A Computational Study of Biological and Optical Materials

by

©Shaheen Fatima

A Thesis submitted to the Department of Chemistry in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Memorial University of Newfoundland

May 2014

St. John's Newfoundland

Abstract

This thesis reports computational studies about protein-ion binding in part I, and optical properties of organic materials in part II.

It is very difficult to investigate a whole protein computationally. So here I proposed smaller models to probe protein-ion binding: a short triple helix (triple chain), a short peptide chain, and individual amino acids. The binding energies, and particularly the differences in binding energy between Na^+ and K^+ ions, do depend on the model and the constraints. I have applied these models to understand experimental observations about the distinct roles of Na^+ and K^+ in collagen aggregation and fibrillogenesis. I have calculated the binding energies for the Na^+ and K^+ with several key amino acids in collagen, selected by analysis of collagen sequence, using density functional theory (DFT).

In part II, I have focused on first and second hyperpolarizabilities of anthraquinoid-type π-extended tetrathiafulvalene, referred to as TTFAQ, and its analogues. This project has employed a wide range of functional groups to exploit the electron donor capability of TTFAQ in order to explore the hyperpolarizabilities of its derivatives. I have assessed size and charge distribution metrics as predictors for NLO response of the TTFAQ derivatives.

Acknowledgements

First I would like to thank Memorial University and the department of chemistry and for giving me the opportunity to pursue my degree. I would like to thank my advisor, Professor Erika Merschrod for accepting me into her group, for her encouragement, her guidance and inspiration during my degree. Prof. Merschrod is always there when we need her. I have learned so much from her. I would like to thank Professor Paul Mezey and Professor Yuming Zhao who served on my Ph.D. committee. I am thankful for their guide, help and constructive suggestions on my thesis. I would like to thank the Prof. Merschrod's group.

I would also like to thank Ms. Mary Flinn and late Ms. Viola Martin for their help. I would like to thank Ms. Linda Thompson for helping me in initial stage of my computational endeavor.

I would like to thank my parents (Mrs. Zarina Wali and Mr. Wali Mohammad) and my brothers and sisters for their continuous support and faith in me. Last but not the least, I am grateful to my husband, Ibrahim Ali Shaik and kids Maryam, Ismaeel and Qasim for their support and encouragement. I am also thankful to all my friends who helped me in babysitting my kids during the period of degree.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant for E.F.M.S), the Canada Foundation for Innovation, and the Industrial Research and Innovation Fund and Memorial University.

For computational resources ACEnet and WestGrid are acknowledged.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A	Acceptor
Ala	Alanine
AQ	Anthraquinone
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
BE	Binding energy
β	Beta (second order polarizability)
$\lvert\beta\mu\rvert$	Beta measured along the dipole moment direction
BSSE	Basis set superposition error
Cys	Cysteine
D	Donor
DFT	Density functional theory
γ	Gamma (third order polarizability)
Gly	Glycine
HF	Hartree-Fock
His	Histidine
HOMO	Highest occupied molecular orbital
Hyp	Hydroxyproline
Pro	Proline
LUMO	Lowest unoccupied molecular orbital
MD	Molecular dynamics
MM	Molecular mechanics
NLO	Non-linear optical
OP	Overlap population
PCM	Polarizable continuum model
QM	Quantum mechanics
Trp	Tryptophan
TTFAQ	Anthraquinone-type π-extended tetrathiafulvalene
TTF	Tetrathiafulvalene
Tyr	Tyrosine
ZPVE	Zero point vibrational energy

Part I

Biological Materials

Chapter 1

Introduction

My area of interest is protein-ion binding, and the system I will talk about here is type-I collagen with Na^+ and K^+ ions. I knew about collagen as a cosmetic material before starting work with Prof. Erika F. Merschrod S. in 2005. When I joined the Merschrod group I came to know that collagen is a not only an anti-wrinkle material but a far more important part of the body. It is a main constituent of our muscle and bones.[1] Collagen is made of small monomers which self assemble to form fibrils.[2] In our group my co-workers experimentally found that fibril formation can be affected by the presence of Na^+ and K^+ ions.[3] This has motivated my theoretical work in this part of the thesis, which deals with the protein-ion association.

1.1 Collagen and ions

In humans and other animals, skin, cartilage, bones, teeth and eyes contain tissue made of collagen. More than 20 types of collagen are known. Here I have studied type-I collagen which is a $\approx$ 300 nm long and $\approx$ 1.5 nm thick right handed triple helix. It has three polypeptide chains which are left handed helices and each chain has $\approx$ 1000 amino acid residues.[1, 4] Out of the three strands there are two identical

α1 (I) chains and one α2 (I) chain.[5] In each chain glycine (Gly) occupies every third position. Along with glycine there are other amino acids which are present in repeating sequence of triplets (Gly-X-Y), but X and Y are mainly proline (Pro) and hydroxyproline (Hyp).[4]

The collagen monomer has two regions. One is the central helical part and the other comprises the two non-helical terminal parts.[1, 5, 6] Gly is mainly present in the central helical region.[7, 8]

Na^+ and K^+ ions are biologically important ions. They play important roles in osmotic balance, information transfer via ion pumps and ion channels and stabilization of bio-molecular conformations.[9, 10] These ions are also important for bone health since they are involved in collagen aggregation.[3, 11–13] Gross, in the 1950's has observed the effect of ions on fibril formation.[14] There is some experimental work on the role of Na^+ and K^+ ions in collagen aggregation.[3, 12, 13]

1.1.1 Prior work on ion binding to amino acids

There is substantial literature focused on amino acid-ion binding, both experimental and computational, with experiments and calculations performed typically for gas phase species. Following are some examples involving Na^+ and K^+. Jensen found a range of binding geometries for Na^+ involving the amine and carboxyl moieties in glycine.[15] More detailed investigations on Na^+-glycine interactions[16] and other amino acids[17] followed from the Armentrout group. Kish et al. also studied binding of Na^+ to a broad range of amino acids with side chains and found that the side chains play a role in binding affinities.[9]

Hoyau et al. found differences in chelation of Na^+ and K^+ for glycine[18] and investigated binding affinities for Na^+ and several amino acids.[19] Aromatic amino acids present additional binding sites, investigated by Ryzhov et al; they also found

differences in binding affinities for Na^+ and K^+.[20]

These studies provide a reference point for the collagen-ion question, but their systems of study are far removed from a full protein interacting with ions.

1.1.2 Prior work on ion binding to collagen

The modeling of collagen and association of ions in the biomineralization process of collagen have been carried out in recent years with the aid of molecular dynamics (MD).[6, 21–23] Schepers and co-workers and Vitagliano et al, mentioned in their papers that the non-helical ends of collagen and ions play an important role in fibril aggregations of collagen.[6, 22] Schepers and co-workers studied calcium, phosphate and fluoride ions binding to the tail of the collagen fibers through MD.[22] Because of the flexibility of the tail region of the collagen, side chains of amino acids participate in protein-ions associations.[6]

Along this line, to really understand these types of collagen-ion interactions, we may need to consider a more protein-like model than a single amino acid.

1.2 Computational chemistry

My work involves the computational study of protein-ion binding. Computational chemistry involves chemical, mathematical, and computational knowledge to solve the chemical problems. Some common computational tasks are: electronic structure determination, geometry optimization, frequency calculations, finding transition state structures, potential energy surface mapping and determination of rate constants for chemical reactions, and thermodynamic properties such as enthalpies, entropies and Gibbs energies of reaction as well as activation energies.[24, 25]

There are different tools or methods available which are used by computational

chemists presently, including molecular mechanics and dynamics as well as semi-empirical and *ab initio* quantum mechanics. Here I will discuss those computational tools and concepts which I have used.

1.2.1 Molecular Mechanics (MM)

Molecular mechanics (MM) applies classical mechanics to model molecular geometry and motion, meaning that it does not directly involve electrons or orbital interactions. MM is one of the techniques used to obtain starting geometries of a system for use in subsequent quantum-mechanical geometry optimizations. It is fast and therefore good for large molecules. One of the challenges with this tool is that its accuracy depends on the availability of parameters for the system.[24, 26]

MM models atoms in a molecule as particles which are held together by elastic springs (modeling bonds). The regulation of bond distances and angles is determined by the characteristics of the springs, which can be formulated as forces or as potential energies. The potential energy (V) of interaction between the particles via the spring can be written as a function of geometrical parameters such as bond angles θ, bond lengths r, dihedral angles ϕ, and non-bonded distances d.

$$V = V_\theta + V_r + V_\phi + V_{nb} \tag{1.1}$$

Here the energies V_θ, V_r and V_ϕ are for bond angle, length and torsion respectively, and V_{nb} represents the energy of non-bonded interactions such as van der Waals or electrostatic interactions. In my calculations, these functions were described by the following equations:[25, 27, 28]

$$V_r = \frac{1}{2} K_r (r - r_{eq})^2 \tag{1.2}$$

where K_r is the force constant of the spring or bond, r is the bond length, and r_{eq}

is its equilibrium length. This equation assumes a harmonic potential. While some force fields include anharmonic terms, my calculations applied a harmonic bond length potential.

$$V_\theta = \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 \tag{1.3}$$

where K_θ is a force constant of angle bending, θ is the bond angle, and θ_{eq} is the equilibrium bond angle.

$$V_\phi = [V_1(1 + \cos(\phi)) + V_2(1 - \cos(2\phi)) + V_3(1 + \cos(3\phi))] \tag{1.4}$$

where V_1, V_2, V_3 are parameters capturing rotational barriers and their periodicity, and ϕ is the dihedral angle.

The non-bonding interactions considered in my calculations were van der Waals and electrostatic interactions:

$$V_{nb} = V_{vdw} + V_{es} \tag{1.5}$$

$$V_{vdw} = K_{vdw} \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^{6} \right] \tag{1.6}$$

a Lennard-Jones (6-12) potential where R is the distance between two non-bonded atoms and σ is a distance at which V_{vdw} is zero (balancing of attractive and repulsive forces). K_{vdw} is a well-depth parameter (value of the energy minimum).

$$V_{es} = K_{es} \frac{Q_A Q_B}{R_{AB}} \tag{1.7}$$

where K_{es} contains the necessary unit conversion factors as in Coulomb's law, Q_A and Q_B are charges on atoms A and B, and R_{AB} is the distance between those atoms.

The values of parameters such as K_r, r_{eq}, etc. depend on the types of atoms involved. The collection of all of the parameters in the energy terms for the applicable

types of atoms are called a force field. In my work I have used the Ghemical force field as implemented in Avogadro.[29, 30] It is an all-atom force field similar to Tripos-5.2.[31]

1.2.2 Quantum Mechanics(QM)

Quantum Mechanics(QM) attempts to solve the Schrödinger equation using various approximations.[32–34] Here I will discuss density functional theory (DFT),[35] an electron density based method. Wavefunction methods are described in Part II, Chapter 5.

DFT is a very popular method because of its efficiency and improving functionals.[36] It is based on the premise of the Hohenberg-Kohn theorems, which relate the energy of a system to its electron density $rho(r)$, and is implemented in my work through the Kohn-Sham formulation.[24] The energy E as a functional of ρ(r) is given by Equation 1.8.[24, 25, 36, 37]

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \qquad (1.8)$$

T_{ni} is the functional for kinetic energy of non-interacting electrons, V_{ne} is for nuclear-electron interactions, V_{ee} is for electron-electron repulsion, ΔT is the correction to the kinetic energy of the interacting electrons and ΔV_{ee} contains non-classical corrections to the electron-electron repulsion energy. The last two correction terms are collectively called exchange-correlation energy, E_{xc}. To get E_{xc} different approximations are used, including SVWN (Slater (exchange),Vosko, Wilk, Nusair (correlation)), B3LYP (Becke 3-Parameter (Exchange), Lee, Yang and Parr (correlation)) and BPW91 (Becke (Exchange), Perdew, Wang (correlation)). B3LYP is one of the most widely used hybrid functionals with DFT.[24, 25, 36, 37] In my first project of modeling of the collagen-ion interactions, I have used B3LYP functionals with a 6-31+G(d,p) basis set.

1.3 Part I: biological materials

In this part of the thesis I have focused on the role of these ions in collagen-collagen interactions, and the role of constraints in collagen conformation on ion binding.

We have developed a set of efficient models (constrained and unconstrained) to study collagen-ion binding, a potentially key process in the formation of numerous structural tissue such as bone and teeth. I have used DFT with the B3LYP hybrid functional and the 6-31+G(d,p) basis set to simulate protein-ion association in this part of the thesis. I have also modeled the ions binding within a continuum solvent model (with dielectric constants corresponding to water, ether, toluene, and tetrahydrofuran) to see the effect of solvents on such interactions. I have found that the solvent effect on constrained models is not pronounced.

Bibliography

[1] Nimni, M. E. *Collagen;* CRC Press: Boca Raton, Fla, 1988.

[2] Kadler, K. E.; Holmes, D. F.; Trotter, J. A.; Chapman, J. A. *Biochem. J.* **1996,** *316,* 1-11.

[3] Sun, M.; Stetco, A.; Merschrod S., E. F. *Langmuir* **2008,** *24,* 5418-5421.

[4] Gautieri, A.; Russo, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J. *J. Chem. Theory Comput* **2010,** *6,* 1210–1218.

[5] Gelse, K.; Poschl, E.; Aigner, T. *Advanced Drug Delivery Reviews* **2003,** *55,* 1531-1546.

[6] Vitagliano, L.; Nèmethy, G.; Zagari, A.; Scheraga, H. A. *J. Mol. Biol.* **1995,** .

[7] Hofmann, H.; Fietzek, P. P.; Kühn, K. *J. Mol. Bio.* **1980,** *141,* 293-314.

[8] Rich, A.; Crick, F. H. C. *J. Molec. Biol.* **1961,** *3,* 483-506.

[9] Kish, M. M.; Ohanessian, G.; Wesdemiotis, C. *Intl. J. Mass Spect.* **2003,** *227,* 509-524.

[10] Sigel, H. *Metal ions in biological systems;* volume 3 Marcel Dekker, Inc.: NY, USA, 1974.

[11] Verbalis, J. G.; Barsony, J.; Sugimura, Y.; Tian, Y.; Adams, D. J.; Carter, E. A.; Resnick, H. E. *J. Bone Miner. Res.* **2010,** *25,* 554-563.

[12] Kumar, M. R.; Merschrod S., E. F.; Poduska, K. M. *Biomacromol.* **2009,** *10,* 1970-1975.

[13] Loo, R. W.; Goh, M. C. *Langmuir* **2008,** *24,* 13276-13278.

[14] Gross, J.; Kirk, D. *J. Biol. Chem.* **1958,** *233,* 355-360.

[15] Jensen, F. *J. Am. Chem. Soc.* **1992,** *114,* 9533-9537.

[16] Moision, R. M.; Armentrout, P. B. *J. Phys. Chem. A* **2002,** *106,* 10350–10362.

[17] Heaton, A. L.; Moision, R. M.; Armentrout, P. B. *J. Phys. Chem. A* **2008,** *112,* 3319-3327.

[18] Hoyau, S.; Ohanessian, G. *Chem. Eur. J.* **1998,** *4,* 1561–1569.

[19] Hoyau, S.; Norrman, K.; McMahon, T. B.; Ohanessian, G. *J. Am. Chem. Soc.* **1999,** *121,* 8864-8875.

[20] Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 1037-1046.

[21] Kawska, A.; Hochrein, O.; Brickmann, J.; Kniep, R.; Zahn, D. *Angew. Chem. Int. Ed.* **2008,** *47,* 4982-4985.

[22] Schepers, T.; Brickmann, J.; Hochrein, O.; Zahn, D. *Z. Anorg. Allg. Chem.* **2007,** *633,* 411-414.

[23] Zahn, D.; Hochrein, O.; Kawska, A.; Brickmann, J.; Kniep, R. *J. Mater. Sci.* **2007,** *42,* 8966-8973.

[24] Lewars, E. G. *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics;* Springer: 2nd ed.; 2011.

[25] Cramer, C. J. *Essentials of computational chemistry;* John Wiley & Sons Ltd.: West Sussex, UK, 2nd ed.; 2004.

[26] Engler, E.; Andose, J.; Schleyer, P. *J. Am. Chem. Soc.* **1973,** *95,* 8005-8025.

[27] White, D. N. J. *Computers & chemistry* **1977,** *1,* 225-233.

[28] Williams, J.; Stang, P.; Schleyer, P. *Annu. Rev. Phys. Chem.* **1968,** *19,* 531-558.

[29] *Avogadro: an open-source molecular builder and visualization tool. Version 1.0.0 http://avogadro.openmolecules.net/* .

[30] Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. *J. Cheminf.* **2012,** *4,* 17.

[31] Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J Comput Chem* **1989,** *10,* 982-1012.

[32] Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102,* 6648-6653.

[33] Head-Gordon, M. *J. Phys. Chem.* **1996,** *100,* 13213-13225.

[34] Simons, J. *J. Phys. Chem.* **1991,** *95,* 1017-1029.

[35] Kohn, W.; Sham, L. *Phys. Rev. A* **1965,** *140,* 1133.

[36] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007,** *111,* 10439-10452.

[37] Geerlings, P.; Proft, F. D.; Langenaeker, W. *Chem. Rev.* **2003,** *103,* 1793-1874.

Chapter 2

Practical models for ion-collagen binding*

Protein-ion binding forms the basis for numerous important biological processes, but modeling this complex phenomenon in detail can be too resource-intensive. We develop and compare a set of small but biologically-suitable models to probe Na^+ and K^+ binding to type-I collagen, a process leading to collagen fibrillogenesis. Starting from the experimental (X-ray) geometry of a model peptide, we calculate binding energies (BEs) of Na^+ and K^+ to single amino acids (glycine, proline, and hydroxyproline), tripeptides, and triple-chain segments of a triple helix. Constrained and unconstrained geometries allow us to model central (helical) and terminal (loop) portions of the protein. Calculations use Density Functional Theory (DFT) and Hartree Fock (HF) methods.

We find that the trends in binding energies for different ions and amino acids do depend on the model and the constraints. By exploring the different outcomes between

*This chapter is in revision with the *J Comp Chem* as "Practical models for ion-collagen binding", Shaheen Fatima and Erika F. Merschrod S. Shaheen Fatima has developed the models, run the calculations, analyzed the results, prepared the first draft of the manuscript, and revised the manuscript, with feedback and suggestions from the co-author.

the models, we develop an understanding of the key features which may dominate the interactions on a whole-protein scale. Our outcomes also provide the basis for parameters for future coarse-grained models or molecular dynamics simulations.

2.1 Introduction

Collagen is one of the most abundant structural proteins in the animal kingdom and is found in skin, cartilage, bone, teeth and cornea.[1] Although it is not a metalloprotein, it does interact with metal ions.[2] Among the many metal ions which are associated with collagen and play an important role in its biological functions, we are particularly interested in looking at Na^+ and K^+ binding to type-I collagen, although the approach described herein is more broadly applicable to protein-ion binding in general.

