40489382
2	0.09954980	2.56174218	0.16928378
3	-2.00589716	3.68515505	1.74528443
4	-5.23480082	2.18430387	2.80687981
5	-3.97134966	1.87566648	2.49397738
6	-3.23712257	2.39270163	1.25601629
7	-1.01737120	3.66533530	0.20054675
8	0.82855508	2.55660448	-1.22797769
9	1.62407231	1.37354642	-1.31581816
10	0.77332001	3.40482516	-2.14376518
11	0.25751389	0.52010534	0.02375129
12	-0.41164824	1.02264128	1.45325038
13	0.87241060	2.81230199	0.93314315
14	-5.71696979	1.80183253	3.71438498
15	-5.84800079	2.83250546	2.16820826
16	-3.38390364	1.22857452	3.16327964
17	-2.67861285	1.57309923	0.76141099
18	-3.96444808	2.81065932	0.53619489
19	-1.67450772	3.52962215	-0.67587420
20	-0.52667072	4.64934791	0.09745525
21	2.03003084	1.44543924	-2.25743162

Table A.2: Atomic positions of SAMC molecule (in Fig. 3.3 from the optimized structure (DFT/B3LYP/STO-3G). Atom labels are as shown in Fig. 3.4

Atom #	x	y	z
1	-0.04913043	-4.76383042	0.82856068
2	-0.28368485	-3.45712380	0.66290473
3	-1.09155625	-2.86772986	-0.49584652
4	-2.74231448	-2.28849905	0.13621025
5	-3.44803057	-1.25214674	-1.55445778
6	-2.70363850	0.41099460	-1.24693052
7	-3.17795402	1.10570996	0.06550893
8	-2.35436318	2.47878401	0.35849950
9	-3.24139284	3.49855996	0.68481223
10	-1.10731416	2.57805186	0.30231007
. 11	-4.63858286	1.54450503	0.06554655
12	0.53020898	-5.14336924	1.67923508
13	-0.42070416	-5.51484324	0.11975065
14	0.10018720	-2.73058132	1.39530614
15	-0.54355536	-2.01169052	-0.93367647
16	-1.22876245	-3.62852374	-1.28525919
17	-1.59925814	0.35789926	-1.20879197
18	-2.96918204	1.04101136	-2.11509956
19	-3.01926338	0.41347550	0.9187006
20	-4.14585727	2.94849281	0.55216746
21	-5.00403250	1.50558606	-0.92271735
22	-5.22021450	0.85031431	0.60420188

Table A.3: Normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for the SAC molecule for three different levels of computation.

Mode #	AM1	PM3	ab initio (RHF)
1	20.27 (0.09890)	12.30 (0.13863)	39.21 (1.21608)
2	30.73 (0.48209)	26.01 (1.12018)	49.70 (0.55928)
3	36.10 (0.79413)	37.94 (0.35477)	60.32 (1.49930)
4	51.24 (0.11841)	47.26 (0.15979)	92.04 (0.40622)
5	87.38 (1.39222)	70.05 (0.46565)	134.19 (6.79272)
6	153.01 (1.55364)	154.56 (1.13412)	181.04 (3.85534)
7	179.34 (0.22169)	189.03 (1.03026)	212.94 (0.22777)
8	217.96 (1.40557)	193.55 (23.63185)	262.74 (1.18752)
9	242.75 (33.80293)	209.47 (1.55624)	330.20 (7.53729)
10	303.02 (0.70149)	280.39 (1.86040)	345.37 (38.96397)
11	346.32 (2.05022)	332.10 (1.51332)	369.55 (56.33691)
12	395.40 (0.59940)	360.40 (2.76570)	427.92 (5.27031)
13	432.50 (0.11124)	430.99 (0.21850)	465.75 (0.91418)
14	518.87 (10.84130)	471.60 (9.84526)	521.64 (18.74310)
15	552.18 (33.66503)	508.78 (33.96803)	603.00 (67.41998)
16	574.65 (11.13571)	525.29 (26.52614)	619.80 (23.42142)
17	588.57 (18.47851)	586.46 (0.99299)	646.09 (115.47615)
18	724.75 (23.76528)	696.09 (15.26877)	758.54 (110.46301)
19	758.12 (4.79237)	762.70 (4.09430)	795.87 (33.55247)
20	798.96 (3.93839)	777.67 (4.14800)	813.86 (12.81917)
21	874.62 (0.25516)	851.39 (0.09917)	875.51(66.84328)
22	892.58 (4.32300)	866.89 (0.34516)	893.04 (126.34207)
23	958.15 (24.86274)	905.85 (1.53379)	982.14 (19.84485)
24	964.44 (10.40775)	931.86 (19.64824)	989.79 (1.02479)
25	997.93 (1.58407)	955.13 (0.36853)	1004.36 (2.90040)
26	1018.88 (11.40164)	1008.70 (0.02682)	1073.29 (7.54216)
27	1058.15 (20.82254)	1016.89 (6.67361)	1118.59 (78.12666)
28	1084.25 (0.60571)	1046.87 (1.59804)	1162.00 (0.04203)
29	1099.98 (2.04960)	1061.18 (1.32725)	1199.82 (69.13818)
30	1150.14 (3.68297)	1072.20 (2.79115)	1202.66 (19.05556)
31	1206.80 (4.97200)	1126.06 (5.75963)	1217.00 (303.48151)
32	1237.17 (1.90708)	1162.51 (12.35175)	1319.59 (1.63549)
33	1252.73 (2.96402)	1169.84 (0.31036)	1359.61 (2.92624)
34	1259.06 (1.43024)	1194.66 (7.93072)	1366.62 (18.66179)
35	1271.45 (2.35548)	1204.57 (4.31189)	1394.09 (24.71953)
36	1317.36 (0.84903)	1235.43 (19.43651)	1413.06 (7.17734)
37	1372.62 (6.02499)	1263.31 (0.61019)	1454.06 (33.88416)
38	1386.43 (22.99404)	1264.66 (7.16170)	1469.06(0.39061)
39	1392.38 (0.76691)	1302.98 (0.28741)	1490.63 (9.32311)
40	1423.69 (11.46286)	1336.32 (2.73786)	1539.30 (30.64697)
41	1432.89 (1.47577)	1372.91 (5.60723)	1612.43 (3.88963)
42	1453.61 (9.44011)	1397.17 (1.31726)	1618.07 (9.29355)

Mode #	AM1	PM3	ab initio (RHF)
43	1538.65 (62.48066)	1432.82 (69.59943)	1658.18 (3.46581)
44	1707.99 (1.53769)	1654.23 (2.92814)	1830.50 (65.58828)
45	1855.97 (0.58422)	1852.16 (0.40270)	1854.78 (3.34533)
46	2087.01 (124.36231)	1980.35(126.13953)	1967.67 (250.35219)
47	2928.33 (0.14447)	2808.97 (1.33680)	3164.15 (26.70159)
48	3027.93 (4.38690)	2970.04 (2.40824)	3267.95 (11.62026)
49	3035.23 (15.02384)	2973.11(8.46200)	3287.45 (7.00483)
50	3099.05 (2.61114)	3039.61 (6.34504)	3318.25 (8.87690)
51	3105.78 (7.13733)	3044.04 (3.04440)	3331.15 (1.11210)
52	3153.52 (4.19190)	3050.35 (4.93572)	3344.49 (5.84250)
53	3179.91 (8.99878)	3135.03 (3.04379)	3348.80 (1.53298)
54	3213.26 (16.76557)	3146.22 (5.87754)	3397.71 (17.50721)
55	3411.31 (15.52102)	3389.83 (0.07947)	3692.84 (8.62266)
56	3425.33 (65.05240)	3519.33 (2.65931)	3793.31 (18.99084)
57	3453.28 (9.85071)	3851.70 (18.03686)	3872.67 (96.97775)

Table A.4: Normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for the SAMC molecule for three different levels of computation.