Metal ion-amino acid complex formation has been investigated experimentally and theoretically for some time, with substantial work on free amino acids in the gas phase.[3–7] Na^+ and K^+ ions have affinities for various amino acids, with Na^+ having least affinity for glycine (Gly) among the different investigated amino acids.[3] Ryzhov observed that Na^+ binds stronger than K^+ to a range of amino acids (glycine, alanine, phenylalanine, tyrosine, and tryptophan).[8]

Amino acids are not proteins, however. For a structural protein such as collagen, one potential role for ions is in the aggregation of protein molecules to form tissue. In addition to their involvement in different biological processes like osmotic balance and information transfer via ion pumps and ion channels, Na^+ and K^+ ions have great importance in the stabilization of bio-molecular conformations.[3] In a 1960 paper, Taylor mentions that K^+ ions are in lower concentration in bone, a collagen-based composite material, as compared to Na^+ ions, and holds its greater ionic radius responsible for this.[9] Others have noted that these ions are important indicators of bone health, from

early reports by Bergstrom[10] to recent work by Verbalis and Barsony,[11] with chemical connections between Na^+ and carbonate content and decreased Na^+ content correlating with decreased bone mass. While the connection to bone is explained through interactions with the mineral component in bone, it is worth considering the potential role for collagen-ion interactions as well.[12]

Therefore, to really understand these types of collagen-ion interactions, we need to consider a more protein-like model than a single amino acid. The rope-like type-I collagen monomer is $\sim$ 300 nm long and $\sim$ 1.5 nm thick with three α-helical peptide chains which in turn have $\sim$ 1000 amino acid residues in each strand.[1] More than twenty different types of collagen have been characterized and grouped in different categories,[13] all with some triple-helical component. Type-I collagen is at the top of the list of all collagens in terms of abundance and research. Its frequent sequence of triplets (Gly-X-Y) leads to a predominance of triple-helical motif, where X and Y can be any amino acid but are often proline (Pro) and hydroxyproline (Hyp). The sequence in one of the three α-helices is different than the other two, but all three contain the key Gly-X-Y repeats which are very important for the stability of collagen's right handed triple helix.[14]

Given the size of this protein, an atomistic quantum chemical computational study would be untenable for calculating binding energies for the full molecule.[15] Coarse-graining, semi-empirical, or non-quantum approaches, however, would require parametrization,[16–18] which then brings us back to the initial question of a reliable binding energy. To address this issue, we have constructed three types of collagen models designed to bridge single amino acids and the entire protein, and we use those models to calculate ion binding. While water or the solvent environment surely also plays a role in the biological processes in question, this also adds another layer of complexity and size.[19] Furthermore, a poor choice of solvent model or an insuffi-

cient number of explicitly water molecules could introduce further errors and bias. Therefore, we have carried out our calculations in the gas phase.

The smallest model involves a single amino acid which can be constrained in the conformation found in the protein, the next model contains a short peptide chain, and the third is a short triple-chain (from a triple helix). We calculate binding energies for Na^+ and K^+ with Gly, Pro and Hyp in our various models by density functional theory (DFT), and our results show that the size and type of model makes a significant difference in binding energies. Our comparison of different models also sheds light on the role of the secondary and tertiary protein structure on ion binding, including geometrical constraints and longer-range inter-chain effects.

2.2 Models

Starting geometries came from the X-ray structure of a model collagen peptide (PDB ID: 1CAG).[20] This structure, a short triple helix analogous to that of type-I collagen, consists of three identical $(Gly\text{-}Hyp\text{-}Pro)_{10}$ peptide chains. A portion excised from the center of the triple helix served as the starting geometry. Hydrogen atoms were added and the structure was energy minimized, allowing only the H atoms to move. To best mimic the continuation of the protein chains, the termini of the amino acids were in canonical (not zwitterionic) form.

The largest model we chose is a triple-chain which has three amino acids (Gly-Hyp-Pro) in each peptide chain, giving a total of nine amino acids. We refer to this as the triple-chain model. While the constrained geometry is a portion of a triple helix, the unconstrained geometry, achieved by allowing the structure to relax during a geometry optimization, is a bit more open than a triple helix. (Such a short triple-chain outside of the context of a longer protein chain does not favour a triple helix.)

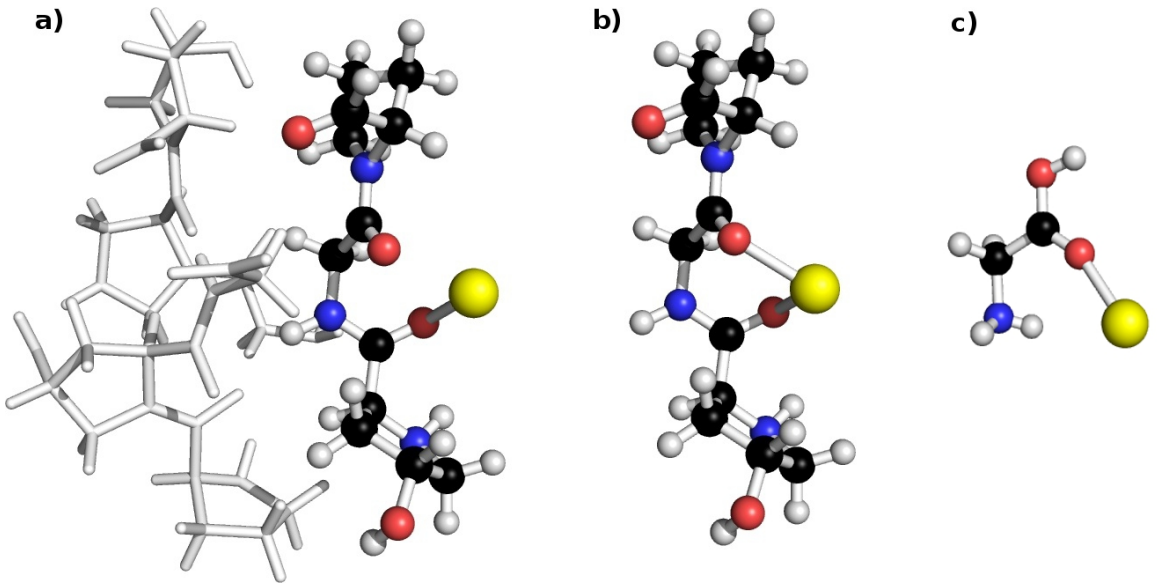

Figure 2.1: a) Triple-chain, b) peptide, and c) amino acid constrained models for glycine extracted from the X-ray structure of a collagen-like peptide, each with a bound Na^+ ion (yellow sphere). For clarity, not all atoms are shown explicitly in the largest model.

Because the chains are staggered relative to each other in forming the triple-chain, each peptide chain has a different amino acid in the middle position (Gly, Hyp and Pro) which is used to bind Na^+ and K^+ ions. Examples of this largest model are given in Figure 2.1a.

The peptide model consists of a single chain taken from the triple-chain, three amino acids long. Like with the triple-chain, here also we attached the ions to the middle amino acid, as seen in Figure 2.1b. Since each chain has a different amino acid in the middle position, we were able to obtain three peptide models from the triple-chain model above for the three amino acids of interest.

The smallest model contains a single amino acid, terminated with hydrogens in the canonical (not zwitterionic) form, as seen in Figure 2.1c. The canonical form is reasonable for the system we have investigated because in a triple helix only the carbonyl oxygen is available for ion binding. (Other potential binding sites in an

amino acid are not accessible in the true protein context.) Kapota and coworkers have reported the stability of canonical form of Gly-Na^+ (denoted as CS1 in scheme 1 of their paper[21]), which is similar to our canonical form in which metal ion binds between carbonyl oxygen and nitrogen as depicted in Figure 2.2d. The N-terminal is simply capped as NH_2, allowing for direct comparison with other amino acid models. (For effects of additional carbonyl groups from further positions on the chain, one can look to the peptide or triple-chain models described above.)

For each model, from amino acid to triple-chain, we calculated the binding energies of Na^+ and K^+ ions using two scenarios: one in which only the ion was allowed to move, constraining the protein subunit to the geometry in the model peptide; and one which also allowed the protein subunit to relax (unconstrained models). Binding energies are defined as the difference between the energy of the optimized ion-model complex and the sum of the energies of the ion and the optimized model alone. Therefore, for the unconstrained models, the binding energy includes any geometry changes to accommodate the bound ion, as would any experimental measurement of such a quantity. The constrained models do not involve any geometry changes within the protein component, mimicking the restricted geometry found within the helical part of the real protein.

There are, of course, many possible unconstrained structures for a system with so many degrees of freedom. Previous work by Wang, Ye and coworkers has detailed structures and binding energies for di-, tri-, and tetrapeptides of glycine and alanine with Na^+ and K^+, based on both experimental and theoretical results.[22, 23] We do not propose exploring or cataloging the extensive potential energy surface for our models but rather focus on energy-minimized structures which evolve from our common starting point, the X-ray structure in the collagen peptide mimic.

These studies has been carried out in the gas phase, to measure intrinsic bond

strengths between alkali metal cations and peptides.[4] Calculations with continuum solvent models SMD[24] and C-PCM[25] resulted in very similar binding energies and identical trends (data for PCM with four solvents are shown in the Supporting Information, Figure A.1). More complex treatment of the solvent would require parametrization (and hence prior knowledge) or explicit solvent molecules which would make the calculation impractical in our case. Therefore, a gas phase calculation can be a suitable alternative to more computationally expensive solvent based calculations.[26]

2.3 Methodology

Starting from initial geometries based on an X-ray structure,[20] hydrogens were added and the structures minimized using molecular mechanics with the Ghemical force field (an all atom force field similar to Tripos-5.2[27]), as implemented in Avogadro.[28, 29] The amino acids were taken in canonical (not zwitterionic) form, since they would not be exhibit the charge separation of a zwitterionic form within the protein context. We have tried different initial positions of ions around the amino acids, peptides and triple-chain (see Supporting Information, Figure A.2). There are only a few accessible binding sites in the protein context, which served as a constraint in selecting initial positions for optimization. The ions always ended up near the carbonyl oxygen in constrained models, while in unconstrained models the ions moved in between nitrogen and carbonyl oxygen. The backbone carbonyl is accessible on the surface (see Supporting Information, Figure A.3).

MacMolPlt was used to build and visualize the input and output files,[30] and figures were generated with PyMOL.[31] Quantum chemical calculations were performed with the FEB 2006 R5 version of Gamess (US).[32, 33] All geometries were optimized by density functional theory (DFT) with the B3LYP[34, 35] hybrid functional, using a

basis set of 6-31+G(d,p). DFT with the B3LYP hybrid functional are widely used and still a popular and efficient choice for studies of this type.[36, 37] Binding energies for ion-protein complexes were calculated by using the same method and basis set. Some of the calculations were performed in parallel. Heaton has reported that basis set superposition error (BSSE) correction is small for DFT calculation on alkali metal cations.[4] So we did not include it in our calculations.

Hartree Fock (HF) calculations with the same basis set were used to check for imaginary frequencies in optimized geometries, because frequency calculations with DFT require a numerical approach which is very time-consuming. However, we did carry out vibrational analyses of sample molecule-ion pairs using DFT with the above mentioned functional and basis set to calculate zero-point vibrational energies. These were at least four orders of magnitude smaller than the total energies, so we neglect them in our calculations.

2.4 Results and discussion

2.4.1 Amino acid–ion complexes

The binding energies and geometries we calculate for the unconstrained ion-amino acid complexes, given in Table 2.1, fall within the range of experimental and computational results in the literature. For these and other models, the binding energy to Na^+ is generally higher than that to K^+. This is to be expected since Na^+, with its smaller size, has a higher charge density. The bond lengths between Na^+ and the amino acid (Figure 2.2c and d) are also shorter than for K^+ (Figure 2.2a and b).

In the case of glycine, we have calculated a 195 kJ mol^{-1} binding energy with Na^+ and 144 kJ mol^{-1} with K^+ for unconstrained models at B3LYP/6-31+G(d,p). Experimental and theoretical results reported by others follow the same general trends,

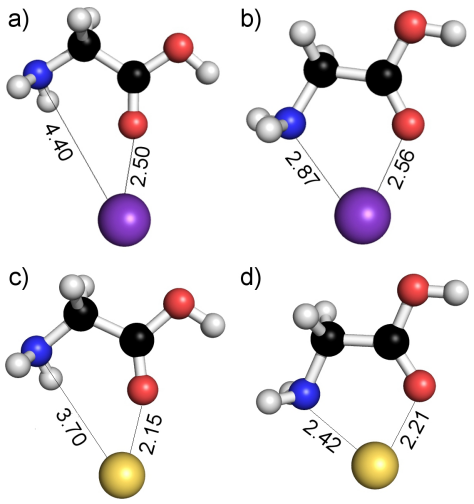

Figure 2.2: Glycine-ion complexes for the amino acid models, obtained at B3LYP/6-31+G(d,p). The glycine-ion distances are longer for K^+ (purple sphere) in the a) constrained and b) unconstrained complexes than for Na^+ (yellow sphere) in the c) constrained and d) unconstrained complexes. Notice that in the unconstrained geometries b) and d) the ion binds to both the amino N and the carbonyl O.

with Gly–Na^+ binding energies of 152 kJ mol^{-1} (MP2/6-311+G(2d,2p)),[6, 38] 168 kJ mol^{-1} (B3LYP/ 6-31+G(d)),[8] and 164 kJ mol^{-1} (experimental value)[38] and Gly–K^+ binding energies of 113 kJ mol^{-1} (B3LYP/6-31+G(d))[8] and 111 kJ mol^{-1} (MP2/6-311+G(2d,2p)).[6]

2.4.1.1 Ion binding to unconstrained and constrained amino acids

The geometries for constrained amino acids and ions are shown in Figure 2.2a) and c). The geometries (low energy conformers) of unconstrained Gly-Na^+ and Gly-K^+ have been studied previously and are given in different literatures.[6–8, 38, 39] The structures for unconstrained amino acid and ions are shown in Figure 2.2b and d, with structures similar to the structure given in literature.[6, 38] The structure of lowest energy conformer of potassiated and sodiated glycine (Gly-K^+)and (Gly-Na^+) given by Hoyau and coworkers has similar orientation as ours (Figure 2.2b, d), with the ion positioned closer to the O than to the N and with the same overall amino acid struc-

	amino acid	peptide	triple-chain
glycine			
Gly-Na^+	195	236	238
Gly-K^+	144	178	171
proline			
Pro-Na^+	203	158	351
Pro-K^+	152	155	269
hydroxyproline			
Hyp-Na^+	231	232	304
Hyp-K^+	177	173	197

Table 2.1: Binding energies (kJ mol^{-1}) for three different unconstrained models, calculated at B3LYP/6-31+G(d,p).

ture (including hydrogen orientation). The N–Na distance is 2.44 Å and the O–Na distance is 2.26 Å in sodiated glycine and the N–K distance is 2.88 Å and the O–K distance is 2.62 Å in Gly-K^+,[6] while in our case they are 2.42 Å (N–Na) and 2.21 Å (O–Na), 2.87 Å (N–K) and 2.56 Å (O–K). We have similar structures for proline and hydroxyproline amino acid complexes with ions, consistent in all models from amino acid to triple-chain. Kapota and coworkers found that a salt bridge form is more stable for proline-ion binding,[21] but this requires a deprotonated terminal carboxylic acid which is not suitable for a model of an amino acid embedded in a protein.

There are significant differences in ion binding geometries depending on whether the amino acid is constrained to the geometry in the helical protein, reinforcing the importance of considering the protein context when calculating ion binding energies. The constrained model necessarily prevents N–ion interactions because of the configuration of the hydrogens enforced by the protein geometry. In the unconstrained model, inversion at the nitrogen allows for greater interaction with the ions (see Figure 2.2b and d). As a result, ions bind to both nitrogen and carbonyl oxygen in unconstrained amino acids, while for the constrained models the ions attach to the backbone carbonyl oxygen only. Therefore, the unconstrained model has higher bind-

ing energies because of multiple binding sites. In the case of constrained amino acids, the NH_2 termination is not capped or relaxed to allow a more direct comparison with the unconstrained model.

2.4.2 Ion binding for larger models: peptide and triple-chain

The size of the model makes a significant difference in binding energies as well as binding geometries. This is in part due to changes in geometry for the unconstrained models. Another contributing factor, however, is charge redistribution due to inter-chain interactions within the protein moiety itself[40] – beyond the local binding site of the ion – for the largest, triple-chain, model. This again underscores the importance of considering the larger protein context when modeling ion-protein interactions.

While the local starting geometry is the same for the amino acids alone or within the context of the peptide or triple-chain models, the optimized geometries for the unconstrained larger models present different binding sites. In the case of the peptide model there are more intra-chain interactions with ions, as illustrated for the case of proline in Figure 2.3. As mentioned in the Models section above, there are many possible geometries for the unconstrained models. We are not proposing to explore that entire configurational space but rather to explore the effects of our constraints by relaxing them for some models. Our starting points are consistently the constrained geometries, to allow some basis for comparison.

For the unconstrained triple-chain model we see both inter-chain (neighboring chains) as well as intra-chain interactions with ions. Therefore, as we move from amino acid to peptide to triple-chain there are more binding sites available for ions, generally resulting in increasing binding energies for the larger models. This is apparent in Figure 2.4, with the exception of the proline-Na^+ case which is discussed below in section effects of side chain on binding. Wang et al. and Ye et al. also reported an

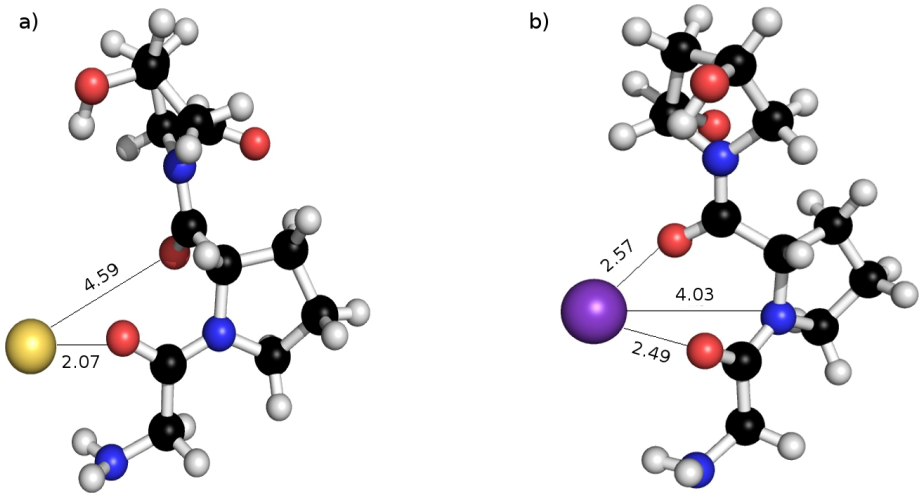

Figure 2.3: Unconstrained peptide model complexes for a) proline-Na^+ (yellow) and b) proline-K^+ (purple), obtained at B3LYP/6-31+G(d,p). Distances are in Å. Notice that the K^+ fits the natural bite of proline, while Na^+, because of its small size, requires more backbone distortion.

increase in binding energy with increased chain length for peptide–Na^+ complexes, although these systems are single-chains and hence cannot include inter-chain interactions.[22, 23]

Interestingly, the binding energies are generally higher for the triple-chain than for the peptide for the *constrained* models as well, as seen in Figure 2.5 and Table 2.2. The local geometry is identical for the constrained triple-chain and peptide models (see, for example, Figure 2.1a and b), and there are no additional close contacts between the ion and the other chains for the triple-chain model. Inter-chain interactions leads to a shift of electron density toward the backbone carbonyl O, making the interaction with the cation more favourable and resulting in a decreased ion–O distance. Therefore, the higher binding energy for the triple-chain relative to the other models also arises from an increased negative charge on the backbone O.

Clearly adding in the other chains within the collagen triple helix is an important factor in assessing the ion interaction with the protein. In addition to our approach, which emphasized expansion of the model to include more protein chains, one could

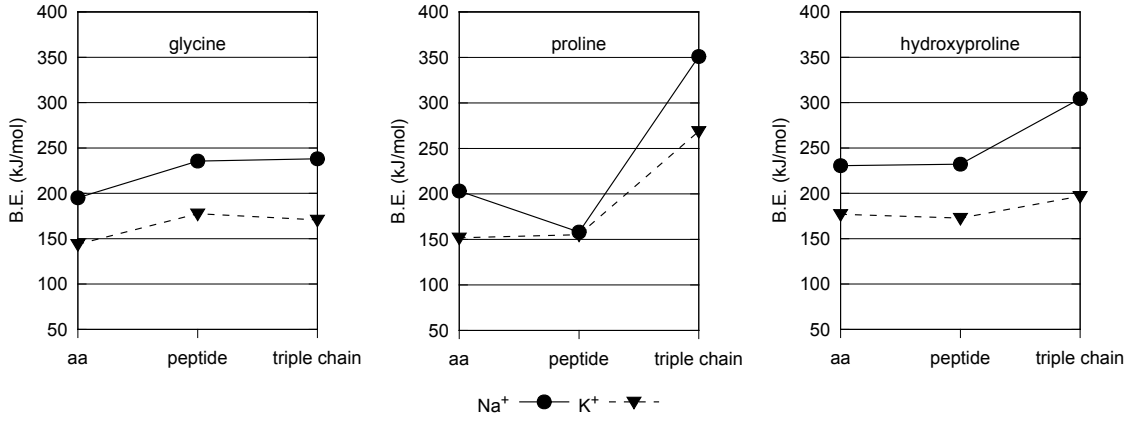

Figure 2.4: Binding energies in unconstrained models for Na^+ (solid, circle) and K^+ (dashed, triangle) ions with a) glycine, b) proline, and c) hydroxyproline, calculated at B3LYP/6-31+G(d,p). The lines are only guides to the eye.