Mode #	AM1	PM3	ab initio (RHF)
1	16.00 (0.22218)	11.01 (0.06843)	24.65 (1.60894)
2	25.31 (1.67047)	24.82 (1.94436)	37.90 (6.00082)
3	31.54 (0.39833)	31.75 (0.98983)	52.55 (1.12925)
4	42.05 (1.10633)	39.45 (0.27022)	68.40 (3.48991)
5	45.96 (0.17700)	47.14 (0.87819)	77.36 (1.32824)
6	72.00 (1.38867)	83.93 (1.53559)	106.34 (2.66736)
7	116.69 (0.64400)	118.78 (1.17726)	128.57 (4.04901)
8	166.53 (4.09380)	175.98 (3.20816)	194.83 (12.76062)
9	186.87 (1.86695)	198.39 (2.16363)	228.79 (4.96135)
10	237.26 (15.65453)	243.95 (12.82039)	300.34 (3.23405)
11	259.38 (0.54253)	266.91 (0.78473)	311.13 (3.12508)
12	279.12 (2.23797)	274.71 (1.37661)	326.19 (17.64317)
13	353.08 (3.77018)	328.62 (0.94856)	385.77 (17.02167)
14	390.79 (0.86354)	402.65 (8.24323)	445.81 (3.34453)
15	413.37 (7.36655)	433.20 (0.33783)	462.84 (1.77158)
16	435.30 (0.28007)	454.80 (38.04403)	477.06 (1.97146)
17	466.07 (42.49999)	475.87 (1.48828)	552.49 (4.33981)
18	514.26 (2.19349)	514.46 (0.62287)	607.11 (4.05102)
19	574.38 (2.79693)	549.26 (7.61197)	643.80 (14.29951)
20	594.60 (10.37771)	583.49 (2.35286)	711.42 (11.78756)
21	710.54 (9.31791)	675.77 (4.34152)	807.91 (20.77764)
22	755.46 (8.38068)	760.22 (3.49802)	819.05 (54.67749)
23	789.52 (2.98512)	761.70 (2.55586)	863.39 (187.90100)
24	865.17 (6.35508)	833.38 (6.13848)	892.24 (4.75846)
25	873.05 (0.23385)	859.34 (0.11650)	917.45 (240.14729)
26	961.78 (15.93113)	875.99 (0.17026)	960.90 (61.24692)
27	971.25 (4.48645)	927.93 (19.61740)	969.73 (5.83514)
28	991.57 (1.86368)	952.77 (0.43631)	997.58 (0.26109)
29	1020.14 (57.99967)	996.14 (5.30899)	1106.68 (79.93375)
30	1062.29 (19.28556)	1006.81 (4.56556)	1127.23 (76.58662)
31	1069.47 (1.36771)	1030.71 (0.34756)	1166.32 (0.57409)
32	1078.75 (0.92788)	1041.30 (4.93773)	1195.61 (0.92113)
33	1170.37 (2.91209)	1074.23 (6.45982)	1197.76 (9.41884)
34	1202.60 (0.14512)	1102.56 (12.25630)	1226.69 (14.35923)
35	1229.59 (3.83616)	1167.28 (1.02326)	1309.70 (7.56478)
36	1233.77 (5.15487)	1176.55 (1.47828)	1345.90 (0.44934)
37	1268.41 (1.43297)	1182.19 (13.35158)	1369.08 (17.69975)
38	1285.89 (4.03384)	1188.45 (4.01794)	1386.75 (67.03938)
39	1311.57 (0.31942)	1256.68 (2.24786)	1405.56 (5.43235)
40	1345.47 (3.90255)	1272.45 (10.95783)	1452.21 (400.03088)
41	1366.99 (8.84984)	1285.31 (5.14504)	1468.99 (2.83336)
42	1392.21 (2.19122)	1315.69 (21.13635)	1476.54 (279.76581)

Mode #	AM1	PM3	ab initio (RHF)
43	1429.78 (1.38422)	1332.04 (2.83661)	1523.70 (5.44987)
44	1431.88 (4.17374)	1366.14 (9.86703)	1609.27 (6.81465)
45	1450.59 (11.15910)	1397.29 (1.77014)	1617.37 (13.12914)
46	1539.95 (151.49809)	1432.12 (113.96259)	1636.98 (8.70450)
47	1718.88 (2.93855)	1653.06 (6.16005)	1849.32 (47.60963)
48	1848.09 (3.66490)	1844.94 (3.00270)	1852.09 (1.51416)
49	2070.03 (117.93750)	1977.87 (109.17891)	1986.90 (267.54187)
50	2949.35 (3.85055)	2819.17 (5.11108)	3245.42(3.98031)
51	3020.03 (22.86767)	2958.14 (13.73251)	3261.36 (1.67246)
52	3038.88 (4.56900)	2985.04 (2.25824)	3278.71 (9.60922)
53	3095.15 (17.00257)	3040.33 (7.52804)	3317.52 (5.93587)
54	3106.42 (2.34716)	3040.67 (12.20362)	3338.71 (3.21394)
55	3153.67 (4.36761)	3054.86 (2.46764)	3352.71 (0.23464)
56	3178.87 (11.59715)	3133.64 (3.75525)	3354.85 (2.78529)
57	3210.82 (20.68442)	3144.40 (6.73161)	3398.05 (12.52708)
58	3406.88 (13.53413)	3388.74 (0.28795)	3666.93 (14.81096)
59	3436.57 (53.83567)	3517.35 (4.32664)	3720.27 (208.33624)
60	3455.97 (9.36782)	3870.14 (15.80288)	3768.23 (14.86009)

Table A.5: Atomic positions of allicin molecule (in Fig. 3.9) from the optimized structure (DFT/B3LYP/STO-3G). Atom labels are as shown in Fig. 3.10

Atom #	x	y	z
1	1.05489945	-3.49028615	0.89646640
2	0.90660251	-2.63744259	-0.12679892
3	0.26735265	-1.26679619	-0.01238989
4	1.59916706	0.14569903	-0.11413695
5	0.49230182	1.44612136	-1.72382358
6	-0.07484262	2.57093083	-0.45101225
7	-1.59183462	2.60259333	-0.23394159
8	-2.22201977	2.00142228	0.78218683
9	1.26140018	1.05866134	1.20487666
10	1.53061486	-4.46983115	0.76529188
11	0.70167007	-3.25141870	1.90757096
12	1.27152751	-2.91443139	-1.12755076
13	-0.43758911	-1.06869352	-0.83913171
14	-0.24164135	-1.12470018	0.95602744
15	0.44283856	2.11694795	0.50813460
16	0.31543470	3.59222837	-0.61128542
17	-2.18285381	3.15715582	-0.98093484
18	-3.31194464	2.03936148	0.89560545
19	-1.67341944	1.44866135	1.55494242

Table A.6: Atomic positions of MPD molecule (in Fig. 3.11) from the optimized structure (DFT/B3LYP/STO-3G). Atom labels are as shown in Fig. 3.12

Atom #	x	y	z
1	0.67050255	-2.32263021	0.33767558
2	0.59367842	-1.44604599	-1.27414913
3	-1.38532852	-0.72755747	-1.22096020
4	-1.11665043	0.89618941	-0.38201405
5	-2.46938807	1.66987666	-0.28897996
6	-2.30513490	3.05146916	0.40876354
7	1.66548649	-2.78829164	0.41594619
8	0.54535946	-1.64019988	1.19457199
9	-0.08752950	-3.11907515	0.41569078
10	-0.71056113	0.74959638	0.63599608
11	-0.38475963	1.50456229	-0.94431166
12	-2.87581929	1.81909760	-1.30484094
13	-3.20210094	1.06231102	0.27135950
14	-3.27340824	3.57313061	0.45955973
15	-1.92460787	2.92722692	1.43443308
16	-1.59837831	3.68644517	-0.14779093

Table A.7: Atomic positions of AMS molecule (in Figure Fig. 3.13) from the optimized structure (DFT/B3LYP/STO-3G). Atom labels are as shown in Fig. 3.14

Atom #	x	y	z
1	-0.56927361	-1.16796587	0.40476403
2	-1.93438039	1.52409291	-0.39393550
3	-2.30066159	2.77794925	0.39902131
4	-0.36235023	0.71142313	-0.33017631
5	1.36322129	-1.68607094	0.21209660
6	-2.80876635	0.96943761	-1.17161439
7	-3.45899727	3.33438626	0.33128360
8	-1.46567582	3.36087423	1.18748694
9	0.15073938	0.69880958	-1.54581251
10	0.46227172	1.44287776	0.38999274
11	1.50932818	-2.90910735	0.62490990
12	1.80237013	-1.64210214	-1.02270759
13	2.18481268	-0.94402670	0.91467425

Table A.8: Normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for allicin molecule for three different levels of computation.