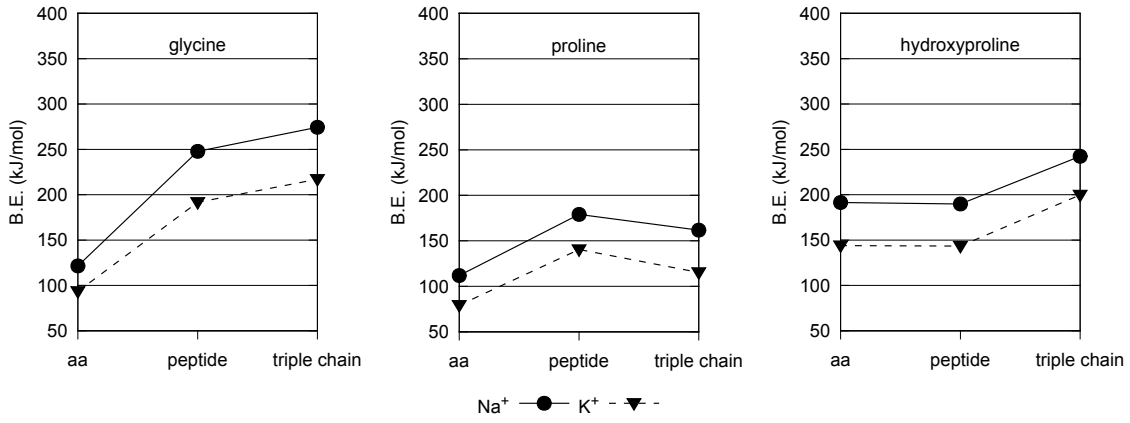

Figure 2.5: Binding energies in constrained models for Na^+ and K^+ ions with a) glycine, b) proline, and c) hydroxyproline, calculated at B3LYP/6-31+G(d,p). The lines are only guides to the eye.

	amino acid	peptide	triple-chain
glycine			
Gly-Na^+	122	248	274
Gly-K^+	93	192	217
proline			
Pro-Na^+	112	179	162
Pro-K^+	79	140	115
hydroxyproline			
Hyp-Na^+	191	190	242
Hyp-K^+	144	143	200

Table 2.2: Binding energies (kJ mol^{-1}) in three different constrained models, calculated at B3LYP/6-31+G(d,p).

also continue to extend the peptide chains to form pentamers, or create a triple-chain model with 15 amino acid subunits. For unconstrained models this really becomes unwieldy, given the large number of possible conformations. While it would be interesting to more fully model the terminal (non-helical, more open) regions of collagen with longer chains, that type of study would be better suited for other computational methods such as molecular dynamics.[41] For constrained models, we are confident that next-nearest-neighbour amino acids would not play a significant role: the terminal atoms in the peptide or triple-chain models do not sense the presence of the ion, as seen in the unchanging charge calculated for those atoms upon the addition of the ion. The exceptions are the terminal oxygen atoms, but the subsequent (next-nearest-neighbour) oxygens would be 6-7 Å away, making their participation negligible.

We also tried a bigger model of triple-chain where we took five amino acids per chain and there we found that it is very time consuming, second adding more amino acids to triple-chain trimer did not change the binding energy much. For pentamer triple-chain model with sodium ion BE is 23 kJ mol^{-1} less than with the trimer triple-chain model.

2.4.3 Effects of constraints on larger models

Constraints placed on the models obviously lead to differences in binding geometry and energy, particularly for the larger models. The importance of the constrained models lies in their connection to the larger protein structure, where said constraints are imposed by the compact triple-helical structure of the collagen monomer. The unconstrained models are more than just a theoretical exercise, however, as their more open structure can mimic the environment in the terminal regions of collagen or in cases where the triple helical structure is otherwise disrupted.[42, 43]

The primary change in allowing the peptide or triple-chain geometries to relax is that they can optimize contacts with the ions. For example in the unconstrained geometry, the glycine peptide model chain wraps around the K^+ ion, enhancing contacts with the carbonyl and hydroxyl oxygens. A backbone carbonyl oxygen from a neighbouring residue moves toward the K^+ (Figure 2.6b) creating more contacts and increasing the binding energy with respect to the constrained model (Figure 2.6a).

This effect is even more pronounced for the triple-chain model where additional inter-chain as well as intra-chain bonding is possible, as seen in Figure 2.7. This intra-chain bonding arises from the reconfiguration of the chains, leading to a new binding geometry for the ion.

2.4.4 Effects of side chain on binding

The effects of side chain on binding energy are somewhat subtle, since the binding occurs primarily with the backbone. Here we discuss the anomalously low binding energies for some proline complexes and the very high binding energies for glycine complexes.

Proline binding models do not follow all of the trends discussed above, emphasiz-

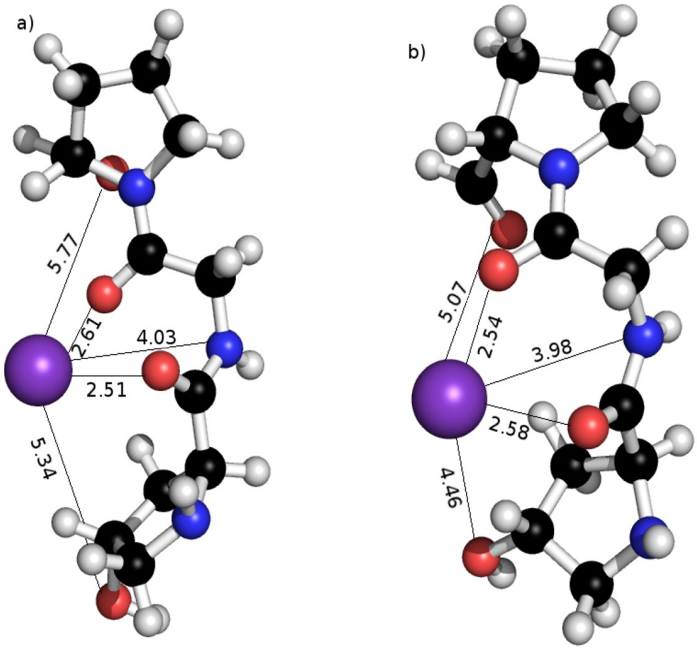

Figure 2.6: Gly-K^+ complexes for the peptide model comparing the a) constrained and b) unconstrained geometries, obtained at B3LYP/6-31+G(d,p). Distances are in Å. In the unconstrained geometry the peptide chain wraps around the ion with additional and closer contacts with backbone carbonyl O, hydroxyl O, and amino N.

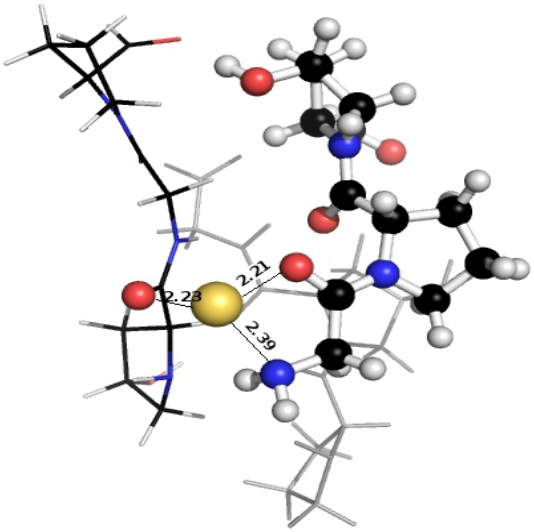

Figure 2.7: Unconstrained proline-Na^+ complex showing inter-chain as well as intra-chain interactions, obtained at B3LYP/6-31+G(d,p). Distances are in Å.

ing the significant effect that the choice of model can have on binding energy. Most dramatic is the very low binding energy of the unconstrained Pro-Na^+ peptide relative to both the Pro-K^+ binding energy and to the unconstrained amino acid model (Figure 2.4). Although the model is unconstrained, there can be an energy cost of stability of the protein moiety in maximizing the binding to the ion.

The binding energy reflects not only the strength of the ion-protein interaction but also the penalties in reorganization of the protein moiety. In the case of proline, its ring imposes some limits on the backbone flexibility. (The hydroxyproline ring demonstrates more flexibility.) As a result of this reduced flexibility, the backbone is not able to wrap as tightly around the Na^+. The larger K^+ ion fits nicely into the natural "bite" of the proline peptide model (Figure 2.3b), leading to more, closer contacts which would require more backbone distortion to achieve for Na^+ ((Figure 2.3a).

The binding energies for ions to glycine are quite high compared to proline and hydroxyproline, as seen for example in the binding energies for the constrained triple-chain models plotted in Figure 2.8. Glycine is the smallest amino acid and its H side chain does not confer the special chemical functionality seen for other amino acids. One might expect stronger cation binding to, for example, the hydroxyl group in the hydroxyproline side chain. We have seen, however, that the ion binding mode is primarily to the backbone groups which are appropriately oriented in this structure. Precisely because of its small side chain, glycine can better accommodate the ion binding to the backbone. While this certainly applies to collagen in that the triple helical domain would be tightly packed, it could also be an important consideration in studying ion binding to other proteins as well.

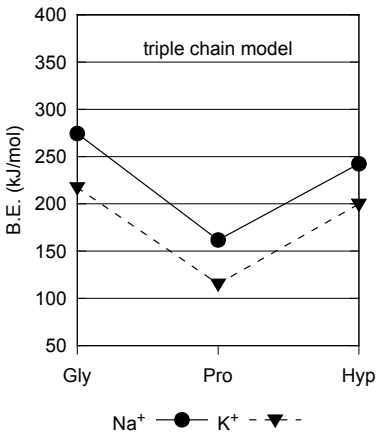

Figure 2.8: Binding energies for Na^+ and K^+ to glycine, proline and hydroxyproline in the constrained triple-chain model, showing the highest binding energy to glycine, calculated at B3LYP/6-31+G(d,p). The lines are only guides to the eye.

2.5 Conclusions

We have developed a set of efficient models to study collagen-ion binding. The binding energies we calculate confirm general trends seen previously for amino acid-ion binding, such as the higher binding energy for the higher-charge-density Na^+ than for K^+. By using protein fragments rather than only single amino acids, we have accounted for the effects of structure beyond the immediate binding environment. For example, our results bring new information about importance of inter-chain interactions in enhancing ion-protein binding, directly through providing cross-chain binding sites, or indirectly by increasing the negative charge on the external carbonyl oxygens.

Our work can provide more realistic estimates of collagen-ion binding energies than single amino acid studies, but it also highlights the importance of defining "realistic" in selecting the model, since the choice of the model can make a dramatic difference in BE, even reversing trends. For example, geometrical constraints can provide a good model for accessibility and compactness in the helical portion of the protein, while relaxing those constraints can model the non-helical parts, which pre-

sumably also play an important role in collagen aggregation. Our set of models allow consideration of the amino acid environment throughout the whole protein, with differences between constrained and unconstrained models indicating the effects of the constraint imposed in the middle (helical) region which would not be present in the terminal (loop) portions.

We have also highlighted the importance of the backbone carbonyls in binding K^+ and Na^+. Even when the amino acid side chain provides additional chemical functionality (such as the hydroxyl in hydroxyproline), the tertiary structure of the protein provides more favourable binding pockets with the backbone carbonyls, particularly around glycine. Therefore, the smallest amino acid with no extra side chain functionality can play a significant role in ion binding.

2.6 Supporting Information

Additional Supporting Information may be found in Appendix A.

Bibliography

[1] Nimni, M. E. *Collagen;* CRC Press: Boca Raton, Fla., 1988.

[2] Sigel, H. *Metal Ions in Biological Systems;* volume 3 Marcel Dekker, Inc.: NY, USA, 1974.

[3] Kish, M. M.; Ohanessian, G.; Wesdemiotis, C. *Intl. J. Mass Spect.* **2003,** *227,* 509-524.

[4] Heaton, A. L.; Moision, R. M.; Armentrout, P. B. *J. Phys. Chem. A* **2008,** *112,* 3319-3327.

[5] Hoyau, S.; Norrman, K.; McMahon, T. B.; Ohanessian, G. *J. Am. Chem. Soc.* **1999,** *121,* 8864-8875.

[6] Hoyau, S.; Ohanessian, G. *Chem. Eur. J.* **1998,** *4,* 1561–1569.

[7] Jensen, F. *J. Am. Chem. Soc.* **1992,** *114,* 9533-9537.

[8] Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 1037-1046.

[9] Taylor, T. G. *ExPERIENTIA* **1960,** *16,* 109-110.

[10] Bergstrom, W. H. *J. Biol. Chem.* **1954,** *206,* 711-715.

[11] Verbalis, J. G.; Barsony, J.; Sugimura, Y.; Tian, Y.; Adams, D. J.; Carter, E. A.; Resnick, H. E. *J. Bone Miner. Res.* **2010,** *25,* 554-563.

[12] Keinan-Adamsky, K.; Shinar, H.; Shabat, S.; Brin, Y. S.; Nyska, M.; Navon, G. *Magn. Reson. Med.* **2010,** *64,* 653–661.

[13] Gelse, K.; Poschl, E.; Aigner, T. *Adv. Drug Delivery Rev.* **2003,** *55,* 1531–1546.

[14] Gautieri, A.; Russo, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J. *J. Chem. Theory Comput* **2010,** *6,* 1210-1218.

[15] Pálfi, V. K.; Perczel, A. *J Comp Chem* **2008,** *29,* 1374-1386.

[16] Project, E.; Nachliel, E.; Gutman, M. *J Comp Chem* **2009,** *31,* 1864-1872.

[17] Roux, B.; Karplus, M. *J Comp Chem* **2004,** *16,* 690-704.

[18] Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S. *J Comp Chem* **2005,** *26,* 1612-1616.

[19] Pálfi, V. K.; Perczel, A. *J Comp Chem* **2009,** *31,* 764-777.

[20] Bella, J.; Eaton, M.; Brodsky, B.and Berman, H. *Science* **1994,** *266,* 75-81.

[21] Kapota, C.; Lemaire, J.; Maître, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004,** *126,* 1836–1842.

[22] Wang, P.; Wesdemiotis, C.; Kapota, C.; Ohanessian, G. *J. Am. Soc. Mass. Spectrom.* **2007,** *18,* 541-552.

[23] Ye, S. J.; Armentrout, P. B. *J. Phys. Chem. A* **2008,** *112,* 3587-3596.

[24] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem B* **2009,** *113,* 6378-6396.

[25] Klamt, A. *J Phys Chem* **1995,** *99,* 2224-2235.

[26] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2009,** *113,* 14231-14236.

[27] Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J Comput Chem* **1989,** *10,* 982-1012.

[28] Avogadro: an open-source molecular builder and visualization tool. Version 1.0.0 http://avogadro.openmolecules.net/.

[29] Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. *J. Cheminf.* **2012,** *4,* 17.

[30] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Mod.* **1998,** *16,* 133-138.

[31] PyMOL: The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC http://www.pymol.org/.

[32] Schmidt, M. W.; Baldridge, K. K.; Boatz, J.; Elbert, S.; Gordon, M.; Jensen, J.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.and Su, S.; Windus, T.; Dupuis, M.; Montgomery, J. *J. Comput. Chem.* **1993,** *14,* 1347-1363.

[33] Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

[34] Becke, A. D. *J. Chem. Phys.* **1993,** *98,* 5648–5652.

[35] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988,** *37,* 785.

[36] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007,** *111,* 10439-10452.

[37] Burt, M. B.; Decker, S. G. A.; Atkins, C. G.; Rowsell, M.; Peremans, A.; Fridgen, T. D. *J Phys Chem B* **2011,** *115,* 11506–11518.

[38] Moision, R. M.; Armentrout, P. B. *J. Phys. Chem. A* **2002,** *106,* 10350–10362.

[39] Klassen, J. S.; Anderson, S. G.; Blades, A. T.; Kebarle, P. *J. Phys. Chem.* **1996,** *100,* 14218-14227.

[40] Shoulders, M. D.; Raines, R. T. *J Biol Chem* **2011,** *286,* 22905–22912.

[41] Schepers, T.; Brickmann, J.; Hochrein, O.; Zahn, D. *Z Anorg Allg Chem* **2007,** *633,* 411–414.

[42] Baum, J.; Brodsky, B. *Curr Opin Struct Biol* **1999,** *9,* 122–128.

[43] Slatter, D. A.; Bihan, D. G.; Farndale, R. W. *Biomaterials* **2011,** *32,* 6621–6632.

Chapter 3

Computational models for ion-induced aggregation pathways of type-I collagen

Based on a previously-refined short triple-chain model for collagen, we have investigated the binding of Na^+ and K^+ with several important amino acids in type-I collagen, some of which exist solely in the terminal (loop) regions of the sequence and some of which predominate in the central (helical) portion. We have calculated the binding energies for Na^+ and K^+ with alanine, tryptophan, tyrosine, cysteine and histidine by density functional theory (DFT), adding to previous data on glycine, proline and hydroxyproline. For each structure we have used constrained and unconstrained triple-chain geometries. In the constrained model only the ion was allowed to move, while constraining the protein subunit to a triple helix geometry. In the unconstrained model the triple-chain and the ion were allowed to move. The unconstrained geometry mimics the non-helical (terminal) part of collagen with the side chains of amino acids in the unconstrained model providing additional binding sites for ions. Our findings

suggest that Na^+ and K^+ ions can promote different monomer interaction modalities in collagen fibril formation.

3.1 Introduction

Type-I collagen is the most abundant collagen of the animal kingdom.[1, 2] It is the main building block of skin, cartilage, bones and connective tissues. The type-I collagen molecule is about 300 nm long and about 1.5 nm thick with three left handed helical polypeptide chains which form a right handed triple helical structure.[2–4] Each monomer has two structurally different regions - a central triple helix and two terminal non helical portions. The subunits of this monomer are two identical $\alpha 1$ (I) chains and one $\alpha 2$ (I) chain.[2]

Collagen is not a metalloprotein but it interacts with metal ions.[4] For example, ions play a important role in protein aggregation. In the 1950's, Gross pointed out the role of ions in the fibril formation of collagen. He found that Li^+, I^-, Br^-, Cl^- and other ions increased the fibril formation rate.[5] In the last couple of years, the role of Na^+ and K^+ in collagen aggregation has been studied experimentally in our group and elsewhere.[3, 6–9] Other ions such as Ca^{2+} have also been investigated and their role in fibrillogenesis has been studied.[8, 10–12]

Ions play different roles in collagen aggregation. In our group, the role of Na^+ and K^+ in the formation of small collagen fibrils has been studied experimentally. It has been observed that sodium ions lead to alignment and potassium ions lead to lateral growth or thickening of collagen monomers (Na^+ lengthens, K^+ widens).[8] This observation has motivated the present theoretical study. Note that we are not invoking any specific model for long-range order of the monomers such as the quarter-stagger model, in which collagen monomers arrange in a staggered manner with 67

nm periodicity.[13]

Ions have different affinities for different amino acids, and this might explain their different roles in aggregation. We are interested in knowing if Na^+ promotes lengthening by bridging the terminal regions or inhibiting lateral growth by blocking the central regions or if K^+ promotes widening by bridging the central regions or inhibiting lengthening by blocking the terminal regions. Research about ions binding to collagen is also important to understand collagen degradation and disturbed metabolism, which occurs in osteoarthritis and osteoporosis.[2]

In the middle triple helical region, glycine (Gly) occupies every third position (X-Y-Gly) where X and Y can be any amino acid, often proline and hydroxyproline.[14, 15] Besides glycine (Gly), proline (Pro) and hydroxyproline (Hyp), there are other amino acids also present in collagen and the sequence of amino acids is quite different in the terminal regions. Based on an analysis of the amino acids sequence of type-I collagen,[14, 15] we have chosen the following amino acids: Tryptophan (Trp), tyrosine (Tyr), cysteine (Cys) and histidine (His) which are mainly present in the terminal region, alanine (Ala) which is present in the terminal as well as in the central region (see Table 3.1).