Mode #	AM1	PM3	RHF/3-21G
1	21.81 (0.23088)	21.49 (0.02618)	51.09 (0.68275)
2	30.14 (0.38524)	32.18 (0.94640)	90.50 (0.21322)
3	41.69 (0.57654)	34.19 (0.26350)	99.20 (0.68172)
4	51.80 (0.11304)	53.05 (0.01838)	115.49 (6.10954)
5	66.00 (0.45683)	57.94 (0.46681)	140.05 (0.27581)
6	133.87 (2.35834)	120.07 (1.83299)	167.24 (0.66079)
7	145.11 (3.26242)	165.13 (1.68058)	238.60 (11.75523)
8	213.94 (0.60025)	221.70 (0.44605)	271.97 (17.21932)
9	258.62 (2.17258)	277.83 (1.19784)	309.01 (1.45596)
10	325.22 (3.60751)	323.00 (5.78025)	327.38 (3.14581)
11	372.52 (7.76525)	370.80 (4.20393)	354.52 (6.74158)
12	419.68 (4.45162)	392.11 (6.07662)	432.76 (7.10728)
13	426.07 (9.13555)	430.24 (1.60256)	464.63 (19.04489)
14	443.91 (5.49933)	441.19 (4.22944)	480.37 (0.42154)
15	574.27 (4.18481)	572.13 (5.55830)	604.08 (13.86565)
16	586.83 (2.88949)	590.97 (1.64276)	617.56 (2.69900)
17	740.19 (16.86527)	732.24 (0.88649)	727.56 (44.67854)
18	743.96 (7.78587)	746.00 (8.81620)	804.07 (38.49809)
19	855.78 (30.69740)	774.23 (47.64508)	814.05 (41.88302)
20	873.13 (1.74062)	853.75 (2.76282)	960.72 (4.62187)
21	904.61 (51.16890)	879.81 (9.40219)	982.27 (5.56276)
22	960.20 (11.42352)	927.62 (15.09196)	1003.82 (2.71387)
23	966.29 (11.17291)	931.03 (15.54926)	1008.89 (7.95120)
24	998.40 (1.04107)	958.50 (0.62762)	1132.09 (77.28049)
25	1002.94 (3.08721)	958.78 (0.97645)	1159.67 (73.21985)
26	1055.84 (23.13420)	1009.43 (12.44542)	1170.36 (27.52311)
27	1062.29 (19.34299)	1014.80 (4.82025)	1179.90 (3.23266)
28	1081.75 (1.68337)	1042.18 (3.78027)	1222.50 (11.89382)
29	1094.62 (0.69931)	1042.78 (0.99433)	1233.21 (7.34582)
30	1249.76 (8.58027)	1169.07 (0.22140)	1350.98 (0.36811)
31	1253.30 (0.67548)	1170.83 (0.85881)	1369.17 (7.12740)
32	1265.95 (0.55999)	1186.50 (5.21242)	1410.93 (35.02794)
33	1275.78 (6.62991)	1194.12 (1.92164)	1462.99 (2.92936)
34	1312.61 (2.93920)	1251.49 (3.61271)	1473.67 (1.39504)
35	1312.90 (3.21175)	1258.98 (0.52123)	1480.13 (13.19096)
36	1379.93 (1.99159)	1327.32(2.79002)	1589.31 (15.65461)
37	1385.41 (10.36227)	1331.52 (1.61406)	1602.03 (5.73314)
38	1428.01 (5.23939)	1384.37 (1.47518)	1621.11 (7.69994)
39	1429.68 (2.695779)	1391.75 (4.38459)	1633.40 (22.03887)
40	1853.30 (0.57077)	1847.41 (0.59994)	1840.09 (3.01716)
41	1854.47 (0.22012)	1848.87 (0.77192)	1856.21 (1.21745)

Mode #	AM1	PM3	RHF/3-21G
42	3011.88 (15.13048)	2968.35 (8.37698)	3225.61 (9.92279)
43	3016.54 (5.13425)	2978.75 (5.85159)	3280.92 (4.21202)
44	3084.92 (12.97911)	3037.23 (10.27939)	3317.02 (4.93845)
45	3088.98 (11.34299)	3039.79 (6.34890)	3319.71 (7.56092)
46	3150.06 (4.26610)	3041.93 (7.82999)	3323.23 (9.38419)
47	3153.92 (4.24697)	3046.15 (5.26522)	3344.53 (9.90241)
48	3177.42 (10.09376)	3131.52 (3.77287)	3353.22 (1.22718)
49	3179.74 (9.88519)	3133.76 (4.01869)	3372.45 (3.31710)
50	3210.65 (19.60897)	3144.43 (6.73852)	3401.87 (10.19294)
51	3212.91 (13.69313)	3144.90 (4.89194)	3408.73 (6.39661)

Table A.9: Normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for methyl propyl disulfide molecule for three different levels of computation.

Mode #	AM1	PM3	RHF/3-21G
1	23.45 (0.18300)	12.97 (0.14517)	51.12 (0.69877)
2	37.95 (0.37318)	48.18 (0.53858)	59.81 (1.33994)
3	76.48 (0.37821)	81.25 (0.32703)	113.03 (0.44638)
4	92.83 (0.12317)	94.18 (0.21667)	122.58 (1.09426)
5	124.79 (0.73641)	141.15 (0.61176)	158.18 (0.73349)
6	163.33 (0.00628)	185.77 (0.01602)	217.84 (1.16878)
7	195.15 (0.76061)	207.35 (0.64168)	257.21 (0.01810)
8	257.60 (0.76696)	272.12 (0.94222)	289.49 (2.05865)
9	355.39 (0.14152)	360.73 (0.16678)	357.10 (0.10457)
10	410.17 (0.76302)	515.21 (1.18539)	484.90 (0.28623)
11	738.99 (10.08333)	727.43 (2.94659)	679.51 (6.46812)
12	770.29 (2.82477)	767.28 (1.83064)	755.48 (13.97456)
13	772.19 (3.30194)	772.86 (0.67285)	817.28 (5.62652)
14	887.53 (0.53495)	876.41 (0.23606)	950.60 (1.51253)
15	913.93 (1.31729)	905.71 (6.10448)	953.36 (0.14408)
16	918.10 (5.26375)	916.41 (0.57326)	1060.99 (1.61871)
17	1009.41 (0.38603)	963.24 (0.69880)	1069.12 (2.97887)
18	1012.28 (1.55374)	992.78 (0.33275)	1082.06 (12.41559)
19	1186.69 (0.33788)	1115.58 (0.20865)	1179.94 (0.76599)
20	1195.36 (0.07333)	1123.73 (0.00708)	1206.65 (20.88508)
21	1211.91 (3.09676)	1128.95 (0.42865)	1374.09 (0.12629)
22	1240.82 (1.83963)	1137.36 (3.06035)	1401.05 (35.83687)
23	1259.03 (3.11848)	1203.11 (2.40656)	1465.93 (0.47408)
24	1343.49 (2.28655)	1317.76 (1.49120)	1497.81 (13.63618)
25	1353.93 (5.60851)	1329.53 (1.43754)	1504.46 (3.66753)
26	1380.44 (2.60877)	1369.63 (4.06535)	1576.04 (7.05665)
27	1385.32 (2.17662)	1374.52 (1.15303)	1632.77 (15.30903)
28	1391.96 (0.74758)	1391.10 (2.17744)	1635.38 (3.74609)
29	1393.78 (0.41550)	1402.06 (1.38508)	1641.86 (17.48095)
30	1400.98 (1.34814)	1407.66 (0.26850)	1660.85 (3.51323)
31	1405.62 (1.56506)	1408.96 (0.30211)	1672.12 (9.62477)
32	1434.58 (0.21733)	1420.00 (0.96172)	1673.46 (7.14808)
33	3010.43 (0.48613)	2952.86 (0.36134)	3197.61 (27.11296)
34	3047.66 (4.00950)	2990.80 (2.24989)	3221.26 (6.83688)
35	3058.56 (16.81148)	3031.11 (0.77013)	3241.91 (17.41374)
36	3062.12 (0.12197)	3063.45 (1.77659)	3244.84 (2.74186)
37	3065.15 (0.93610)	3080.16 (0.00874)	3261.40 (32.40405)
38	3075.86 (4.16638)	3085.73 (0.14928)	3269.27 (27.46600)
39	3085.98 (1.19953)	3095.14 (11.82432)	3273.09 (26.02818)
40	3115.89 (1.89440)	3111.52 (2.72061)	3333.54 (2.01707)
41	3157.35 (0.15483)	3181.99 (0.18815)	3338.58 (11.23215)
42	3161.67 (6.40617)	3197.85 (5.93322)	3357.45 (4.08774)

Table A.10: Normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for allyl methyl sulfide molecule for three different levels of computation.

Mode #	AM1	PM3	RHF/3-21G
1	33.82 (0.56190)	23.40 (0.56779)	60.61 (0.89471)
2	53.18 (0.04862)	56.95 (0.14786)	94.04 (0.69974)
3	104.93 (0.49998)	103.46 (0.38982)	160.29 (0.43488)
4	175.18 (0.30413)	179.65 (0.47481)	201.38 (0.63549)
5	299.37 (0.43459)	298.02 (0.29518)	324.33 (2.55096)
6	427.49 (0.11209)	431.21 (0.04679)	458.33 (1.11316)
7	579.42 (0.69146)	584.72 (0.87636)	619.98 (1.62256)
8	756.11 (10.33988)	754.34 (3.27445)	697.06 (5.58043)
9	777.96 (3.05476)	772.79 (3.86852)	818.51 (28.14312)
10	871.46 (0.08989)	856.65 (0.16173)	962.77 (3.16990)
11	919.31 (0.01217)	919.97 (0.00683)	997.02 (0.82273)
12	941.54 (5.90385)	926.93 (17.02402)	1059.32 (3.92275)
13	968.46 (6.85171)	945.94 (2.34771)	1088.24 (2.61666)
14	997.71 (1.29968)	955.01 (0.50798)	1119.50 (81.51910)
15	1055.56 (21.08837)	1008.66 (5.10288)	1163.11 (0.25928)
16	1081.45 (0.95685)	1042.91 (4.02755)	1200.48 (3.19575)
17	1249.68 (3.38228)	1167.44 (0.27781)	1354.55 (0.78463)
18	1268.74 (1.93721)	1190.83 (2.59134)	1392.66 (40.50105)
19	1316.79 (0.98760)	1264.17 (2.67365)	1467.76 (0.12754)
20	1352.73 (1.18493)	1329.35 (1.98093)	1500.55 (6.62855)
21	1364.63 (4.50668)	1336.90 (2.42400)	1611.70 (4.16054)
22	1382.51 (2.59209)	1381.80 (3.08435)	1640.30 (12.76952)
23	1391.74 (1.45143)	1394.50 (1.82048)	1643.72 (9.34755)
24	1433.10 (1.32988)	1398.96 (1.15307)	1648.19 (8.22017)
25	1857.30 (0.23292)	1852.84 (0.15907)	1856.50 (4.12436)
26	3025.53 (2.28356)	2972.76 (1.00319)	3234.13 (28.23415)
27	3069.60 (1.72089)	3039.56 (7.10951)	3247.89 (18.75519)
28	3073.36 (9.32847)	3048.12 (1.65639)	3308.52 (1.52626)
29	3097.03 (1.40410)	3103.59(1.15164)	3318.04 (12.58122)
30	3152.80 (5.13748)	3106.96 (6.38015)	3325.87 (10.43640)
31	3164.71 (2.66786)	3135.58 (2.97793)	3329.92 (3.53788)
32	3180.23 (8.50080)	3146.90 (5.51854)	3341.57 (5.66294)
33	3214.36 (14.94421)	3200.03 (2.82121)	3395.99 (17.14719)

Table A.11:   Atomic positions of DPS molecule (in Fig. 3.21) from the optimized structure (DFT/B3LYP/STO-3G).

Atom #	x	y	z
1	1.17854301	-2.25783347	0.00000822
2	-0.77852275	-0.26652621	0.00000940
3	-1.94736087	2.26999640	0.00000804
4	-1.80902496	3.82024748	-0.00000525
5	-0.55947980	1.55634947	-0.00000093
6	1.00401087	-0.70699293	0.00000944
7	2.67695067	-2.67918172	0.00000650
8	0.67962158	-2.68213606	0.88916374
9	0.67962230	-2.68213340	-0.88915054
10	-2.52134439	1.95445665	0.88914919
11	-2.52136444	1.95444272	-0.88911445
12	-2.80253923	4.29481386	0.00000519
13	-1.26241152	4.16255785	0.89225522
14	-1.26243651	4.16254496	-0.89228654
15	0.01372463	1.87828783	-0.88972482
16	0.01374047	1.87829573	0.88970915
17	1.50798861	-0.28488040	-0.88973127
18	1.50798888	-0.28488244	0.88975209
19	2.76815330	-3.77637065	0.00000815
20	3.19089224	-2.28933348	-0.89220145
21	3.19089391	-2.28933133	0.89221510

Table A.12: Calculated normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for DPS molecule for three different levels of computation. (Modes with first five largest intensities (marked by star) in each group have been included).

Mode #	AM1	PM3	RHF	DFT
14	786.27(4.373)*	775.97(3.439)*	814.03(0.020)	800.73(1.162)
15	805.78(6.713)*	811.41(2.453)*	821.26(11.485)	801.89(15.104)
28	1243.40(1.946)	1134.22(1.720)	1383.42(46.008)*	1317.17(21.405)
30	1263.98(0.439)	1204.35(2.476)*	1465.65(0.001)	1368.66(0.021)
31	1269.57(4.230)*	1212.46(1.790)	1467.91(0.195)	1370.73(0.175)
33	1388.18(4.298)*	1336.11(1.713)	1509.07(0.458)	1407.93(0.733)
36	1394.56(0.275)	1403.96(2.029)*	1637.71(15.225)	1549.15(17.117)
44	3010.12(0.642)	2954.03(0.534)	3196.12(37.454)*	2958.43(20.362)
46	3038.84(4.654)*	2978.37(2.169)*	3218.43(11.418)	2989.07(12.693)
49	3062.23(0.036)	3033.53(1.151)	3241.82(35.082)*	2992.27(3.497)
50	3064.72(0.922)	3054.70(1.648)	3242.28(2.936)	3000.36(27.862)*
51	3064.74(0.011)	3056.87(0.060)	3243.38(1.229)	3001.43(26.144)*
52	3086.24(0.999)	3079.64(0.001)	3258.93(47.138)*	3024.87(15.373)
54	3109.67(1.948)	3085.55(0.081)	3265.42(32.649)	3064.20(30.834)*
55	3110.95(0.576)	3085.57(0.080)	3266.49(23.497)	3066.56(29.834)*
56	3157.44(0.089)	3182.15(0.168)	3302.74(59.555)*	3068.81(7.131)
57	3157.44(0.036)	3182.15(0.074)	3303.59(0.922)	3069.56(79.977)*

Table A.13: Atomic positions of cyfluthrin molecule (in Fig. 3.31) from the optimized structure (PM3).

Atom label	x	y	z
1	0.87776	-3.42789	0.48704
2	0.14231	-1.60837	-0.94642
3	-0.83448	2.68920	0.17868
4	1.73292	0.45548	-2.54437
5	0.49840	2.29686	6.41609
6	0.35725	2.83417	4.82701
7	-0.95159	3.83935	4.54071
8	1.24156	2.47076	3.89149
9	1.16236	2.91723	2.49271
10	1.10572	1.86246	1.41276
11	2.34355	2.72451	1.56191
12	2.63024	3.80623	0.55763
13	3.59238	2.04091	2.04956
14	0.17223	2.02248	0.25831
15	0.33585	1.35819	-0.93046
16	1.48653	0.52139	-1.10159
17	1.94249	0.39797	-3.68150
18	1.25431	-0.86805	-0.54833
19	2.16887	-1.39738	0.36194
20	1.98813	-2.67429	0.87983
21	-0.05538	-2.89221	-0.43157
22	0.70561	-4.65881	0.98927
23	-1.11748	-3.71817	-0.73554
24	-2.01506	-3.23455	-1.69177
25	-1.78302	-3.49231	-3.04631
26	-2.75368	-3.12774	-3.97267
27	-3.19577	-2.62515	-1.25691
28	-4.15269	-2.27016	-2.20141
29	-3.93244	-2.51850	-3.55280
30	-0.57621	-1.18202	-1.66110
31	2.08843	1.81076	4.14537
32	0.54542	3.83100	2.35986
33	1.18022	0.79582	1.71459
34	1.71326	4.27761	0.17857
35	3.24629	4.59592	1.00757
36	3.17962	3.40185	-0.30345
37	3.96347	1.33129	1.29688
38	4.39013	2.76725	2.25129
39	3.41131	1.48162	2.98310
40	2.38682	0.99556	-0.62531
41	3.04091	-0.80781	0.67425

Atom #	x	y	z
42	2.71027	-3.09179	1.59080
43	-0.85601	-3.97767	-3.37143
44	-2.58727	-3.32201	-5.03761
45	-3.36720	-2.43699	-0.19116
46	-5.08311	-1.79173	-1.87743
47	-4.69056	-2.23308	-4.28982

Table A.14: Calculated normal mode frequencies of vibrations (in cm^{-1}) and integrated infrared band intensities (in km/mol) (in parenthesis) for cyfluthrin molecule for AM1 and PM3. (Modes with first ten largest intensities (marked by star) in each group have been included).