The modeling of protein aggregation is a daunting task, therefore small segments of the collagen have been studied computationally in recent years.[11, 16] We have developed a small model for collagen which preserves the possibility of intra- and inter-chain interactions while remaining computationally tractable.[17] Our prior results indicate that the primary ion-protein interactions are with the backbone, a feature which is not apparent in single amino acid-metal ion studies. Experimental and computational studies related to individual amino acid-metal ion interactions emphasize cation-π interactions,[18–24] while we have since found that in the protein context (beyond the single amino acid level), the side chains play a secondary role

with the primary binding is to the backbone.[17]

The non-helical ends of collagen and the ions present play an important role in fibrillogenesis of the collagen molecule.[10–12, 16] Schepers and co-workers studied calcium, phosphate and fluoride ion binding to the tail of the collagen fibers through molecular dynamics (MD). Because of the flexibility of the tail region, the side chains of amino acids can participate in protein-ion associations.[16] It is therefore important to consider how different the binding is for these ions in the terminal region as opposed to the central region of the collagen monomer. A more open version of our collagen model, not constrained to maintain a triple-helical geometry, is applied to represent the terminal regions.

In this paper we build on previous work[17] and calculate binding energies for Na^+ and K^+ with Trp, Tyr, His, Cys and Ala in our triple-chain models by DFT. Here we choose a triple-chain model because our previous results have shown that the size and type of model makes a significant difference in binding energies. By considering constrained and unconstrained models, we also shed light on the role of the secondary and tertiary protein structure on ion binding, including longer-range inter-chain effects.[17]

3.2 Computational approach

Starting from initial geometries based on an X-ray structure,[25] hydrogens were added and the structures minimized using molecular mechanics with the Ghemical force field (an all-atom force field similar to Tripos-5.2[26]), as implemented in Avogadro.[27, 28] The Ghemical force field was used to generate starting geometries. There are only a few accessible binding sites in the protein context, which serve as constraints in selecting initial positions for optimization. The ions always end up near the carbonyl oxygen

in constrained models, while in unconstrained models the ions lie between nitrogen and carbonyl oxygen. The backbone carbonyl is accessible on the surface of the triple-helical monomer.[17]

MacMolPlt was used to build and visualize the input and output files,[29] and figures were generated with PyMOL.[30] Quantum chemical calculations were performed with the FEB 2006 R5 version of Gamess (US).[31, 32] All geometries were optimized using DFT with the B3LYP[33, 34] hybrid functional and the 6-31+G(d,p) basis set. DFT with the B3LYP hybrid functional is widely used and still a popular and efficient choice for studies of this type.[35, 36] Binding energies for ion-protein complexes were calculated by using the same method and basis set. Some of the calculations were performed in parallel (multiple processors were used). We did not include basis set superposition error (BSSE) in our calculations since Heaton has reported that BSSE correction is small for DFT calculation on alkali metal cations.[37] All the calculations are in the gas phase. The gas phase calculations are reported to give a better understanding of intrinsic properties.[36]

Hartree Fock (HF) calculations with 6-31+G(d,p) basis set were used to check for imaginary frequencies in all the optimized geometries because frequency calculations with DFT require a numerical integration which is very time-consuming. We have not considered zero-point vibrational energies (ZPVE) in these calculations because they are much smaller than the total energies.[17]

3.3 Modeling: protein-ion binding

We took the starting structures from our previous work.[17] A hydroxyproline was replaced with other amino acids in the polypeptide structure to serve as the starting geometry. The structure was energy minimized using B3LYP/6-31+G(d,p) method,

allowing only the replaced amino acid's atoms to move. This structure was then used to bind Na^+ and K^+ ions, with either all protein atoms frozen (constrained) or all atoms allowed to move (unconstrained).

The triple-chain (part of triple helix) we chose has three amino acids (Gly-X-Pro) in a peptide chain for a total of nine amino acids in the triple-chain. In one peptide, glycine is in the middle position and in the other two peptide chains, Pro and other amino acids (Ala, Trp, Tyr, Cys, His) occupy the middle position. We chose these amino acids based on an analysis of the sequence of bovine dermal type-I collagen.[14]

Collagen has a tightly packed middle region and loose terminal regions. The terminal region has about 200 amino acids while in the central region, the number of the amino acids is nearly 1000.[16] In our previous work, we have focused on the middle part of the collagen which is tightly packed and has glycine, proline and hydroxyproline mainly.[17] In the present work, we have included other amino acids based on their availability in the terminal region of the collagen molecule. The terminal regions are more open (non-helical) and play an important role in aggregation.[16]

Based on the sequence analysis, we decide which is a helical region and which is non-helical, as follows. The helical (or central) region is defined where every third position is occupied by glycine. Where this periodicity of occurrence of glycine ends, the non-helical (or terminal) region begins. The amino acids Cys, Trp and Tyr, are mainly present in the terminal region or rarely or never found in the central region, while His shows a split between the two and Ala is primarily in the central region (see Table 3.1).

For each amino acid under study in the triple-chain, we measured the binding energies of Na^+ and K^+ ions using two scenarios: one in which only the ion was allowed to move, constraining the protein subunit to the geometry in the model; and one which also allowed the protein subunit to relax (in unconstrained models).

Amino acids	Terminal(%)	Middle(%)
Cysteine	100.00	0.00
Tryptophan	100.00	0.00
Tyrosine	88.89	11.11
Histidine	60.00	40.00
Alanine	14.32	85.68
Proline	7.93	92.07
Glycine	6.74	93.26

Table 3.1: Percentage of occurrence of amino acids in terminal and central regions of type-I collagen.[14]

3.4 Results and discussion

The binding energies for the constrained and unconstrained ion-triple-chain complexes are given in the Table 3.2 and Table 3.3 respectively. The binding energy to Na^+ is generally higher than that to K^+. This is to be expected since Na^+, with its smaller size, has a higher charge density (charge/volume). The smaller size is also seen in the shorter bond lengths between Na^+ and the amino acids as compared to K^+, which is consistent with our previous work.[17]

Amino acids	B.E with Na^+	B.E with K^+
Alanine	217	179
Tyrosine	255	209
Tryptophan	254	209
Cysteine	248	199
Histidine	246	210

Table 3.2: Binding energies (kJ mol^{-1}) in constrained models, calculated at B3LYP/6-31+G(d,p) level.

3.4.1 Binding energies: constrained and unconstrained models

We have investigated constrained and unconstrained geometries for our models to capture different aspects of the true protein geometry. The central portion of the

Amino acids	B.E with Na^+	B.E with K^+
Alanine	286	217
Tyrosine	292	257
tryptophan	331	238
Cysteine	244	217
Histidine	323	251

Table 3.3: Binding energies (kJ mol^{-1}) in unconstrained models, calculated at B3LYP/6-31+G(d,p) level.

collagen monomer is tightly packed, and there would be little opportunity for relaxation or motion when binding with ions. An unconstrained model is good to study the looser, non-helical terminal regions of the protein. Although models are referred to by their "active" amino acid, the models are always the triple-chain set of 9 amino acids.

In constrained models the affinities for Na^+ are ranked as Ala < His < Cys < Trp < Tyr. For K^+ the order of binding energies is different due to the His case: Ala < Cys < Trp < Tyr < His. (See Table 3.2. For unconstrained models (Table 3.3), affinities for Na^+ are ranked as Cys < Ala < Tyr < His < Trp. There are more differences for the unconstrained models in switching to K^+ with binding energies for Ala < Cys < Trp < His < Tyr. This is depicted graphically in Figure 3.1, which also includes data from our previous work on Pro, Hyp and Gly.[17] The shift for His is explained in the next section.

3.4.2 Side chain effects on geometry

Other researchers have identified cation-π interactions for amino acids with aromatic side chains.[18–24] These interactions are very weak in our models, especially in the constrained geometry because of the repulsion between the ion and the hydrogen of the side chain. This repulsion is assessed by examining overlap population (OP) values between the ions and those hydrogens on the side chains. The OP between ions and

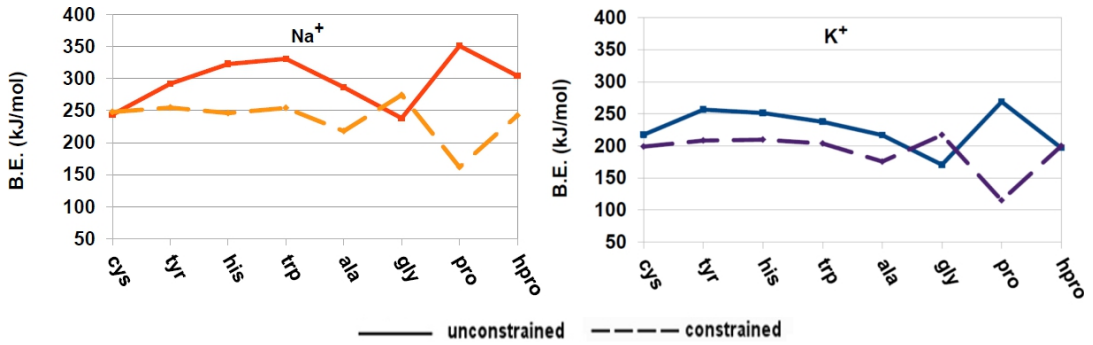

Figure 3.1: Binding energies for constrained and unconstrained models of Na^+- (LHS) and K^+- (RHS) amino acid complexes. The lines are merely guides to the eye. Data on Pro, Hyp and Gly from previous work.[17]

the side chains of Trp, Tyr and His in the constrained models are negative, indicating repulsion. Therefore the relatively short distance observed, for example, between Na^+ and H in Figure 3.4a, is repulsive, not stabilizing. The values are more negative for Trp-K^+ than Trp-Na^+ȮP analysis also helps to explain the higher binding energy for His than for other potassiated complexes, and this complex is further stabilized with a positive OP between the histidine ring and the hydroxyl oxygen of the neighboring chain.

The side chains of Trp, Tyr and His move toward the ions in the unconstrained models, leading to higher binding energies for the triple-chain unconstrained models with the Trp, Tyr and His amino acids than for the constrained models. This stabilization is also seen in the OP where, for example, Trp-Na has the most positive OP values between Na^+ and atoms in the side chain. The Tyr-K^+ system, similar to the His-K^+ constrained system, shows more positive values of OP between the Tyr ring and the hydroxyl oxygen of the neighboring chain. Figure 3.2b shows that the relaxation of neighbouring chains in the unconstrained model leads to a close and direct interaction between K^+ (shown as a purple sphere) and the oxygen to its left in a neighbouring chain (shown as sticks)

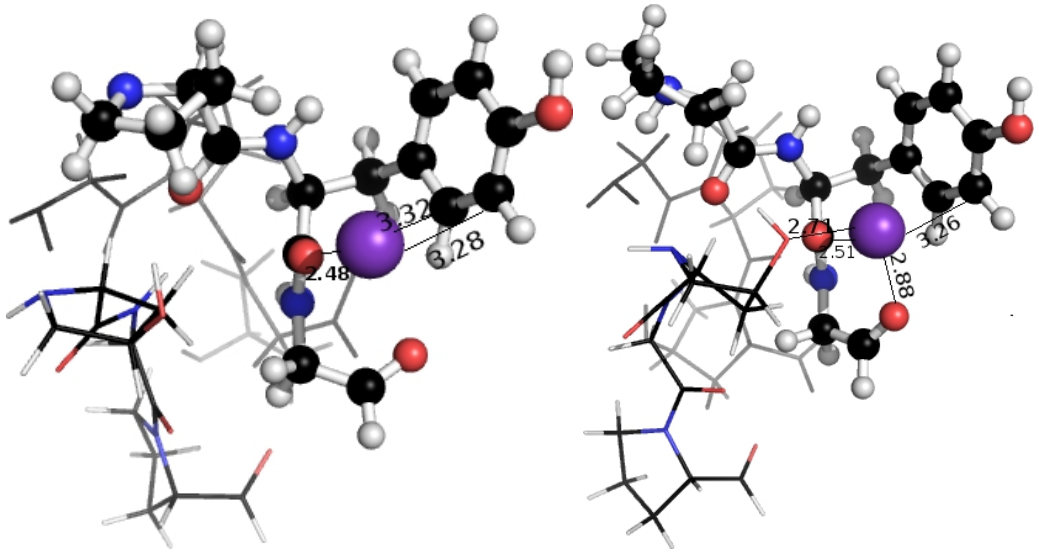

Figure 3.2: Tyr-K^+ (a) constrained and (b) unconstrained complexes, obtained at B3LYP/6-31+G(d,p). Black spheres or lines represent C, red is O, blue is N, white is H, and purple is K. Distances are in Å.

3.4.3 Different affinities for Na^+ and K^+

The order of binding energies for Na^+ and K^+ is different in constrained and unconstrained models. If we look at the unconstrained models, Tyr-K^+ has the highest binding energy among potassiated complexes, while for sodiated complexes, the binding energy is the highest for tryptophan. In the unconstrained models, the ion can facilitate or mediate protein rearrangement, leading to additional stabilization factors (higher binding energies) for certain complexes.

In the unconstrained Tyr-K^+ complex, K^+ interacts with the hydroxyl oxygen of the neighboring chain (Figure 3.2b), while Na^+ in sodiated complex of tyrosine does not (Figure 3.3b). The potassiated complex of tyrosine has inter-chain as well as intra-chain interactions.

The unconstrained sodiated tryptophan complex has the highest binding energy among the sodiated complexes because the Na^+ ion is closer to the Trp ring (Figure 3.4b). This is not the case in potassiated unconstrained tryptophan complexes (Figure

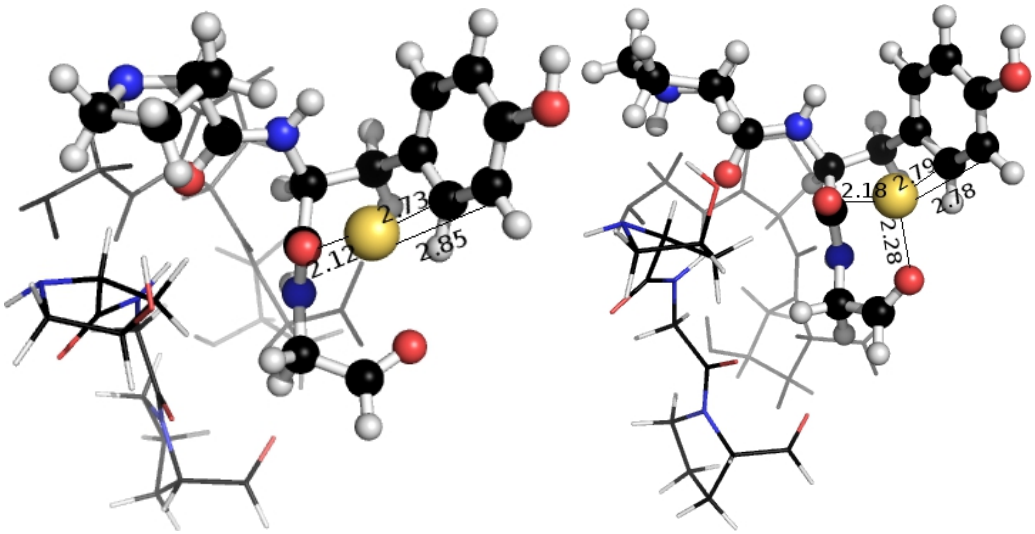

Figure 3.3: Tyr-Na^+ (a) constrained and (b) unconstrained complexes, obtained at B3LYP/6-31+G(d,p). Black spheres or lines represent C, red is O, blue is N, white is H, and yellow is Na. Distances are in Å.

3.5b). Therefore among unconstrained potassiated complexes, tryptophan ranks third after tyrosine (first) and histidine (second).

3.4.4 Binding of ions in central vs terminal regions

So we return to our question: do ions have different preferences for terminal or middle regions? From Figure 3.1, we can see an overall preference for Na^+ binding to Pro in the constrained geometry and for Na^+ binding to Gly for the unconstrained geometry. There is no obvious preference for binding K^+ nor a clear difference between amino acids predominantly in the terminal region and those predominantly in the central region. However, in considering the protein context, one can eliminate some models. For example, the constrained Trp model would represent tryptophan within a compact, helical context whereas this amino acid occurs entirely in the terminal, more open region of the monomer.

Figure 3.6 shows a subset of the data, selected to represent the majority (or

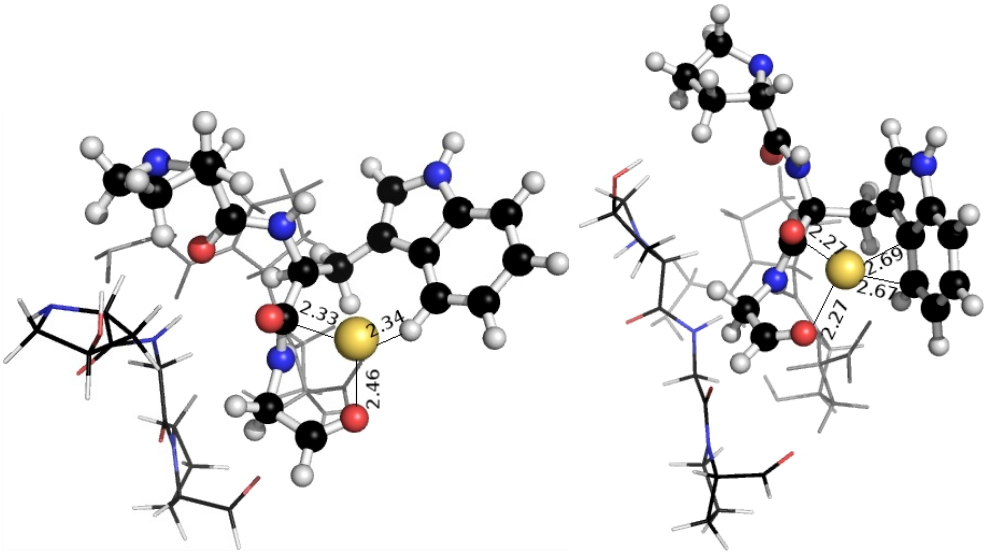

Figure 3.4: Trp-Na^+ (a) constrained and (b) unconstrained complexes, optimized at B3LYP/6-31+G(d,p). Black spheres or lines represent C, red is O, blue is N, white is H, and yellow is Na. Distances are in Å.

sole) geometrical context for each amino acid considered. Here we see some clearer differences. While in general, the Na^+ binding energies are higher, and the binding energies for unconstrained complexes are higher as would be expected, a more subtle difference emerges within these data. Na^+ shows a stronger preference for binding to the terminal region, whereas the preference is not so strong for K^+.

We propose that Na^+ could promote elongation by binding to the looser terminal regions and bridging one monomer to another longitudinally. The lower binding energies for Na^+ and the compact central amino acids would reduce the tendency to bind there. The experimental data shows more lateral fibril growth for K^+ than for Na^+, and our data indicates more comparable binding energies for K^+ in the central regions. Furthermore, the larger K^+ ions would be better able to bridge the more compact regions which would be less likely to interdigitate.

A higher binding affinity does not necessarily mean that an ion will promote aggregation. Ion binding could actually block aggregation through a local accumula-

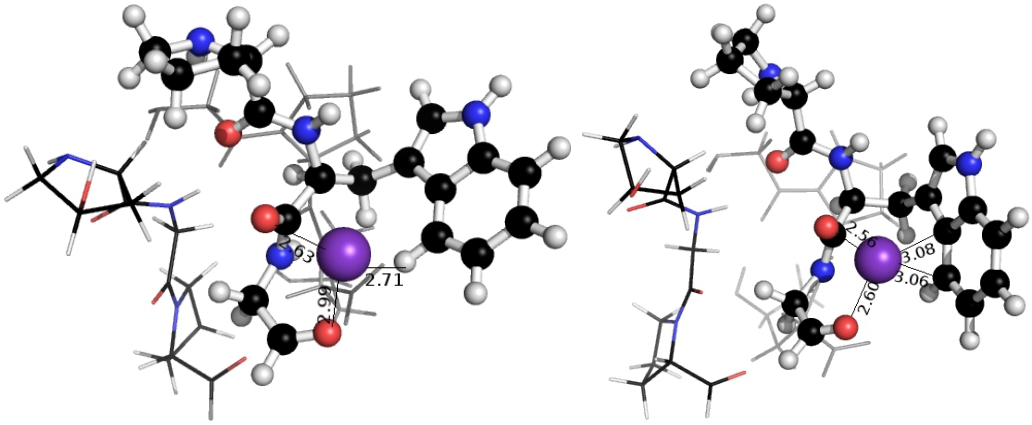

Figure 3.5: Trp-K^+ (a) constrained and (b) unconstrained complexes, obtained at B3LYP/6-31+G(d,p). Black spheres or lines represent C, red is O, blue is N, white is H, and purple is K. Distances are in Å

tion of positive charge in two regions which could otherwise interact but would now be electrostatically repulsive. However, the binding geometries which we present in this paper provide room for further coordination in the presence of another protein monomer.

3.5 Conclusions: a distinct role for ions in aggregation

In summary, we can gain an understanding of even very complex biological assemblies with a careful choice of a model. Our constrained model is efficient to apply and represent the compactness found in the triple-helical portion of the collagen monomer. By also considering the unconstrained model, however, we get a better picture of the whole monomer, including its terminal regions. In general, the ions are binding to the protein backbone, but in the terminal region, the side chains for Trp, Tyr and His also contribute to the binding.

We have applied these models to understand experimental observations of thin

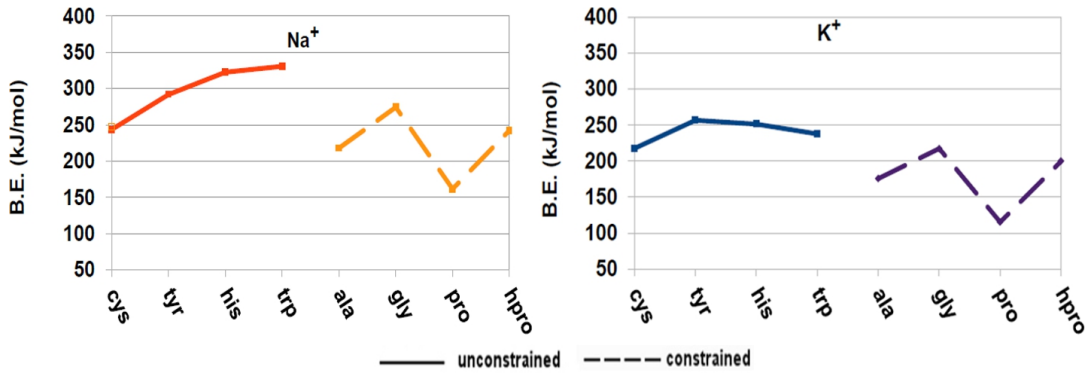

Figure 3.6: Binding energies for Na^+ (LHS) and K^+ (RHS) with selected triple-chain models. Amino acids found primarily in the terminal region (Cys, Tyr, His, Trp) are modeled with the unconstrained triple-chain, while those found primarily in the central region (Ala, Gly, Pro, Hyp) are modeled with the constrained triple-chain. The lines are merely guides to the eye. Data on Pro, Hyp and Gly from previous work.[17]

surface-aggregated fibrils in the presence of Na^+ and thicker ones in the presence of K^+. It seems that Na^+ could promote bridging of terminal regions as there is strong preference for Na^+ in terminal regions. This helps us to understand our experimental data of elongation of fibrils in the presence of Na^+. For K^+ there is lateral and longitudinal growth of fibrils because, for K^+ ions, there is no clear preference.

Finally, a better understanding of ion binding to a range of amino acid sites within a complex protein context is of use to answer a wider range of questions, including ion-induced collagen degradation and other important biomedical matters.

Bibliography

[1] Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009,** *78,* 929-958.

[2] Gelse, K.; Poschl, E.; Aigner, T. *Adv. Drug Delivery Rev.* **2003,** *55,* 1531–1546.

[3] Tenboll, A.; Darvish, B.; Hou, W.; Duwez, A.-S.; Dixon, S. J.; Goldberg, H. A.; Grohe, B.; Mittler, S. *Langmuir* **2010,** *26,* 12165-12172.

[4] Sigel, H. *Metal Ions in Biological Systems;* volume 3 Marcel Dekker,Inc.: NY, USA, 1974.

[5] Gross, J.; Kirk, D. *J. Biol. Chem.* **1958,** *233,* 355-360.

[6] Leow, W. W.; Hwang, W. *Langmuir* **2011,** *27,* 10907-10913.

[7] Kumar, M. R.; Merschrod S., E. F.; Poduska, K. M. *Biomacromol.* **2009,** *10,* 1970-1975.

[8] Sun, M.; Stetco, A.; Merschrod S., E. F. *Langmuir* **2008,** *24,* 5418-5421.

[9] Loo, R. W.; Goh, M. C. *Langmuir* **2008,** *24,* 13276-13278.

[10] Kawska, A.; Hochrein, O.; Brickmann, J.; Kniep, R.; Zahn, D. *Angew. Chem. Int. Ed.* **2008,** *47,* 4982-4985.

[11] Schepers, T.; Brickmann, J.; Hochrein, O.; Zahn, D. *Z. Anorg. Allg. Chem.* **2007,** *633,* 411-414.

[12] Zahn, D.; Hochrein, O.; Kawska, A.; Brickmann, J.; Kniep, R. *J. Mater. Sci.* **2007,** *42,* 8966-8973.

[13] Hodge, A. J.; Schmitt, F. O. *Proc. Natl. Acad. Sci. 46,* 186-197.

[14] Hofmann, H.; Fietzek, P. P.; Kühn, K. *J. Mol. Bio.* **1980,** *141,* 293-314.

[15] Rich, A.; Crick, F. H. C. *J. Molec. Biol.* **1961,** *3,* 483-506.

[16] Vitagliano, L.; Nèmethy, G.; Zagari, A.; Scheraga, H. A. *J. Mol. Biol.* **1995,** .

[17] Fatima, S.; Merschrod, E. *J. Comput. Chem.* **2013,** *na,* in review.

[18] Remko, M.; Soralová, S. *J. Biol. Inorg. Chem.* **2012,** *17,* 621-630.

[19] Shankar, R.; Kolandaivel, P.; Senthilkumar, L. *J Phys Org Chem* **2011,** *24,* 553-567.

[20] Jover, J.; Bosque, R.; Sales, J. *Dalton Trans.* **2008,** 6441-6451.

[21] Rezabal, E.; Mercero, J.; Russo, N.; Ugalde, J. *J. Chem. Theory Comput.* **2007,** *3,* 1830-1836.

[22] Rimola, A.; Rodríguez-Santiago, L.; Sodupe, M. *J Phys Chem B* **2006,** *110,* 24189-24199.

[23] Marino, T.; Russo, N.; Toscano, M. *J Phys Chem B* **2003,** *107,* 2588-2594.

[24] Gokel, G. W. *Chem. Commun.* **2003,** 2847-2852.

[25] Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994,** *266,* 75-81.

[26] Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J. Comput. Chem.* **1989,** *10,* 982-1012.

[27] Avogadro: an open-source molecular builder and visualization tool. Version 1.0.0 http://avogadro.openmolecules.net/.

[28] Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. *J. Cheminf.* **2012,** *4,* 17.

[29] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Mod.* **1998,** *16,* 133-138.

[30] PyMOL: The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC http://www.pymol.org/.

[31] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993,** *14,* 1347-1363.

[32] Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

[33] Becke, A. D. *J. Chem. Phys.* **1993,** *98,* 5648–5652.

[34] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988,** *37,* 785.

[35] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007,** *111,* 10439-10452.

[36] Burt, M. B.; Decker, S. G. A.; Atkins, C. G.; Rowsell, M.; Peremans, A.; Fridgen, T. D. *J Phys Chem B* **2011,** *115,* 11506–11518.

[37] Heaton, A. L.; Moision, R. M.; Armentrout, P. B. *J. Phys. Chem. A* **2008,** *112,* 3319-3327.

Chapter 4

Conclusions and future directions

4.1 An improved model for collagen

In this part of my thesis I have studied Na^+ and K^+ ion affinities for various amino acids in type-I collagen. While keeping my focus on collagen-ion association, the experimental and theoretical aspects of metal ion-amino acid interactions in the biomineralization process of collagen have been discussed. In Chapter 2, while realizing the need for a smaller model to understand such interactions I have developed a set of efficient models (constrained and unconstrained geometries for amino acids, peptides and triple chain). The constrained model mimics the restricted geometry of the helical part and the unconstrained model mimics the non-helical part of the protein. Therefore these models provide a better understanding of the amino acid environment in the whole protein. I have found that constrained geometries in case of the triple chain model are good for studying such type of interactions in big proteins like collagen, because inter-chain interactions play a significant role in such interactions by providing cross-chain binding sites.

4.2 A new view of the role of ions in collagen aggregation

Our models provide a more realistic approach to calculate collagen-ion binding energies (BEs). The constrained triple chain model is good for the central part of the protein since it can address the accessibility and compactness issues in the tightly held helical part. The unconstrained model is useful for modeling the terminal, non-helical part. In Chapter 3, with the help of the unconstrained model, I examine ion binding in the terminal region of the collagen monomer. I have found that these metal ions play role in collagen aggregation. Thus, our models provide the scope to consider the amino acid environment in the whole protein (middle helical and terminal loop portions).

I have found that Na^+ binds stronger than K^+ to the studied amino acids within a larger protein context, which has also been shown by previous studies.[1–3] However, I have found that the binding modes within the protein are different: ions bind primarily with backbone carbonyls, not side chains(e.g. glycine). I have also found that implicit solvent models do not have a significant effect on binding energies of the constrained model of triple chain.[4]

I have studied distinct roles for ions in aggregation, based on experimental data[5] indicating that different fibril growth modes are observed with Na^+ compared to K^+. Does Na^+ promote elongation by blocking lateral aggregation sites or by bridging the terminal regions of monomers? Does K^+ lead to wider fibrils through lateral growth or suppression of elongation? Based on binding affinities, it appears that Na^+ promotes bridging in terminal regions, leading to elongated, narrower fibrils.

4.3 Suggestions for future work

In the future, one can look into binding of Ca^{2+} and Pb^{2+} ions to type-I collagen. Ca^{2+} ions help in fibrillogensis of collagen,[6–8] and lead is also interesting because it is also bivalent and similar in size to Ca^{2+}. One important aspect of lead poisoning is lead storage in bones, which contain a significant amount of collagen.[9, 10] Therefore, it would be interesting to see how Pb^{2+} might be interacting with the collagen and affecting its fibril integrity.

One can also build on the initial hypothesis of ion-induced aggregation by studying potential bridging scenarios. The small models which I have developed are even more crucial once one is considering a calculation involving more than one helix, with each helix having some degrees of freedom relative to each other and to the ion.

One could imagine taking this a step further, to create aggregates of short oligomers, but the calculations quickly become unreasonably time-consuming with quantum-mechanical methods. Molecular dynamics can help to deal with multiple oligomers or with larger sections of or even the whole protein,[7, 11] but it needs parametrization.[12–14] Reliable binding energies are a prerequisite for the parametrization, which can come from our results. The binding energies we have calculated with the smaller protein like model by quantum chemical approach may be helpful to parametrize the data for non-quantum approach.

In summary, we can understand very complex biological assemblies with a careful choice of model. We have developed a set of efficient models which are more realistic. The choice of models can make a dramatic difference: the constrained model is a good model for compactness, but by considering the unconstrained model also, we get a better picture of the whole molecule. Our models allow consideration of the amino acid environment of the whole protein (middle vs terminal). For the central region, the backbone plays a important role in binding these ions, while in terminal regions,

the side chains can contribute to the binding of these ions.

Bibliography

[1] Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 1037-1046.

[2] Kish, M. M.; Ohanessian, G.; Wesdemiotis, C. *Intl. J. Mass Spect.* **2003,** *227,* 509-524.

[3] Hoyau, S.; Ohanessian, G. *Chem. Eur. J.* **1998,** *4,* 1561–1569.

[4] Fatima, S.; Merschrod S., E. F. *J. Comput. Chem.* **2013,** *N/A,* in revision.

[5] Sun, M.; Stetco, A.; Merschrod S., E. F. *Langmuir* **2008,** *24,* 5418-5421.

[6] Kawska, A.; Hochrein, O.; Brickmann, J.; Kniep, R.; Zahn, D. *Angew. Chem. Int. Ed.* **2008,** *47,* 4982-4985.

[7] Schepers, T.; Brickmann, J.; Hochrein, O.; Zahn, D. *Z. Anorg. Allg. Chem.* **2007,** *633,* 411-414.

[8] Zahn, D.; Hochrein, O.; Kawska, A.; Brickmann, J.; Kniep, R. *J. Mater. Sci.* **2007,** *42,* 8966-8973.

[9] Meirer, F.; Pemmer, B.; Pepponi, G.; Zoeger, N.; Wobrauschek, P.; Sprio, S.; Tampieri, A.; Goettlicher, J.; Steininger, R.; Mangold, S.; Roschger, P.; Berzlanovich, A.; Hofstaetter, J. G.; Streli, C. *J. Synchrotron Rad.* **2011,** *18,* 238-244.

[10] Silbergeld, E. K. *Environ. Health Perspect.* **1991 February,** *91,* 63-70.

[11] Vitagliano, L.; Nèmethy, G.; Zagari, A.; Scheraga, H. A. *J. Mol. Biol.* **1995,** .

[12] Project, E.; Nachliel, E.; Gutman, M. *J Comp Chem* **2009,** *31,* 1864-1872.

[13] Roux, B.; Karplus, M. *J Comp Chem* **2004,** *16,* 690-704.

[14] Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S. *J Comp Chem* **2005,** *26,* 1612-1616.

Part II

Optical Materials

Chapter 5

Introduction: Nonlinear optical (NLO) properties of π-extended tetrathiafulvalene derivatives

In this section of the thesis, I have focused on computational predictions of nonlinear optical (NLO) properties of π-extended tetrathiafulvalene (π ex-TTF) derivatives. These molecules are used in the fabrication of cation sensors, intra molecular charge transfer and nonlinear optical materials and electro-optical devices.[1–5]

5.1 Polarizability

In nonlinear optics, we study the interaction between light and matter. When a material interacts with electromagnetic waves it becomes polarized, which means a separation of charges occurs in a molecule. This tendency of charge separation is called polarizability. When this tendency is large then molecules are easily polarizable.[6]

The nonlinear optical properties of a material arise when an intense electric field

is applied. Polarization, P, induced in a molecule by an external electric field, E, is expanded in powers of electric field as shown in Equation 5.1.[7]

$$P = \alpha E + \beta E^2 + \gamma E^3 + ... \tag{5.1}$$

The first term, α, is responsible for linear optical properties, and the second, β, and third, γ, and further terms in the expansion are responsible for nonlinear optical properties of a medium. β and γ are also referred to as the first- and second-order molecular hyperpolarizabilities.

Equation 5.1 can be written in vector form,

$$P = \alpha_{ij} E_j + \beta_{ijk} E_{jk} + \gamma_{ijkl} E_{jkl} + ... \tag{5.2}$$

where α_{ij} is the polarizability, β_{ijk} is the first hyperpolarizability (second order polarizability), γ_{ijkl} is the second hyperpolarizability (third order polarizability), and i, j, k, l are molecular coordinates.

The first hyperpolarizability, β, must be zero for centrosymmetric systems. The only way to ensure that $|P(E)| = |P(-E)|$ is if all even order terms are equal to zero, that is to say, even order coefficients (β, etc) must be zero. For example, in benzene which is centrosymmetric, β is zero, while for non-centrosymmetric substituted benzene β is non-zero.[7, 8] This may be more clear from the following explanation.

For centrosymmetric media, regardless of the direction of the electric field, we should observe polarization of the same magnitude, i.e. $|P(E)| = |P(-E)|$. If electric fields $+E$ or $-E$ are applied, then the first term in the Equation 5.1 becomes $+\alpha E$ or $-\alpha E$ respectively. The second term in the Equation 5.1 becomes $+\beta E^2$ whether we apply $+E$ or $-E$. Similarly, higher odd order terms are $\pm\chi_n E^n$ whereas higher even order terms will always be $+\chi_n E^n$.

There is a problem if we have $+\chi_n E^n$ for even order terms, because then $|P(E)|$ is no longer equal to $|P(-E)|$. For example, for $+E$ electric field, $P(E)$ is equal to $+\alpha E$ $+\beta E^2$ $+\gamma E^3$ and for $-E$ electric field, $P(-E)$ is equal to $-\alpha E$ $+\beta E^2$ $-\gamma E^3$. So to make them equal we need an extra factor of $-2\beta E^2$, i.e. $|P(E)| = |P(-E) - 2\beta E^2|$, which is not true experimentally. Therefore β must be zero for centrosymmetric systems to ensure that $|P(E)| = |P(-E)|$.

5.1.1 Size effects

Polarization is related to the number of electrons and in turn to molecular mass. Thus the size and the polarizability are co-related.[9] In the last two decades there has been ongoing research to determine the size and the polarizability relationship.[10–14] In Chapter 7, we also look at the relationship of system size and nonlinear optical properties.

Ghanty and his co-workers reported in 1993 that there is a correlation between polarizability and system size (atom, molecule, cluster). Earlier papers show a similar correlation between polarizability and atomic size, where the number of electrons indicates the size.[13, 14] Huang and co-workers studied macrocyclic thiophene derivatives. By adding thiophene units (consisting of trithiophene and diacetylene) and hence increasing size, they were able to increase the first hyperpolarizability of the thiophene derivatives.[10]

Lan and his co-workers have reported that polarizability and hyperpolarizability of silicon carbide are related to size and shape of the Si_nC_n cluster. This paper gives a good example of how the size and number of electrons are related but different. According to their paper, polarizability is not directly related to the geometrical size but depends on the composition of the SiC cluster. The polarizability is higher when more Si atoms are present as compared to C atoms. The carbon atom with atomic

number 6 is smaller than the Si with atomic number 14, so it is less polarizable. The polarizability also depends on the shape of the cluster not only on the number of the electrons. The prolate shape is more polarizable than a compact shape cluster.[11, 15] Lan and his co-workers have also found that prolate shaped C-rich clusters have higher β and smaller γ.[11, 16]

May et al. have reported that γ increases with extension of conjugation and donor-acceptor substitution of small cyanoethynylethene molecules.[17, 18] The largest γ value reported by May et al. was 3800 (10^{-36} esu) for a cyanoethynylethene molecule which was donor substituted and most conjugated.[19] Alain and co-workers in 1999 found that γ of push-pull phenylpolyenes of intermediate size increases more with conjugation length than β.[20]

5.1.2 Substituent effects

There are several strategies to enhance NLO properties of π-conjugated systems. NLO properties can be affected by variations within π-conjugated systems, through the addition of donor and acceptor (D-A) groups.[17, 21] The strength of the donor and acceptor groups plays an important role in determining the NLO properties. For example amine based donor groups enhance NLO responses (β and γ) in chromophores having thiophene derivatives.[22]

5.2 Applications of NLO materials

Materials with large NLO properties are useful in telecommunications, optical computing and information processing.[6, 23, 24] The π-extended tetrathiafulvalene (π ex-TTF) derivatives as organic opto-electronics materials are useful for these applications.[24] π-conjugated organic systems with large β are used in actual devices for

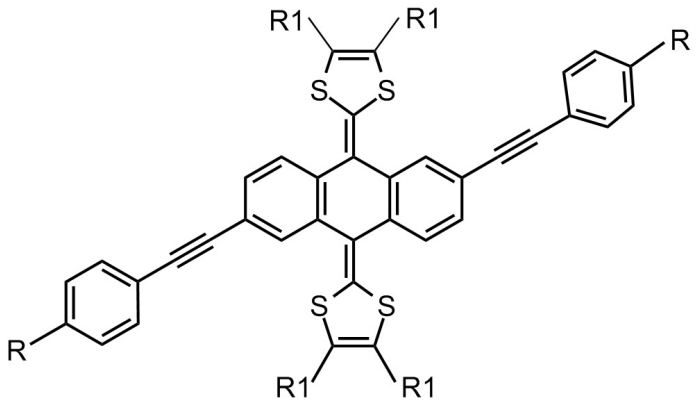

Figure 5.1: TTFAQ-IV with different R1 (R1 = H, SH, SMe) and R (R = H, NO_2, CN, ^tBuS, NMe_2, OCOMe, OH, CH_3, SH, OCH_3, SCH_3, NH_2)

practical applications and there are many references available for β.[25] For γ, less literature is available.[26] A large second hyperpolarizability, γ is needed for photonic applications, for example in optical switching and data processing.[25] γ values are usually too low for most materials to be used in devices for practical applications.[25, 27]

In general, there is a lack of data available for NLO properties of π-extended tetrathiafulvalene, which also provides the motivation for this project.