Mode #	AM1	PM3
53	846.45 (19.926)	792.86 (24.478)*
65	998.52 (48.366)*	974.77 (1.067)
90	1349.50 (40.211)*	1268.28 (28.778)*
91	1365.40 (20.076)	1289.29 (25.774)*
98	1403.29 (1.462)	1388.11 (36.217)*
99	1416.50 (10.682)	1394.01 (85.008)*
102	1461.22 (29.433)*	1408.52 (1.685)
103	1468.68 (11.675)	1435.81 (28.410)*
104	1496.89 (217.708)*	1456.35 (3.785)
107	1623.75 (126.946)*	1573.23 (68.806)*
110	1698.25 (59.626)*	1703.32 (39.247)*
116	2072.36 (98.891)*	2001.15 (90.785)*
128	3162.58 (29.466)*	3067.13 (20.880)
129	3180.75 (9.757)	3071.44 (34.428)*
132	3192.05 (66.691)*	3078.99 (0.974)
134	3196.60 (46.271)*	3157.08 (0.555)
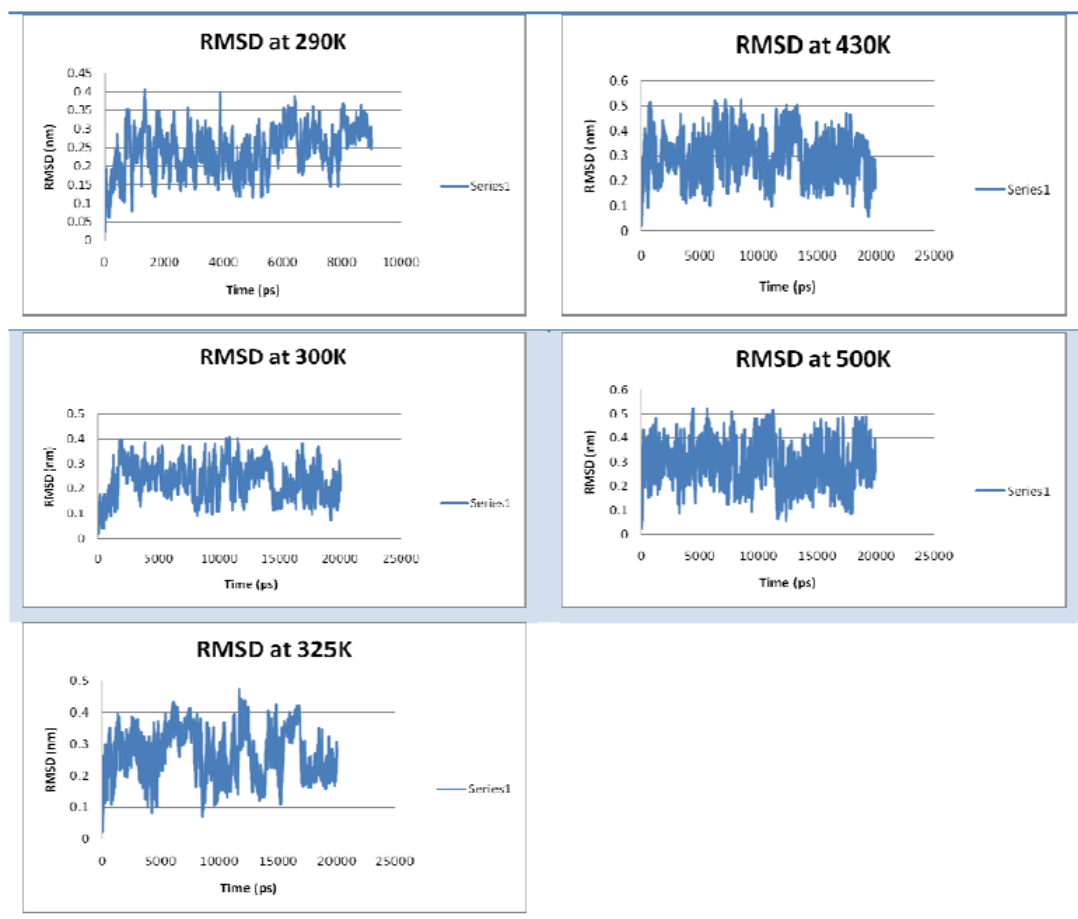
Figure A.1: RMSD of the backbone atoms of the immunoglobulin peptide at five different temperatures.

DDATKT

a)

b)

c)

112

DDATKTF

d)