5.3 π-extended tetrathiafulvalene

In this part of the thesis I have focused on π-extended tetrathiafulvalene derivatives which have conjugated π-spacers, referred to as ex-TTF, with various functional groups, as depicted in Figure 5.1. When these ex-TTFs are expanded via an anthraquinone (AQ)-type π-spacer, they are referred to as TTFAQs.[4, 5, 28] These compounds are interesting because of their redox, electron delocalization and structural properties, and their dication form is thermodynamically stable. Because of all these qualities, TTFAQ has been used in making charge-transfer complexes, NLO materials, cation sensors and in many other applications.[29–32]

In the last three decades, material chemists have extended TTFAQ in different ways. Some have linked different acceptor groups to one side of anthracene through single and/or double bonds to form D-A systems which are used to form charge-transfer complexes. These complexes exhibit properties such as nonlinear optical effects.[3, 31, 33] Highly extended TTFAQ has been prepared by Diaz and her co-workers by placing a bi-anthracene spacer between dithiole rings.[31] Zhao and co-workers have synthesized various TTFAQ derivatives, studied their optical and electronic properties.[4, 5]

I have undertaken a computational study of NLO properties of various ex-TTF's and further extended TTFAQ and its derivatives in this part of the thesis.

5.4 The role of computational chemistry

With the help of computational chemistry, one can predict and investigate many chemical properties of systems in a cost effective way with no worries of chemicals and laboratory waste.[34] In designing and developing nonlinear optical materials, labor, facilities and time are needed, and progress is sometimes hampered by uncertain results. Computational approaches to design and predict properties of materials have been shown to be very effective in saving time and energy.[34–36] Computational methods can be used to reduce costly inputs and harmful by-products by designing materials with the most desirable properties.

Therefore computational studies can help in designing systems with desirable hyperpolarizabilities.[35] In the last few decades, quantum mechanical techniques have proved very useful in improving hyperpolarizabilities via rational design of the NLO systems.[37] Furthermore, computational work can extend our knowledge of system properties beyond what has been currently synthesized, such as our work here which

examines additional substituents beyond those which were created by Zhao group.[4, 5]

5.4.1 The Hartree-Fock method

I have calculated the nonlinear optical properties of TTFAQ and its derivatives with an *ab initio* quantum mechanical approach. By definition *ab initio* methods do not require parameterization, and they give reliable results.[35] In order to exactly solve the Schrödinger equation,

$$\hat{H}\psi = E\psi \tag{5.3}$$

in which $\hat{H}$ represents the Hamiltonian operator, E is the total energy, and ψ is the wavefunction, for a multi-particle system, we apply several approximations.

The Hamiltonian $\hat{H}$ consists of kinetic and potential energy terms for the nuclei and electrons. The Born-Oppenheimer approximation simplifies $\hat{H}$ by separating the nuclear motion from electronic motion. Since nuclei are heavier than the electrons, they can be considered stationary. Therefore according to the Born-Oppenheimer approximation the nuclear kinetic energy term is zero and the nuclear-nuclear Coulombic energy term is constant.[38]

The Hartree-Fock (HF) approximation simplifies the electronic terms. It treats each electron as interacting with a mean field created by all other particles. The resulting equation

$$\hat{F}\phi_i = \epsilon_i\phi_i, \qquad i = 1, 2, 3, ...N \tag{5.4}$$

involves the one-electron Fock operator $\hat{F}$, single-determinant wavefunctions ϕ_i, and the energies ϵ_i.[38–40]

In this part of the work I have applied HF theory with the 6-31G basis set to optimise geometries, to find imaginary frequencies, and to calculate NLO properties.

5.4.2 Finite field calculations

I have used the FFIELD[8] algorithm as implemented in GAMESS[41, 42] to calculate hyperpolarizabilities. This is a numerical, finite field approach wherein the energy of the system is calculated in the presence and absence of applied fields with components in the x, y, z, and combined directions. By writing the energy $\mathcal{E}$ of the system in an electric field as a series expansion in field E, analogous to Equation 5.2, one obtains

$$\mathcal{E}(E) = \mathcal{E}(0) - \mu_i E_i - \frac{1}{2!}\alpha_{ij}E_iE_j - \frac{1}{3!}\beta_{ijk}E_iE_jE_k - \frac{1}{4!}\gamma_{ijkl}E_iE_jE_kE_l - ... \qquad (5.5)$$

Knowing $\mathcal{E}$ for various fields E, one can solve for the different elements of the polarizability and hyperpolarizability tensors (α_{ij}, β_{ijk}, γ_{ijkl}).[8] The average hyperpolarizabilities β and γ, as well as $\beta\mu$, are calculated using formulas as given in the paper by Kurtz[8] and also defined by Equations 6.2 and 6.3 in Chapter 6.

5.5 Part II: optical materials

I have used computational methods to model a range of substituted ex-TTFAQ molecules and their NLO properties, which can provide targets to experimentalists. It is not possible to synthesize every possible potentially useful derivatives of targeted compounds to examine the trends in nonlinear optical properties of all these ex-TTFAQs. Some of these derivatives have been synthesized by the Zhao group and some of them I have designed by adding different donor and or acceptor groups on both sides of the anthracene spacer and different groups on dithiole rings, shown in Figure 5.1.

Our goal is to develop a consistent relationship between structure and NLO properties (both β and γ) for ex-TTFAQ that could be used to design ex-TTFAQ with

large β and γ values.

Bibliography

[1] Bryce, M. R. *J. Mater. Chem* **2000,** *10,* 589-598.

[2] Bendikov, M.; Wudl, F.; Perepichka, D. F. *Chem. Rev.* **2004,** *104,* 4891-4946.

[3] Herranz, M.; Sánchez, L.; Martín, N. *Phosphorus, Sulfur, and Silicon and the Related Elements* **2005,** *180,* 1133-1148.

[4] Shao, M.; Zhao, Y. *Tetrahedron Lett.* **2010,** *51,* 2892–2895.

[5] Chen, G.; Mahmud, I.; Dawe, L. N.; Daniels, L. M.; Zhao, Y. *J. Org. Chem.* **2011,** *76,* 2701-2715.

[6] Marder, S. R.; Kippelen, B.; Jen, A. K.; Peyghambarian, N. *Nature* **1997,** *388,* 845-851.

[7] Williams, D. J. *Angew. Chem. Int. Ed.* **1984,** *23,* 690-703.

[8] Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. *J. Comput. Chem.* **1990,** *11,* 82-87.

[9] Ayers, P. W. *Faraday discussions* **2007,** *135,* 161-190.

[10] Huang, S.; Ren, A.-M.; Zou, L.-Y.; Zhao, Y.; Guo, J.-F.; Feng, J.-K. *J Mol Model* **2012,** *18,* 393-404.

[11] Lan, Y.-Z.; Kang, H.-L.; Niu, T. *Int. J. Quantum Chem.* **2013,** *113,* 949–958.

[12] Chandrakumar, K. R. S.; Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem. A* **2004,** *108,* 6661-6666.

[13] Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem.* **1996,** *100,* 17429-17433.

[14] Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem.* **1993,** *97,* 4951-4953.

[15] Karamanis, P.; Pouchan, C.; Weatherford, C. A.; Gutsev, G. L. *J. Phys. Chem. C* **2011,** *115,* 97-107.

[16] Lan, Y.-Z.; Feng, Y.-L. *Chem. Phys. Lett.* **2011,** *501,* 409-413.

[17] May, J. C.; Biaggio, I.; Bures, F.; Diederich, F. *Appl. Phys. Lett.* **2007,** *90,* 251106.

[18] May, J.; Lim, J.; Biaggio, I.; Moonen, N.; Michinobu, T.; Diederich, F. *Opt. Lett.* **2005,** *30,* 3057-3059.

[19] May, J.; Lim, J.; Biaggio, I.; Moonen, N.; Michinobu, T.; Diederich, F. *Optics Letters* **2005,** *30,* 3057-3059.

[20] Alain, V. *Adv. Mater.* **1999,** *11,* 1210-1214.

[21] Krawczyk, P. *J. Mol. Model.* **2010,** *16,* 659-668.

[22] Kwon, O.; Barlow, S.; Odom, S. A.; Beverina, L.; Thompson, N. J.; Zojer, E.; Brédas, J.-L.; Marder, S. R. *J. Phys. Chem. A* **2005,** *109,* 9346-9352.

[23] Dalton, L. R.; Sullivan, P. A.; Bale, D. H. *Chem. Rev.* **2010,** *110,* 25-55.

[24] Segura, J. L.; Martin, N. *Angew. Chem. Int. Ed.* **2001,** *40,* 1372-1409.

[25] Marder, S. R.; Torruellas, W. E.; BlanchardDesce, M.; Ricci, V.; Stegeman, G. I.; Gilmour, S.; Bredas, J. L.; Li, J.; Bublitz, G. U.; Boxer, S. G. *Science* **1997,** *276,* 1233-1236.

[26] Tykwinski, R.; Gubler, U.; Martin, R.; Diederich, F.; Bosshard, C.; Gunter, P. *J. Phys. Chem. B* **1998,** *102,* 4451-4465.

[27] Meyers, F.; Marder, S. R.; Pierce, B. M.; Brèdas, J. L. *Chem. Phys. Lett.* **1994,** *228,* 171-176.

[28] Shao, M.; Chen, G.; Zhao, Y. *Synlett* **2008,** *3,* 371–376.

[29] Godbert, N.; Bryce, M. R. *J. Mat. Chem.* **2002,** *12,* 27-36.

[30] Sarhan, A. E. A. O. *Tetrahedron* **2005,** *61,* 3889-3932.

[31] Diaz, M. C.; Martin, N.; Viruela, R.; Viruela, P. M.; Ortí, E.; Brede, O.; Zilbermann, I.; Guldi, D. M. *Chem. Eur. J.* **2004,** *10,* 2067-2077.

[32] Christensen, C. A.; Bryce, M. R. *J. Org. Chem.* **2007,** *72,* 1301-1308.

[33] Diaz, M.; Seoane, C.; Martin, N. *J. Org. Chem.* **2004,** *69,* 4492-4499.

[34] Cramer, C. J. *Essentials of computational chemistry: theories and models;* John Wiley & Sons ltd.: West Sussex, UK, 2nd ed.; 2004.

[35] Kanis, D.; Ratner, M.; Marks, T. *Chem. Rev.* **1994,** *94,* 195–242.

[36] Lewars, E. G. *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics;* Springer: NY, USA, 2nd ed.; 2011.

[37] Dalton, L. R. *J. Phys. Condens. Matter* **2003,** *15,* R897-R934.

[38] Simons, J. *J. Phys. Chem.* **1991,** *95,* 1017-1029.

[39] Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102,* 6648-6653.

[40] Szabo, A.; Ostlund, N. S. *Modern quantum chemistry : introduction to advanced electronic structure theory;* Dover: Mineola, NY, 1996.

[41] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993,** *14,* 1347-1363.

[42] Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

Chapter 6

Optical and electronic properties of anthraquinone-type π-extended tetrathiafulvalene (TTFAQ) derivatives

The optical and redox properties of anthraquinone-type π-extended tetrathiafulvalene (TTFAQ) derivatives are computationally investigated here. We find that the nature of the end capping groups has an effect on the nonlinear response of TTFAQ derivatives and that these compounds appear to be viable building blocks for NLO applications. The TTFAQ core acts as an acceptor or a donor depending on the strength of the capping groups. The acetylenic bridge is more than just an electron conduit: when a weak acceptor group like ^{t}BuS is present, the sulfur atom interacts with the acetylenic bridge. The change in conjugation in the acetylenic bridge then leads to electron transfer from the acetylenic bridge to the ^{t}BuS group, resulting in anomalous properties.

6.1 Introduction

TTF (tetrathiafulvalene) and its derivatives are redox active, strong electron donors and have wide ranges of applicability in organic conductors and superconductors, magnets, molecular switches and cation sensors, molecular machines, organic optoelectronics, and plastic solar cells.[1–4] The variety of potential applications arises in part from the tunability of its reversible two electron oxidation by introducing a π-conjugated spacer between the two dithiole rings or by peripheral substitutions.[5] Recently, many materials have been prepared using TTF (1,3 dithiole) as an electron donor linked to some acceptor groups via σ and/or π linkage owing to their electronic interactions.[6]

The π-extended tetrathiafulvalene derivatives which have conjugated π-spacers, referred to as ex-TTF, have been synthesized with various functional groups in the last 20 years since Bryce started investigating their properties in the 1980s.[2, 7] Zhao and his group have synthesized and measured electrochemical properties of various ex-TTFs including the ex-TTFs expanded via an anthraquinone (AQ)-type π-spacer, generally referred to as TTFAQs.[1, 2, 8] TTFAQ and its derivatives exhibit NLO properties and are used as redox-active building blocks for molecular-based electronic and optoelectronic devices and materials.[2, 3, 9–12] Also, because of the higher thermal stability of TTFAQ compounds, these compounds are good for practical applications.[7]

TTFAQ is known to form a dication by a reversible two electron oxidation process.[3] The combination of redox, charge delocalization and the structural properties of TTFAQs have made these compounds attractive candidates for building blocks of electronic and photonic devices, nonlinear optical materials, cation sensors, etc.[13] For example, TTFAQ-donor-acceptor triads can be used in electrochemical sensors for transition metal ions.[2]

Little information about NLO properties of these kinds of TTFAQs is available

but we know they are synthetically achievable and stable. More generally, π extended TTF are less studied organic metals.[14, 15] Therefore, they represent an attractive system for further study.Computational studies can provide useful information about new materials whose development is still in infancy. Computational methods are also useful in understanding and interpreting the trends of experimental results. The role of computational strategies in predicting properties of new materials has been emphasized by various researchers involved in development of novel organic metals.[7, 16, 17] Recently, Santos et. al. have synthesized the first fully conjugated tetrathiafulvalene-tetracyano-π-quinodimethane (TTF-TCNQ) type system.[6] Their theoretical calculations show electronic communication between substituents in ground-state wavefunctions, but geometrical changes upon excitation and under redox result in a loss of this coupling as seen in spectroscopic and electrochemical experiments.

In this chapter, we explore the electronic structure of TTFAQ and its derivatives with an *ab initio* approach. We analyze our findings in relation to previous measurements.[2] However, our purpose here is not to reproduce experimental data but to understand prior experimental results (particularly trends) and to develop predictive power for the properties of this important class of ex-TTFs.

6.2 Computational Details

The Ghemical force field (an all-atom force field similar to Tripos-5.2, which has been tested on organic molecules with a range of functionalities[18]), as implemented in Avogadro,[19, 20] was used to generate the starting geometries. MacMolPlt was used to build and visualize the input and output files.[21] Quantum chemical calculations were performed with the JAN 2009 (R3) version of Gamess (US), provided by ACEnet.[22, 23] The geometries of TTFAQ derivatives are optimized at the RHF/6-31G level. Cal-

culations with the same basis set were used to check for imaginary frequencies in optimized geometries. The second order polarizability (β) and third order polarizability (γ) have been calculated by applying a finite field.

We chose this level of theory and basis set based on literature findings. Champagne *et al.* have reported that density functional theory (DFT) did not give good results for NLO properties of large π-conjugated systems.[24] They have assessed density functional methods to investigate linear and non-linear responses to electric fields of a range of π-conjugated systems and similary found problems with DFT results.[25–28] Recently, Galán and co-workers have shown that Hartree Fock (HF) calculations produced NLO properties of thiazole derivatives which are consistent with experimental results.[29] Jacquemin *et al.* found that diffuse and polarization functions are not very important for extended systems and the 6-31G basis gave satisfactory results which are very close to values obtained by large basis sets (for example aug-cc-pVTZ).[30] Therefore, we believe that the HF/6-31G approach is suitable for our compounds.

6.3 Geometry: evidence for aromaticity in the dication

The optimized structures are shown in Figure 6.1. In TTFAQ two dithiole rings are separated by anthracene. The neutral molecule of TTFAQ has a non-planar butterfly or saddle-like conformation because this type of conformation reduces the steric hindrance between the sulfur atoms of 1,3-dithiole moieties and the peri hydrogens of neighboring benzene rings.[7, 31, 32] The saddle-like structure of neutral TTFAQ planarizes when two-electron oxidation occurs. The dication conformation gains aromaticity when the central part (anthracene) becomes planar. π electron delocalization is clear in the molecular orbitals for the dication, and the decreased bond alternation

in the rings (see Table 6.1) provides further evidence for increased aromaticity.

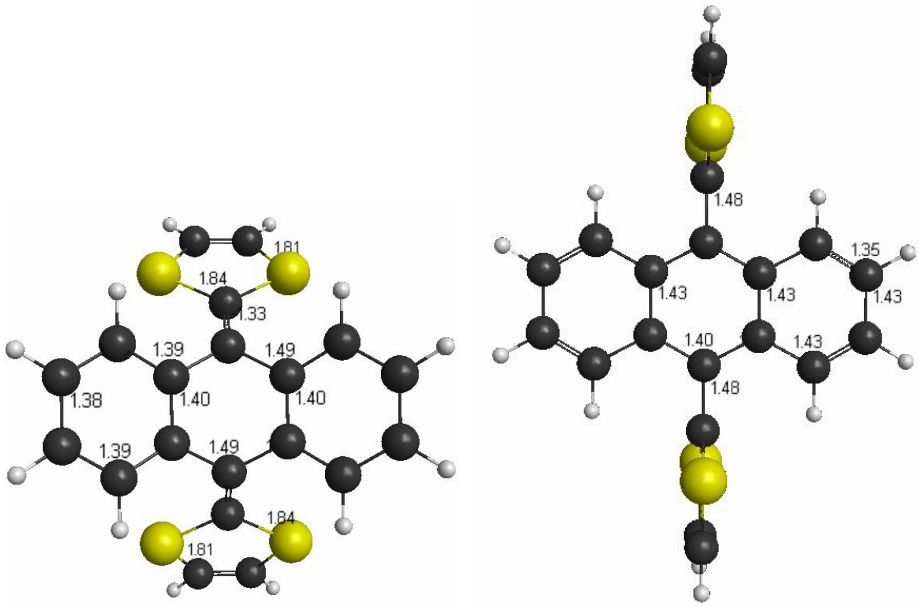

Figure 6.1: Optimized geometries for (a) neutral and (b) dicationic TTFAQ (HF/6-31G). Bond distances are in Å.

The dithiole rings in the dication are perpendicular to the anthracene, removing the steric hindrance in the dication and thus supporting the planarity of the anthracene.[32] They also gain in aromaticity as seen by equalization of bond lengths (Table 6.1), and the orthogonality allows for greater electronic communication through coupling of the σ system in the dithioles with the π system in the anthracene and vice versa.

6.4 Linkage with donor and acceptor groups

TTFAQ derivatives can be functionalized at the dithiole ring or at the anthraquinone spacer via conjugated π-linkages, such as C=C and C≡C bonds.[1, 2, 8] Here we study π-extended TTFAQ with acetylene linkages to different donor and/or acceptor groups, as depicted in Figure 6.2. Our systems can be described as A-B-TTFAQ-B-A, D-B-

	Bond length (Å)	
	neutral	dication
dithiolium ring		
C-C	1.31 (d)	1.33 (d)
C-S	1.81	1.79
S-C	1.84	1.72
Anthracene spacer		
C-C 1	1.40 (c)	1.43 (c)
C-C 2	1.49 (s)	1.40 (c)
C-C 2a	1.39 (c)	1.35 (d)
C-C(bridging) 3	1.33 (d)	1.48 (s)

Note: d-double, s-single, c-conjugated bond

Table 6.1: Bond lengths in neutral and dicationic TTFAQ showing more equal bond lengths (gain in aromaticity) in the dication, obtained with HF/6-31G.

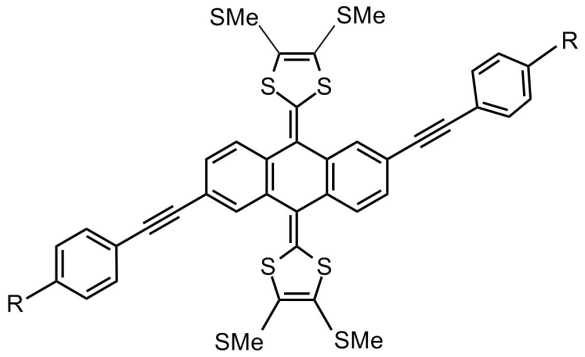

Figure 6.2: TTFAQ with different substituents (R = H, $N(CH_3)_2$, ^{t}BuS, CN, $COOCH_3$, NO_2, SCH_3, CH_3, OH, OCH_3, NH_2, SH.)

TTFAQ-B-D, and D-B-TTFAQ-B-A, where B is the acetylene bridge, A is an acceptor, and D is a donor. Substituents do not significantly change the geometry of TTFAQ derivatives, which means that the molecule is still butterfly-like in the neutral form and planar in the dication. We investigated the following substituents: donors H, SCH_3, CH_3, OH, OCH_3, NH_2, $N(CH_3)_2$; and acceptors ^{t}BuS, SH, CN, $COOCH_3$, NO_2. Some of these have been successfully synthesized already[2] and some have not (SCH_3, CH_3, OH, OCH_3, NH_2, SH,).

Isosurfaces of frontier orbitals show good electronic communication with R groups

via acetylene linkages, as seen in Figure 6.3 (R=H). There is also large spatial overlap between frontier molecular orbitals, meaning that both the HOMO and LUMO extend over the same region of the molecule. The highest occupied molecular orbital (HOMO) extends from TTFAQ moiety to donor or acceptor R groups along acetylenic bridges on both sides of the TTFAQ unit. The lowest unoccupied molecular orbital (LUMO) also extends from the R groups to the anthracene part of TTFAQ in the symmetrically substituted derivatives.

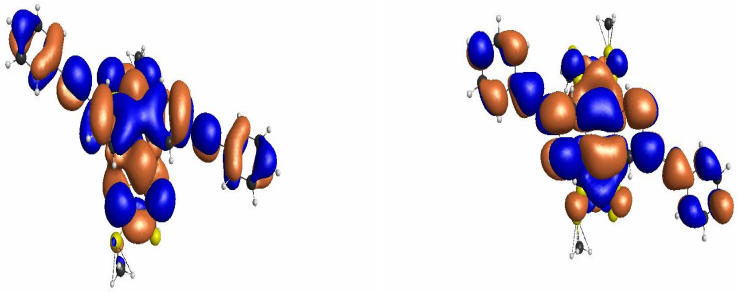

Figure 6.3: Molecular orbital isosurfaces for neutral TTFAQ with R = H (*i.e.* unsubstituted) (HF/6-31G). The HOMO is at left and the LUMO at right; different shading of the lobes indicates different phases. Note that there is a good electronic communication with R groups via the acetylene linkages.

There is an interesting difference in the molecular electronic structure when adding donor or acceptor substituents, which turns out to impact the optical properties of these molecules. The core TTFAQ acts as a donor or an acceptor depending on the type of substituent. When an acceptor group is present the core acts as a donor, and when a strong donor group like $N(CH_3)_2$ is present it behaves as an acceptor. Evidence for the ability of the core to act as a donor and as an acceptor is seen in the optical properties described below as well as in shifts in the distribution of electron density between the core and the substituents in the occupied and unoccupied

orbitals. The HOMO is based more on the donor and the LUMO is based more on the acceptor group, confirming our assignments of donor and acceptor qualities of the substituents.

6.5 Tuning electronic and optical properties

As some of these substituted molecules have been synthesized and characterized,[2, 3] we can use experimental data to validate or calibrate computational predictions on the remaining substituents and for the whole class of molecules.

6.5.1 Tuning redox properties

These TTFAQ derivatives show interesting redox properties which change with the substituents. It is challenging to predict redox properties, particularly for multi-electron oxidation (these undergo two-electron oxidation), using gas-phase ground-state calculations. Solvent effects can be significant for charged systems, and we have already shown above that there are significant conformational changes between the neutral and dicationic species. However, it may still be possible to identify a trend based on experimental data which can be used to predict properties for as-yet-unsynthesized variants.

Using a very simple model based on the ideas of Koopman,[33] the energy of the HOMO can be used to approximate the ionization energy which in turn can be related to oxidation (loss of an electron). Therefore, one might expect the HOMO energies to track the half wave potential $E_{1/2}$ for the first oxidation. This is not the case, as seen in Table 6.2, which is not entirely surprising for several reasons. Koopman's theorem comes from a one-electron theory, although of course one-electron theories formed the basis for much of our intuitive chemical understanding. Perhaps more

significant, however, are the changes in conformation and the potential contributions of solvation which affect experimental measurements but which are not accounted for in the calculation of the HOMO energy of the neutral species.

R	E_{pa} (mV)	$E_{1/2}$(mV)	HOMO (hartree)
–$N(CH_3)_2$	560	–	-0.251
–H	583	507	-0.268
–S^tBu	585	437	-0.270
–NO_2	608	491	-0.283
–$COOCH_3$	610	540	-0.273
–CN	619	498	-0.279

Table 6.2: Electrochemical data (half wave potential $E_{1/2}$ and peak anodic potential E_{pa}) for the first oxidation of several substituted neutral TTFAQ,[2, 3] with the calculated energies of the HOMO for comparison (HF/6-31G).

That said, there is a clear trend in Table 6.2 showing a correlation between HOMO energy and peak anodic potential (E_{pa}) for the first oxidation step. The HOMO energies for the rest of the substituted TTFAQs for which experimental data is not available indicate a lower E_{pa} for the first oxidation step for R = NH_2, OCH_3, OH, and a higher E_{pa} for R = SH. This data is collected in Table B.1.

6.5.2 Tuning the energy gap

UV visible spectra provide a simple and direct experimental handle on the optical band gap (energy gap or HOMO-LUMO gap for molecules) for a material. The key features from the experimental UV visible spectra[2, 3] are that the lowest energy peak (λ_{max}), given in Table 6.3, is red shifted for a stronger acceptor group as compared to unsubstituted TTFAQ (R=H). The red shift is largest for NO_2 which is the strongest electron withdrawing group, followed by CN, $COOCH_3$ and ^{t}BuS. Conversely, the absorption band is blue shifted for a strong donor group ($N(CH_3)_2$).

This lowest energy peak is likely to be a HOMO-LUMO transition, hence we

R	λ_{max} (nm)	Abs. edge (cm^{-1})	E_{HOMO} (au)	E_{LUMO} (au)	ΔE_{H-L} (eV)
$-N(CH_3)_2$	453	22075	-0.251	0.070	8.735
–H	458	21830	-0.268	0.061	8.953
$-COOCH_3$	460	21740	-0.273	0.048	8.735
$-^{t}BuS$	461	21690	-0.270	0.057	8.898
–CN	470	21280	-0.279	0.041	8.708
$-NO_2$	470	21280	-0.283	0.018	8.191

Table 6.3: Experimental[3] optical properties of TTFAQ (neutral) with various R groups, and related computational parameters (HF/6-31G).

checked for matching shifts in the HOMO-LUMO gap (ΔE_{H-L}), the absorption edge, and the peak wavelength (λ_{max}). ΔE_{H-L} generally trends with the absorption edge and λ_{max}, but there are issues with the strong electron withdrawing groups. For strong electron withdrawing groups, there is a more significant shift in electron density even with this symmetrical substitution since the core is acting as a donor and the molecule is not planar. These strong changes are observed in the LUMO energies but not the HOMO (which is naturally based more on the donor group).

E_{HOMO} tracks λ_{max} quite well, however. For R=$N(CH_3)_2$, E_{HOMO} is higher than other TTFAQ derivatives considered here, as see in Table 6.3. Conversely, E_{HOMO} for R=NO_2 is lower as compared to others.These computed values turn out to correlate with other optical properties as well, as shown below.

6.5.2.1 Oxidative spectra

Oxidative spectra (UV-visible spectra during oxidation of the material) provide additional insight into the transformation from the neutral molecule to the dication. The key features from the experimentally measured oxidative UV visible spectra[2, 3] are that, upon oxidation of the TTFAQ derivatives, the lowest energy peak at about 21000 cm^{-1} discussed above, disappears, with a new peak forming at 23000 cm^{-1}.The HOMO-LUMO gap shrinks in going from the neutral molecule to the dication, so we

would not expect the HOMO-LUMO peak to shift to higher energy. Therefore, this new peak does not result from the HOMO-LUMO transition.

In fact, the former HOMO-LUMO peak shifts to much lower energy and has a lower intensity because of the symmetry mismatch. The HOMO of the dication is of π character. It is based on the anthracene and extends along the conjugation bridge to the end capped substituents as shown in Figure 6.4. There is a higher energy empty orbital (the LUMO+2) which is also of π character and based on anthracene as seen in Figure 6.5b. In contrast, the LUMO and LUMO+1 have primarily σ character based on the anthracene and some π character based on the dithiole rings as showed in Figure 6.5 a. Therefore, the band which emerges upon oxidation is likely a transition from HOMO to LUMO+2 or LUMO+3 on the dication, a $\pi-\pi$ transition. The values for $\Delta E_{H-(L+2)}$ and $\Delta E_{(H-1)-(L+2)}$ of the dications are given in Table 6.4.

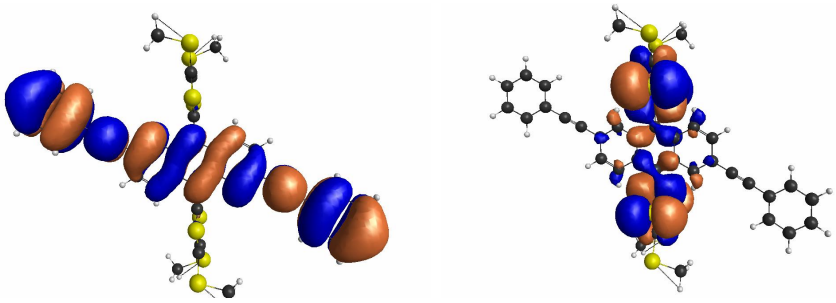

Figure 6.4: Molecular orbital isosurfaces for the HOMO (left) and LUMO (right) of TTFAQ dication with R = H, (HF/6-31G). Different shading of the lobes indicates different phases.

	λ_{max} (nm)	$\Delta E_{H-(L+2)}$ (eV)	$\Delta E_{(H-1)-(L+2)}$
–H	310	6.857	7.320
–^{t}BuS	326	6.776	7.075
–NO_2	333	7.293	7.919
–CN	328	7.048	7.483

Table 6.4: A comparison of λ_{max} and energy transitions from HOMO-n to LUMO+m on for cationic substituted TTFAQ, calculated at HF/6-31G.

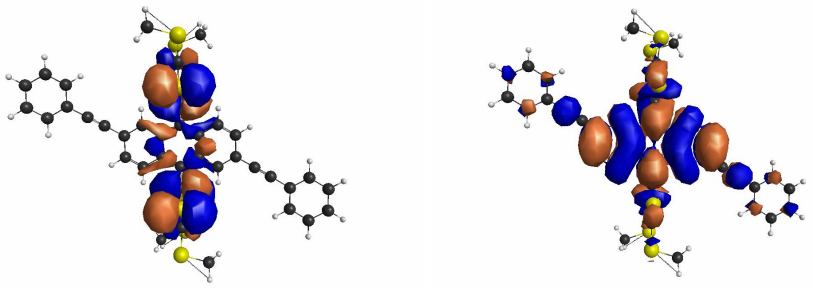

Figure 6.5: Isosurfaces for frontier orbitals a) LUMO+1 and b) LUMO+2 of TTFAQ dication with R = H, (HF/6-31G). Different shading of the lobes indicates different phase.

6.5.3 Tuning polarizability: NLO properties

TTFAQ and its derivatives have captured the attention of modern materials chemists because of their remarkable electronic, optical and electrochemical properties.[12] Two approaches to tuning the polarizability would be to extend the π system, or to try different substituents (including doubly-substituted molecules exhibiting a push-pull effect). The net goal is enhanced hyperpolarizabilities (non-linear response), increasing β and γ which are defined as follows:[17, 34, 35]

$$P = \alpha E + \beta E^2 + \gamma E^3 + ... \tag{6.1}$$

where P is the polarization, E is the electric field, α is the first order polarizability, β is the second order polarizability (first hyperpolarizability), and γ is the third order polarizability (second hyperpolarizability). Here we also calculate $|\beta\mu|$, the first hyperpolarizability projected along the dipole moment direction, since that is a better comparison to experiment where the measurement is along the dipole moment.[34] $|\beta\mu|$ and γ are defined in Equations 6.2 and 6.3 respectively.[34]

$$|\beta\mu| = \frac{\frac{3}{5}(\beta \cdot \mu)}{|\mu|} \tag{6.2}$$

$$\gamma = \frac{1}{5}[\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz} + 2(\gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz})] \tag{6.3}$$

6.5.3.1 Symmetric substitution: D-D / A-A

As mentioned above, the core of the molecule can act as an acceptor or as a donor. When coupled with strong donor or acceptor substituents, and keeping in mind the butterfly geometry, we observe a push-pull effect even with symmetrical substitution (both donor or both acceptor substituents). We calculate enhanced β and $|\beta\mu|$ values as a result, whether the substituents are *either* a strong electron withdrawing or donating group. The TTFAQ derivative with the nitro (NO_2) end caps have the highest first order hyperpolarizability among all the symmetrically substituted compounds (A-B-TTFAQ-B-A or D-B-TTFAQ-B-D) considered in this study. γ is the highest for the $N(CH_3)_2$ (strongest donor group) substituted TTFAQ. The second highest γ value is for the NO_2 (strongest acceptor group) substituent.

The trends observed for NLO response directly relate to the electronic excitation data discussed above, where the strongest red- or blue-shifts in optical absorbance result in the highest $|\beta\mu|$ values. In general, a smaller energy gap makes a molecule more easily polarizable,[36] which explains why TTFAQ with R=NO_2 has a higher second hyperpolarizability than others. A similar trend is not evident for γ, which does not track λ_{max} nor the ΔE_{H-L}. The relation between ΔE_{H-L} and $|\beta\mu|$ therefore goes beyond overall polarizability, with a smaller ΔE_{H-L} reflecting more push-pull or better donor-acceptor match, which results in a higher $|\beta\mu|$. This is shown in Table 6.5.

The second hyperpolarizability, γ, depends on the donor or acceptor ability of the TTFAQ and the end capping groups. Larger values of γ arise from substituents with more diffuse electron clouds. Also, the smallest values of γ are seen for R=H.

We found that donor strength has a predominant role in enhancing γ values for the TTFAQ derivatives studied.

R	ΔE_{HL} (eV)	β ($\times 10^{-30}$ esu)	$\lvert\beta\mu\rvert$ ($\times 10^{-48}$ esu)	γ ($\times 10^{-36}$ esu)
–H	8.953	0.513	0.774	136.975
–^{t}BuS	8.898	0.525	0.561	217.235
–$N(CH_3)_2$	8.735	17.778	25.111	251.508
–$COOCH_3$	8.735	12.91	9.4	222.735
–CN	8.708	11.451	17.492	245.487
–NO_2	8.191	23.929	36.558	245.608

Table 6.5: Calculated ΔE_{HL} and hyperpolarizabilities β, $|\beta\mu|$, and γ (HF/6-31G).

The TTFAQ with R = ^{t}BuS is an interesting case, with lower polarizability than expected. The sulfur modifies the acetylenic bridge through electron withdrawal. By looking at the charges on the acetylenic bridge and on the TTFAQ unit, we found that there is less negative charge on the acetylenic bridge in the case of ^{t}BuS substituent as compared to others. For other substituents, the acetylenic bridge acts as a passive conduit for electrons between the R group and the central TTFAQ core, while for R=^{t}BuS, it is acting as a donor because it is coupling with the ^{t}BuS group. The ^{t}BuS substituent changes the whole back bone of the compound. The C-S overlap populations between the acetylenic bridge and ^{t}BuS group are more positive than for the linking atoms of other substituents. Therefore we can say that there is interaction between ^{t}BuS group and acetylenic bridge which leads to the anomalous behaviour of ^{t}BuS containing compounds.

6.5.3.2 Asymmetric (D-A) substitution

We have considered one case of asymmetrically substituted (D-B-TTFAQ-B-A), with NH_2 (strong donor) and NO_2 (strong acceptor) groups. There is significant enhancement of first order hyperpolarizability because of intra-molecular charge transfer from NH_2 to NO_2. The donor and acceptor nature of the substituents is shown in Figure

6.6. There we see the HOMO is based on the NH_2 side extended along acetylenic bridge and the LUMO is on the NO_2 side. The second order polarizability (γ) is not enhanced by this charge transfer and localization of orbitals. γ is highest for $N(CH_3)_2$ (strongest donor group) symmetrically substituted TTFAQ. The second highest value is for NO_2 symmetrically substituted TTFAQ. These data are summarized in Table 6.6.

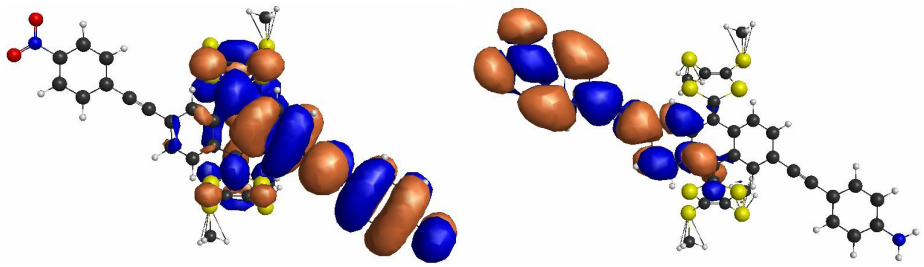

Figure 6.6: Molecular orbital isosurfaces for the HOMO (left) and LUMO (right) of asymmetrically substituted (-NH_2, -NO_2) TTFAQ (HF/6-31G). Different shading of the lobes indicates different phases.

molecule	β ($\times 10^{-30}$ esu)	$\lvert\beta\mu\rvert$ ($\times 10^{-48}$ esu)	γ ($\times 10^{-36}$ esu)
NH_2–TTFAQ–NH_2	14.664	22.247	200.679
NO_2–TTFAQ–NO_2	23.929	36.558	245.608
NH_2–TTFAQ–NO_2	61.026	91.954	213.637

Table 6.6: Comparison of NLO response between symmetrically and asymmetrically substituted TTFAQ (HF/6-31G).

6.6 Conclusions

The electronic and optical response of symmetrically and asymmetrically substituted π ext-TTFAQ analogues have been studied. Our computational approach has enabled the interpretation of trends in prior experimental results and points to approaches for tuning redox and optical properties of this useful class of materials. Our calculations

indicate good electronic communication across these large species, whether neutral or dicationic.

Varying the donor or acceptor strength of the substituents on symmetrically-substituted molecules can be used to tune the oxidation potential and the optical response of these molecules. Despite strong conformational changes between neutral and dicationic structures, we have identified a strong correlation between the experimentally measured peak anodic potentials for the first oxidation step and the energy of the HOMO for the neutral molecules, providing a simple measure to screen potential molecules for tuned redox activity. The energy of the HOMO for the neutral molecules also tracks λ_{max} in the optical absorption spectra regardless of the nature of the substituent, making it a better screening tool than the HOMO-LUMO gap which responds differently to donor and acceptor substituents.