e)
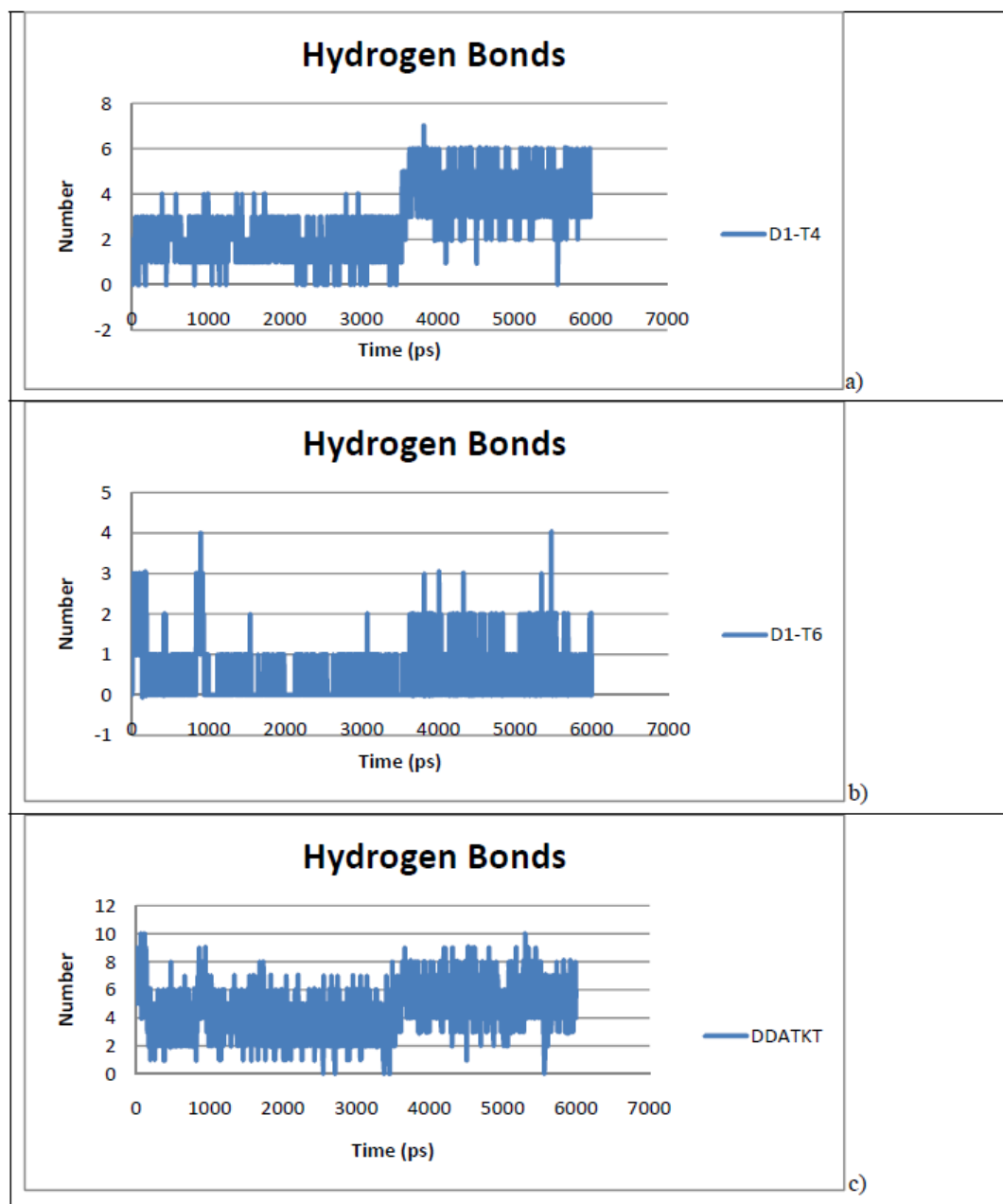
f)
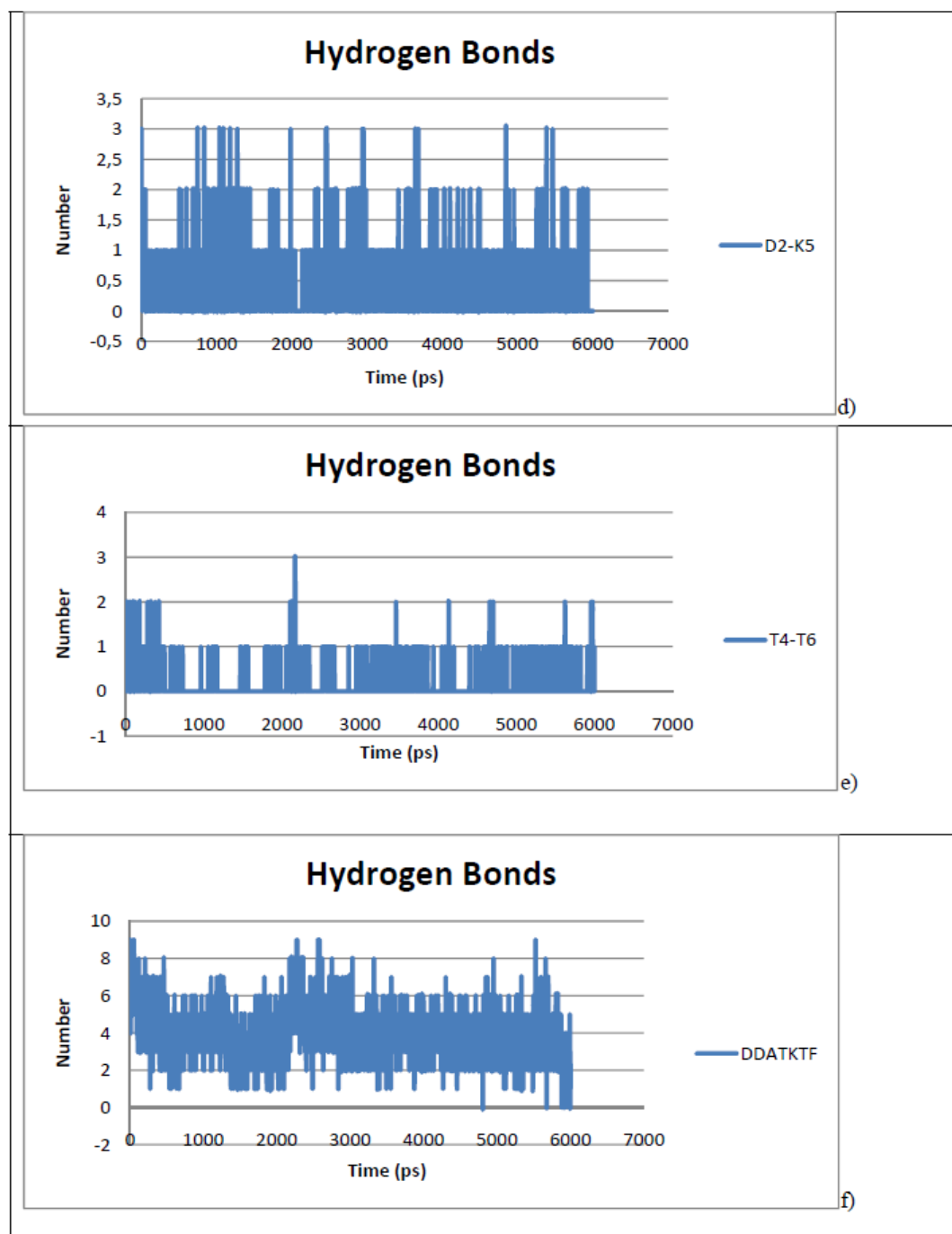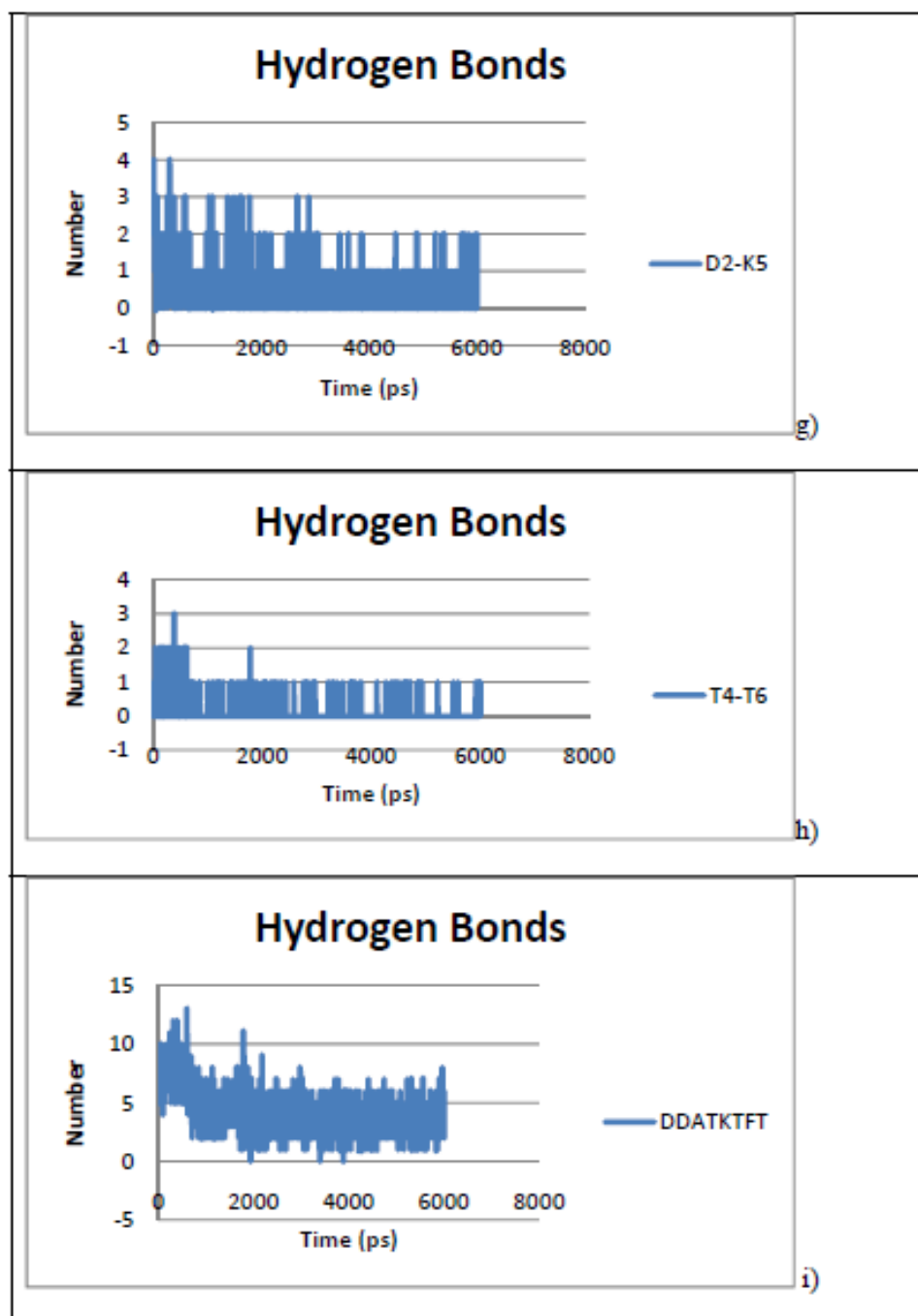
DDATKTFT

Figure A.2: Hydrogen bond time series: a-c) DDATKT a) D1-T4, b) D1-T6, c) all; d-f) DDATKTF d) D2-K5, e) T4-T6, f) all; g-i) DDATKTFT g) D2-K5, h) T4-T6, i) all
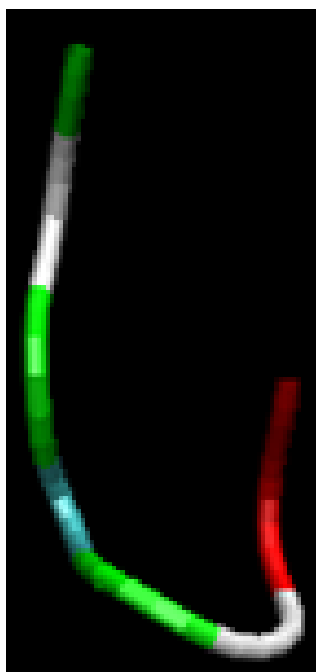
Figure A.3: The native structure of the DDATKTFT peptide taken from the PDB.
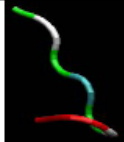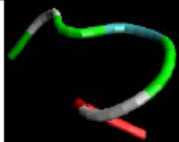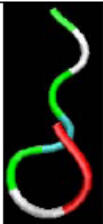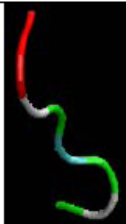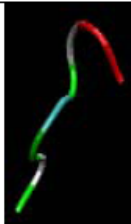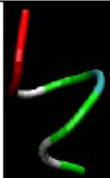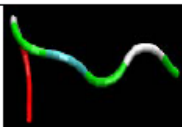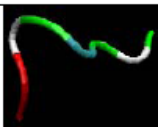
290K	300K	325K	430K	500K
2ns	4ns	4ns	4ns	4ns
4ns	8ns	8ns	8ns	8ns
6ns	12ns	12ns	12ns	12ns
8ns	16ns	16ns	16ns	16ns
9ns	20ns	20ns	20ns	20ns

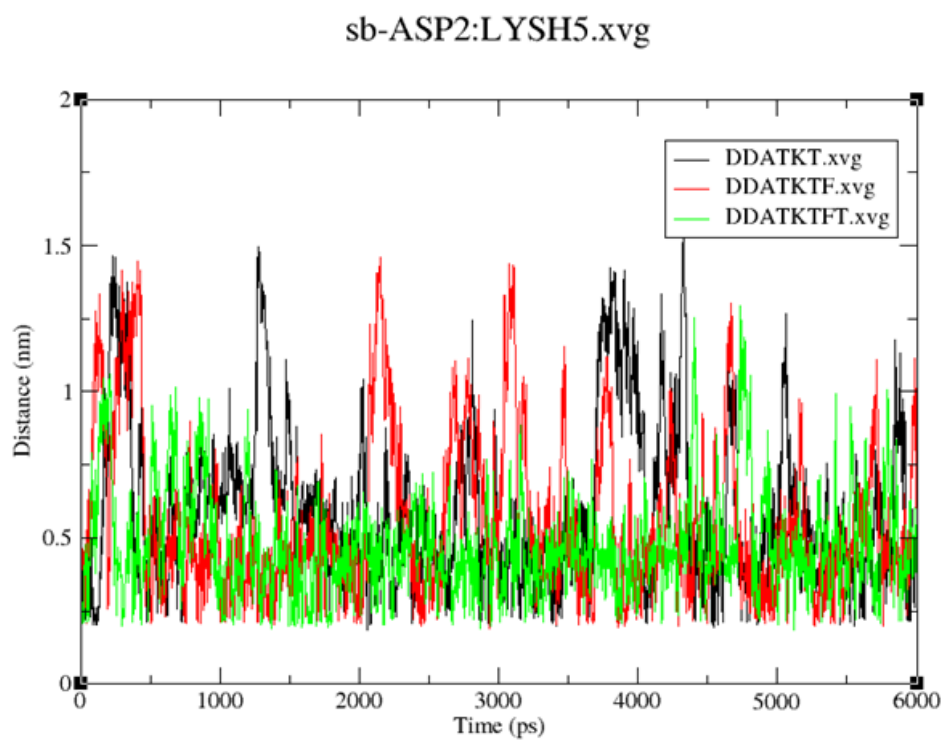Figure A.4: Snapshots of the DDATKTFT peptide at different temperatures.

116

Figure A.5: Comparison of the salt bridge ASP2-LYS5 (D2-K5) for the three different variants of the immunoglobin sequence.
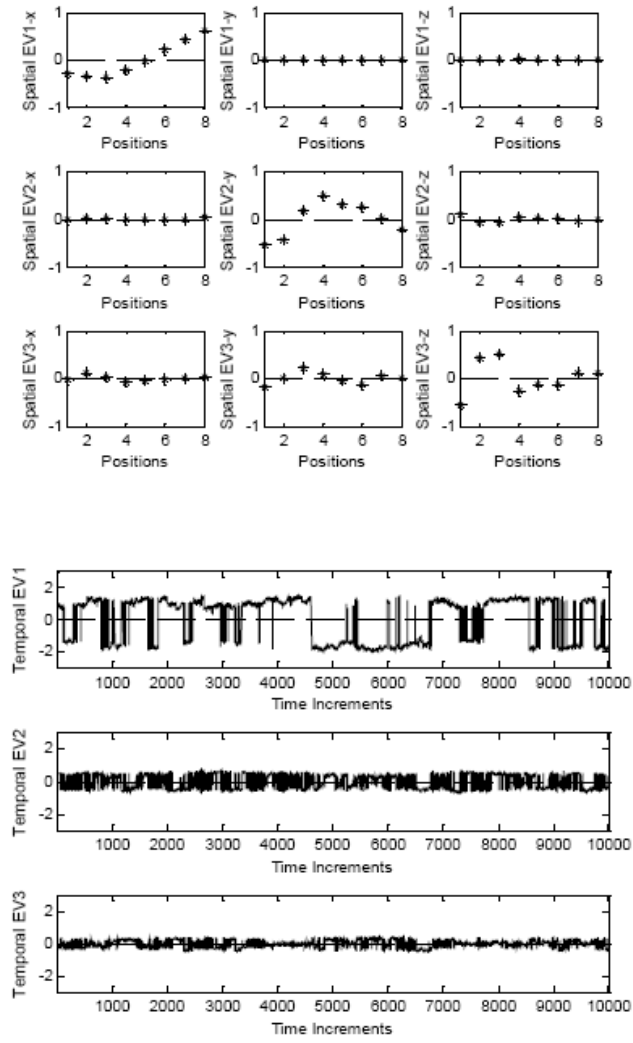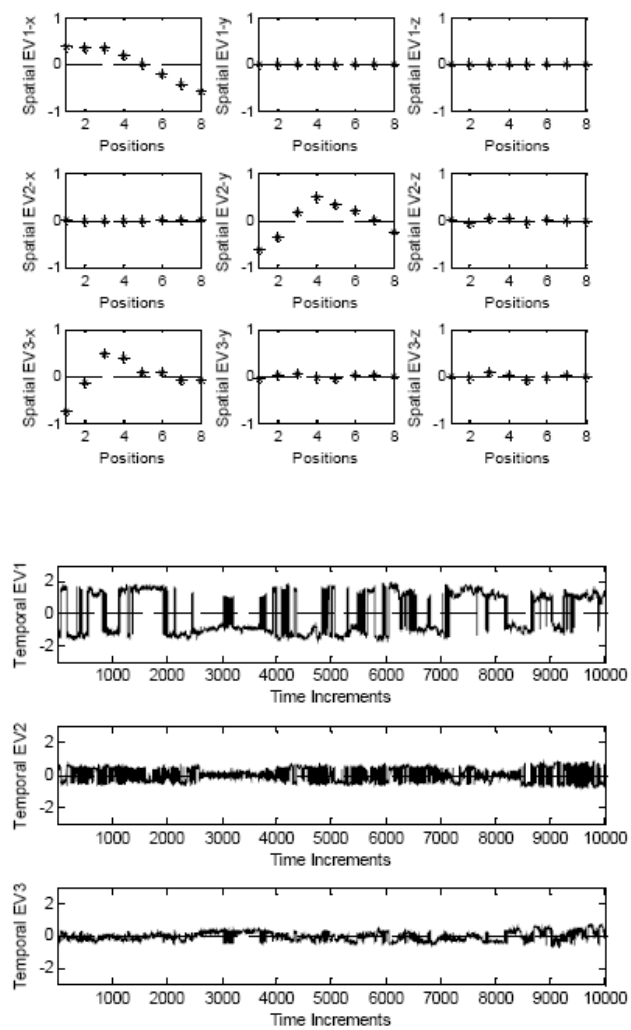
Figure A.6: Eigenmodes of protein G at 300K
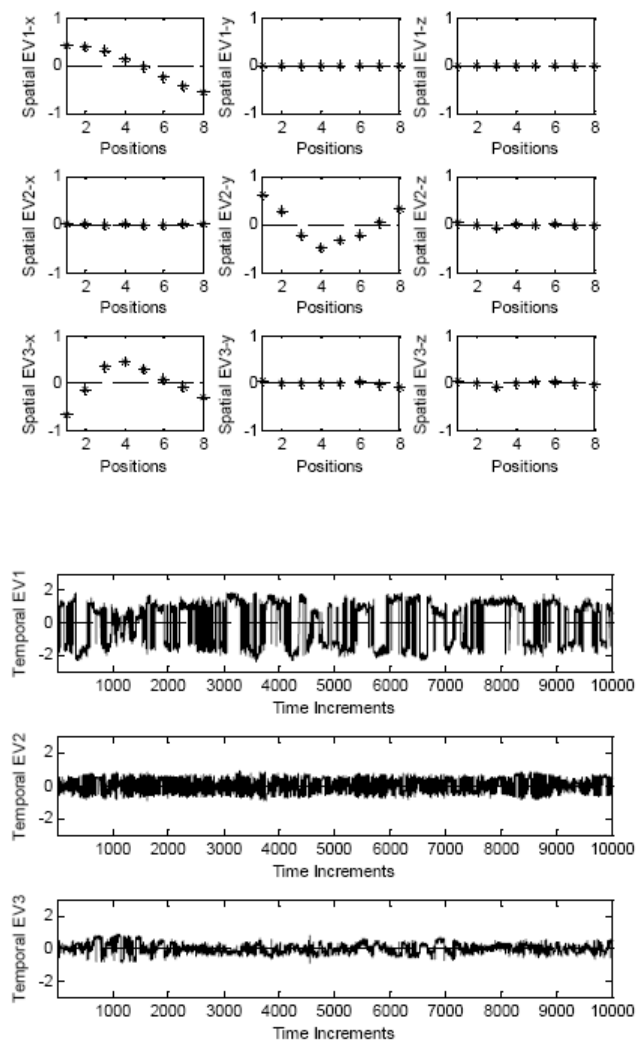
Figure A.7: Eigenmodes of protein G at 325K

Figure A.8: Eigenmodes of protein G at 430K

APPENDIX B

PRINCIPAL COMPONENT ANALYSIS

MD simulations produce a large amount of data on the dynamics of the protein molecule during folding or unfolding processes. Simple time-series plots or 3D movies, while useful, are not very effective means to extract correlations among backbone atoms and identify key mechanistic patterns. Principal Components Analysis (PCA) or Karhunen-Loeve Expansion (KLE) has been used extensively to explain the salient features of MD trajectories [24, 25, 38, 39]. The goal of PCA is to reduce the dimensionality of large scale data sets and summarize the motion in a small number of modes. For this analysis, only the backbone C_α atoms are considered here. Before the data is analyzed the center of mass motion has to be removed to eliminate the trivial translation modes. Furthermore, we rotate all molecules into a common frame so that rotational degrees of freedom are removed and the x-direction is the direction of the largest eigenvalue of the gyration tensor. To accomplish this, we determine at every simulation step the axes of the gyration tensor and align them with the Cartesian coordinates. The resulting matrix has the dimensions Mx$3N$, where M represents the number of time steps and N is the number of residues in the peptide chain. Often, we find $M \gg 3N$. The final form of the data matrix can be expressed as the following array:

$$\mathbf{R} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & ... & y_1(t_1) & y_2(t_1) & ... & z_N(t_1) \\ x_1(t_2) & : & ... & : & : & : & : \\ ... & : & ... & : & : & : & : \\ x_1(t_M) & x_2(t_M) & ... & y_1(t_M) & y_2(t_M) & ... & z_N(t_M) \end{bmatrix} \qquad (B.1)$$

Then, the elements of the spatial covariance matrix Φ for Eq. B.1 are computed from:

$$\Phi_{ij} = \frac{1}{M} \sum_{m=1}^{M} R_i(t_m)R_j(t_m) \tag{B.2}$$

An eigenvalue decomposition of yields,

$$\Phi\phi_j = \lambda_j\phi_j \tag{B.3}$$

where λ_j is the jth eigenvalue, and ϕ_j is the jth eigenvector of Φ, respectively. ϕ_j is also referred to as the spatial eigenvector or spatial mode of the matrix. The matrix R can be expressed in terms of its finite expansion as

$$R(t_m) = \sum_{j=1}^{3N} c_j(t_m)\phi_j \tag{B.4}$$

where m=1,2,...M. The temporal mode $c_j(t_m)$ represents the time varying amplitude of the spatial mode and is computed by projecting the data onto the spatial modes

$$c_j(t_m) = R^T(t_m)\phi_j \tag{B.5}$$

If this calculation is repeated for all sampling times, m=1,2,...M, one gets the time series $c_j = [c_j(t_1)c_j(t_2)...c_j(t_M)]$ for the jth modal amplitude. The modal amplitudes c_j are zero mean, and orthogonal, i.e., $< c_j(t) >= 0$ and $< c_i^T c_j >= \lambda_i \delta_{ij}$.

The PCA, or KLE, will have K modes (eigenvector directions) and each eigenvalue measures the mean amplitude of projection of its corresponding mode and establishes the relative importance of the mode. Among the class of all linear expansions, KLE is optimal in the sense that, on a subspace of lower dimension $K \ll 3N$, it offers optimal fidelity. As each mode explains a certain amount of variance associated with the data, one can calculate the cumulative variance attributed to choosing a small number of modes that would capture the key dynamic information:

$$\%CumVar = 100x\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{3N} \lambda_i} \tag{B.6}$$

One can retain only the first K modes that extract the important trends and filter out the details deemed insignificant by the user. For analysis, it is important to determine which residues contribute to each mode, thus possibly identifying coherent motions by different parts of the peptide. Furthermore, the temporal mode represents the time-series of each mode (eigenvector) that elucidates how a specific motion is distributed throughout the folding/unfolding process.

VITA

Personal Data

Surname, Name: Tekin, Emine Deniz

Date of Birth: December 22, 1981

Date of Place: Ankara, Turkey

E-mail: dcalisir@metu.edu.tr

Education and Professional Record

Fall 2007–Fall 2008 Department of Chemical Engineering and Material Science , U. of California, Davis

Fall 2002–Now PhD, Department of Physics, Middle East Technical University

Fall 1998–Spring 2002 BS, Department of Physics, Kahramanmaras Sutcu Imam University, Turkey

Schools & Conferences

- E.D. Calisir, S. Erkoc, H. Yildirim, A. Kara, T.S. Rahman, M. Selvi, F. Erkoc, *"Theoretical Comparative Study of the Structure, Dynamics and Electronic Properties of Five Ally Molecules: Allicin, Methyl Propyl Disulfide (MPD), Allyl Methyl Sulfide (AMS), S-allyl cysteine (SAC) and S-allyl mercaptocysteine (SAMC)"* , E2006 APS March Meeting, March 13-17, 2006; Baltimore, MD, Abstract: C1.00067. (poster presented by H. Yildirim).

- E.D. Calisir, S. Erkoc, *"Structural, electronic and QSAR properties of the cyflutrin molecule: A theoretical AM1 and PM3 treatment"*, 3rd Humboldt Conference on Computational Chemistry, June 24-28, 2006; Varna, Bulgaria, Book of Abstracts, p. 110.(poster presented by E.D. Calisir).

- 1st Workshop on computer aided drug design & development, Feb 2-5, 2007; Koc

University, Istanbul, Turkey.

Publications

- S. Erkoc, H. Tezcan, E.D. **Calisir** , F. Erkoc, *"Synthesis of bis-formazan molecule and quantum chemical calculations",* Int. J. Pure and Appl. Chem. 1, 37(2006).

- E.D. **Calisir**, S. Erkoc, H. Yildirim, A. Kara, T.S. Rahman, M. Selvi, F. Erkoc, *"Theoretical comparative study of the structure, dynamics and electronic properties of two Allyl molecules: S-allyl cysteine (SAC) and S-allyl mercaptocysteine (SAMC)",*
  Int. J. Pure Appl. Chem. 1, 47(2006). (Thesis work)

- H. Yildirim, A. Kara, T.S. Rahman, E.D. **Calisir** , S. Erkoc, M. Selvi, F. Erkoc, *"Theoretical comparative study of the structure, dynamics and electronic properties of two Allyl molecules: Allicin, Methyl Propyl Disulfide (MPD) and Allyl Methyl Sulfide (AMS)",*
  Int. J. Pure Appl. Chem. 1, 171(2006). (Thesis work)

- E.D. **Calisir** and S. Erkoc, *"Structural and electronic properties of dipropyl sulfide: A theoretical investigation",*
  Int. J. Mod. Phys. C 17, 1179(2006). (Thesis work)

- E.D. **Calisir** and S. Erkoc, *"Structural, electronic and QSAR properties of the cyfluthrin molecule: A theoretical AM1 and PM3 treatment"* ,
  Int. J. Mod. Phys. C 17, 1391(2006). (Thesis work)

- E.D. **Calisir** , O.B. Malcioglu, S. Erkoc, *"Density functional theory investigation of 1-chloroborepin and propargyl benzene planar molecules",*
  Int. J. Pure Appl. Chem. 1, 509(2006).

- N. Chakraborti, R. Jayakanth, S. Das, E.D. **Calisir** , S. Erkoc, *"Evolutionary and genetic algorithms applied to $Li^+ - C$ systems: Calculations using differential evolution and particle swarm algorithm",*
  J. Phase Equilibria and Diffusion 28, 140(2007).

- E.D.C. **Tekin**, F. Erkoc, I. Yildiz, S. Erkoc, *"Quantum chemical calculations of warfarin sodium, warfarin and its metabolites"* , Commun. Comput. Phys. 4, 161(2008).

125

- E.D. **Tekin** and S. Erkoc, *"Structural and electronic features of the ubiquinone and ubiquinol molecules: Molecular dynamics and quantum chemical treatments"* , Molec. Simul. (in press).

- A. Maass, E.D.**Tekin**, A. Schuller, A. Palazoglu, D. Reith, R. Faller *"Comparing the Folding and Unfolding Characteristics of Short Beta Strand Peptides under Different Conditions and from Different Starting Points "*,
Biochim. et Biophys. Acta - Proteins and Proteomics. 1804, 2003 (2010). (Thesis work)