For NLO response, the range for $|\beta\mu|$ for the symmetrically-substituted TTFAQ derivatives is 0.56 to 36.56 $\times 10^{-48}$ esu, while for the asymmetrically substituted (donor on one end, acceptor on the other) TTFAQ analogue, it reaches 91.95 $\times 10^{-48}$ esu due to the push-pull effect of the substituents. But even with a symmetrically substituted version, a strong donor or a strong acceptor can lead to enhanced first-order hyperpolarizability, with the TTFAQ acting as either an acceptor or a donor. The first-order hyperpolarizability is generally predicted by ΔE_{H-L}. The large molecular second hyperpolarizability of the symmetrically substituted TTFAQ with R=N$(CH_3)_2$ shows that a strong donor group can be used to create a molecule with large γ.

According to our results, the class of TTFAQ derivatives studied can be used as building blocks for a range of functional materials. They are reversibly oxidized to give dications, thermally stable and have large nonlinear optical properties. Our results will be useful for future experiments in designing and screening potential synthetic targets.

6.7 Supporting Information

Additional Supporting Information may be found in Appendix B.

Bibliography

[1] Chen, G.; Mahmud, I.; Dawe, L. N.; Daniels, L. M.; Zhao, Y. *J. Org. Chem.* **2011,** *76,* 2701-2715.

[2] Shao, M.; Zhao, Y. *Tetrahedron Lett.* **2010,** *51,* 2892–2895.

[3] Chen, G.; Shao, M.; Wang, L.; Zhao, Y. *Asian Chem. Lett.* **2007,** *11,* 185–196.

[4] Bryce, M. R. *J. Mater. Chem* **2000,** *10,* 589-598.

[5] Nielsen, M. B. *Phosphorus, Sulfur & Silicon & the Related Elements* **2011,** *186,* 1055–1073.

[6] Santos, J.; Illescas, B. M.; Martín, N.; Adrio, J.; Carretero, J. C.; Viruela, R.; Ortí, E.; Spänig, F.; Guldi, D. M. *Chem Eur J* **2011,** *17,* 2957–2964.

[7] Bendikov, M.; Wudl, F.; Perepichka, D. F. *Chem. Rev.* **2004,** *104,* 4891-4946.

[8] Shao, M.; Chen, G.; Zhao, Y. *Synlett* **2008,** *3,* 371–376.

[9] Batail, P. *Chem. Rev.* **2004,** *104,* 4887–4890.

[10] Albert, I. D. L.; Marks, T. J.; Ratner, M. A. *J. Am. Chem. Soc.* **1998,** *120,* 11174-11181.

[11] Albert, I. D. L.; Marks, T. J.; Ratner, M. A. *J. Am. Chem. Soc.* **1997,** *119,* 6575–6582.

[12] Kanis, D.; Ratner, M.; Marks, T. *Chem. Rev.* **1994,** *94,* 195–242.

[13] Christensen, C. A.; Bryce, M. R. *J. Org. Chem.* **2007,** *72,* 1301-1308.

[14] Diaz, M. C.; Illescas, B.; Martin, N. *Tetrahedron Lett.* **2003,** *44,* 945–948.

[15] Chen, G.; Dawe, L.; Wang, L.; Zhao, Y. *Org. Lett.* **2009,** *11,* 2736-2739.

[16] Avci, D. *Spectrochimica Acta Part A Mol Biomol Spectrosc.* **2011,** *82,* 37–43.

[17] Pranata, J.; Murray, C. *J. Phys. Org. Chem.* **1993,** *6,* 531-534.

[18] Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J Comput Chem* **1989,** *10,* 982-1012.

[19] Avogadro: an open-source molecular builder and visualization tool. Version 1.0.0 http://avogadro.openmolecules.net/.

[20] Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. *J. Cheminf.* **2012,** *4,* 17.

[21] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Mod.* **1999,** *16,* 133–138.

[22] Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

[23] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993,** *14,* 1347-1363.

[24] Champagne, B.; Perpète, E. A.; van Gisbergen, S. J. A.; Baerends, E.-J.; Snijders, J. G.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. *J. Chem. Phys.* **1998,** *109,* 10489-10498.

[25] de Wergifosse, M.; Champagne, B. *J. Chem. Phys.* **2011,** *134,* 074113.

[26] Champagne, B.; Perpète, E. A.; Jacquemin, D.; van Gisbergen, S. J. A.; Baerends, E.-J.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. *J. Phys. Chem. A* **2000,** *104,* 4763.

[27] Champagne, B.; Kirtman, B. *Int. J. Quant. Chem* **2009,** *109,* 3103-3111.

[28] Borini, S.; Limacher, P. A.; Luthi, H. P. *J. Chem. Phys.* **2009,** *131,* 124105.

[29] Galán, E.; Andreu, R.; GarÃŋn, J.; Mosteo, L.; Orduna, J.; Villacampa, B.; Diosdado, B. E. *Tetrahedron* **2012,** *68,* 6427-6437.

[30] Jacquemin, D.; Champagne, B.; Hättig, C. *Chem. Phys. Lett.* **2000,** *319,* 327-334.

[31] Frère, P.; Skabara, P. J. *Chem. Soc. Rev.* **2005,** *34,* 69–98.

[32] Godbert, N.; Bryce, M. R. *J. Mat. Chem.* **2002,** *12,* 27-36.

[33] Szabo, A.; Ostlund, N. S. *Modern quantum chemistry : introduction to advanced electronic structure theory;* Dover: Mineola, NY, 1996.

[34] Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. *J. Comput. Chem.* **1990,** *11,* 82-87.

[35] Williams, D. J. *Angew. Chem. Int. Ed.* **1984,** *23,* 690-703.

[36] Pearson, R. G. *Proc. Natl. Acad. Sci.* **1986,** *83,* 8440-8441.

Chapter 7

Assessing metrics of size and charge separation as predictors for non-linear optical response of organic materials

Anthraquinone-type π-extended tetrathiafulvalene (TTFAQ) derivatives are computationally investigated to identify relationships between size, charge separation, and nonlinear optical (NLO) response of these compounds. We identify simple predictors for strong NLO response, developing structure-NLO property relationships that could be used to synthesize organic opto-electronics materials with large β and γ values. We evaluate different metrics of size and find that a "quantum mechanical" size (as defined by Hollett et al, *J Phys Chem A* **110**, 13884–8, 2006) provides a more nuanced distinction between molecules of similar molar mass and is a good metric for third-order polarizabilities (γ). Size is a poor predictor for second-order polarizability ($|\beta\mu|$), but charge separation provides a fast indication of the magnitude of second-order

response while taking into account potential contributions from seemingly-innocent "backbone" components.

7.1 Introduction

Electro-optical and photonic devices require materials which have large nonlinear properties. In these devices nonlinear optical (NLO) effects are required to manipulate (process and store) information as well as to transmit that information at high speed.[1–3] More generally, the NLO effects are useful in controlling the phase or the frequency of the light.[3] The second and third order NLO effects are required for optical frequency conversion, optical switching and data processing etc.[1, 3, 4] NLO properties can also be used to directly transduce chemical information. Recently, Champagne and his co-workers have proved that NLO responses are not only important for frequency doubling and intensity-dependent refractive indices but can be used in characterization and detection of metal cations.[5]

Nonlinear optical response is a consequence of nonlinear polarization of a system. In the presence of light the charges in the system are displaced which causes polarization in the system. When the intensity of light is low, the induced polarization is linearly proportional to the electric field strength. In case of high intensity light, the polarization is related in a non-linear fashion to field strength.[1, 3] In other words, the nonlinear optical responses arise in a nonlinear medium where the induced polarization depends nonlinearly on the applied field.[6]

Computational approaches to materials design offer several advantages. Calculations are less time-consuming (and cheaper) than the synthesis-purification-measurement cycle, providing predictive power and guiding subsequent synthetic efforts. Calculations can also yield insights into trends and can lead to a more general understanding

of the characteristics required to optimize a particular molecular property. Therefore, a simple and fast method to screen potential NLO materials would be of great advantage.

Theoretical methods have been used to measure and optimize the molecular second-order polarizability of organic chromophores. These calculations are useful in producing qualitative trends[7] but full computations of non-linear polarizability are time consuming. To achieve an effective screening method, we are seeking simpler metrics which can serve as indicators for NLO properties, providing an initial step which can then be followed up with further computational and experimental work.

In the 70's, methods to correlate atomic and molecular dimensions to atomic and molecular polarizabilities were proposed.[8, 9] Later, polarizability (primarily first-order polarizability, α) was correlated with the cube of the equilibrium distance, size of metal clusters, electronic charge density, hardness, electronegativity and ionization potential of atoms and internuclear distances.[10–13] Correlation of number of carbon atoms with polarizability has also been reported by Lan and his co-worker. They have also found that charge redistribution in Si_mC_n clusters (mn = 1–4) is related to polarizability (α).[14]

Higher-order polarizabilities for larger molecules (particularly those with heteroatoms) take even longer to calculate. However, these more complex molecules do not lend themselves easily to the size metrics described above. Galán and co-workers used Mulliken charges to explain second order nonlinear optical properties of the thiazole derivatives.[15] Hatua and co-workers also looked at issues of charge in the polarizability of metal-hydrocarbon complexes.[16]

In this article, we take a systematic approach to assessing several measures of size and charge in predicting NLO response for anthraquinone-type π-extended tetrathiafulvalene (TTFAQ)-derivatives with different end groups.[17–19] This promising class

of molecules provides a rich field of potential NLO materials, where computational guidance for future synthetic efforts would greatly increase the likelihood of success. We calculate $|\beta\mu|$ as well as β and γ, since this first value (polarizability along the molecular dipole) is most closely related to what would actually be measured experimentally.[20]

7.2 Computational details

The initial structures have been generated with molecular mechanics using the Ghemical force field (an all-atom force field similar to Tripos-5.2[21]) as implemented in Avogadro.[22, 23] MacMolPlt was used to build and visualize the input and output files.[24] Quantum chemical calculations were performed with the JAN 2009 (R3) version of Gamess(US).[25, 26] We used Hartree Fock (HF) methods; B. Champagne et al. have reported that density functional theory (DFT) did not give good NLO results for large pi-conjugated systems.[27]

The geometries of TTFAQ derivatives were optimized with the 6-31G basis set, with vibrational analysis to check for imaginary frequencies in optimized geometries. Jacquemin et al. found that diffuse and polarization functions are not very important for extended systems and the 6-31G basis gave satisfactory results which are very close to values obtained with large basis sets (*e.g.* aug-cc-pVTZ).[28] The second order (β) and third order (γ) polarizabilities have been calculated by applying a finite field with the same basis set. All calculations were done in the gas phase.

To calculate "quantum mechanical size" or electronic size, MUNgauss (version 1.7)[29] was used to calculate the electronic second moment average radius of the TTFAQ derivatives.[30] "Nuclear size" or internuclear distance was measured as the distance between nuclear centres of most-distant atoms. Molecular mass and num-

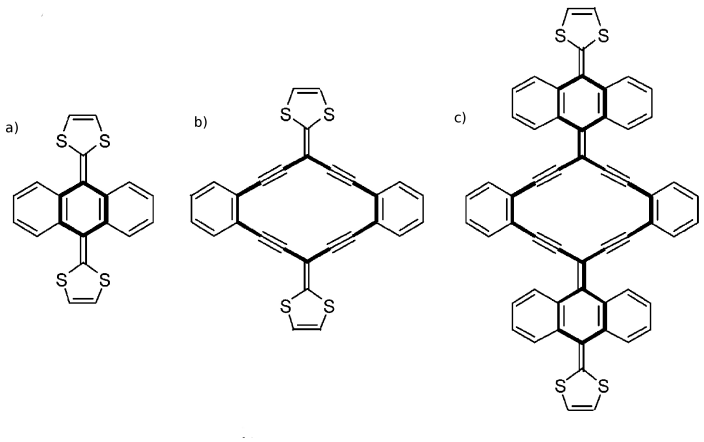

Figure 7.1: Extending the π system of the core: a) TTFAQ-I, b) TTFAQ-II, and c) TTFAQ-III

ber of electrons were also tested as measures of size. Charges for subsections of the molecules are defined as the sum of Mulliken charges for the atoms in those fragments.

7.3 Size and polarizability relation in ex-TTFAQ's

7.3.1 Extending the π system

As a first size-variant, we considered TTFAQ with different extensions of the π system. TTFAQ-II and TTFAQ-III have an expanded core relative to TTFAQ-I, while TTFAQ-IV has also been expanded via acetylenic π bridges. The schematics of these compounds are given in Figures 7.1 and 7.2.

The NLO response upon extending the core π system is given in Table 7.1. While none of these molecules have a particularly large NLO response, we do see a general increase in third order polarizability (γ) with increasing core size. It is important to note that all of these systems do show non-zero second-order polarizabilities (β), despite what could appear to be inversion symmetry when viewed in 2D. These molecules are not planar, as seen in Figure 7.3, thus breaking that inversion symmetry and allowing

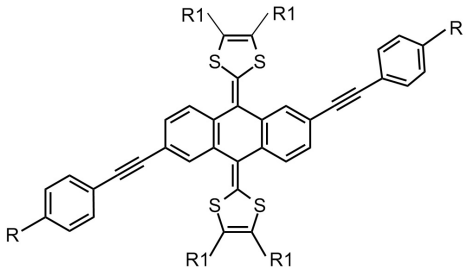

Figure 7.2: TTFAQ-IV, where R = H, OH, CH_3, SH, OCH_3, SCH_3, NO_2, CN, ^{t}BuS, $N(CH_3)_2$, $OCOCH_3$, NH_2, NH_2–NO_2 and R1 = H, SH, SCH_3.

molecule	$\beta(\times 10^{-30} esu)$	$\lvert\beta\mu\rvert(\times 10^{-48} esu)$	$\gamma(\times 10^{-36} esu)$
TTFAQ-I	0.590	0.902	-16.516
TTFAQ-II	0.611	0.882	82.311
TTFAQ-III	3.160	4.797	155.438
TTFAQ-IV	0.545	0.807	130.040

Table 7.1: β and γ values for TTFAQ-I, -II, -III and -IV molecules, calculated at HF/6-31G. (TTFAQ-IV: R and R1 = H).

for even-order non-linear response.

7.3.2 Apparent size effects of substituents

Growing the core did not significantly increase NLO response, so we considered a different way of changing the size of the molecule. In the case of TTFAQ-IV, we wanted to know how NLO properties would be affected if we grew the TTFAQ-IV molecule along the acetylenic bridge and dithiole rings by adding different substituents. Figure 7.2 shows the substituent positions R and R1. Since this also adds heteroatoms, we also needed to consider how the donor or acceptor ability of each substituent influences the NLO properties of the resultant compound.

Adding substituents can change the conformation and structure which would also affect polarizabilities. Substituents on the dithiole rings (R1 = H, SH, SCH_3), for example, showed some differences, as seen in Figure 7.4. When R1 = SCH_3 the R1 group orients perpendicular to the molecular plane, while for R1 = H or SH the

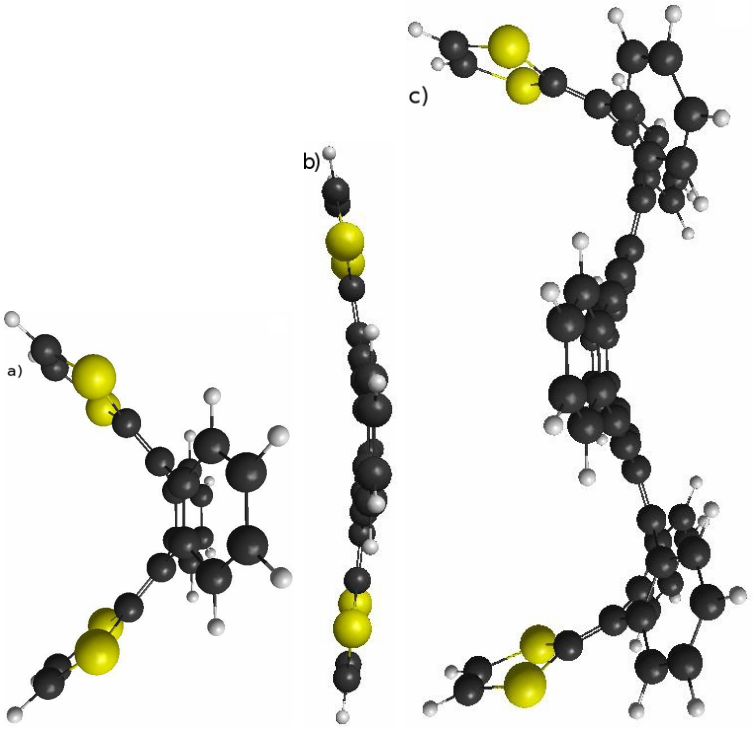

Figure 7.3: Energy-minimum geometries for a) TTFAQ-I, b) TTFAQ-II, and c) TTFAQ-III (HF/6-31G). The black spheres are C atoms, the white spheres are H atoms, and the yellow spheres are S atoms.

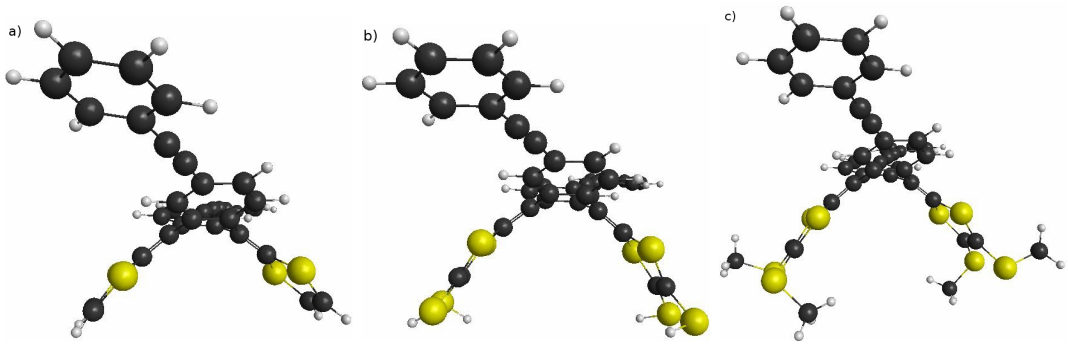

Figure 7.4: TTFAQ-IV with R (= H) and different R1 (a) H; (b) SH ; (c) SMe, optimized at HF/6-31G.

R1 group lies within the plane of the molecule. In most cases, we see increases in $|\beta\mu|$ and γ upon adding thiol substituents to the dithiole group but, because of the conformational changes, there is no direct relationship between NLO response and increased number of electrons or molecular mass, two possible measures of size as shown in Figures 7.5a, b and 7.6a, b. Therefore, a size metric which accounts for these conformational changes would be a definite improvement.

7.3.3 Size based on electronic size

In changing substituents at R, the conformational changes are minimal, which might suggest that a simple measure of size would show good correlation with NLO properties, as has been shown for the simpler systems described in the introduction. However, from Table C.1 we can see that the molecule with ^{t}BuS as a side group (R) does not have the largest NLO values, though this is the largest of the side groups considered, whether considering mass or number of electrons or number of atoms as an indicator of size. We encounter issues with other side groups which are similar or even equal in numbers of electrons (*e.g.* CH_3 and OH) or numbers of atoms (*e.g.* OH and CN), or may be quite similar in inter-nuclear distances (*e.g.* CH_3 and NH_2), but present quite different NLO response.

Electronic size, as defined by Hollet et al,[30] turns out to be a much better option in that it captures and quantifies the chemical intuition which already tells us that –OH and –CN are not the same, and it also serves to rank other groups which are not so obviously different.

In Figure 7.5 showing γ *vs* size, there is a good correlation between γ and electronic size (Figure 7.5c). This is in contrast to Figures 7.5a and 7.5b, comparing γ and other measures of size.

7.4 NLO response and charge separation

We have shown that a size proxy is good for third order polarizability γ, but it is not that good for the second order polarizability β, or more specifically $|\beta\mu|$, as seen in Figure 7.6. We then tried charge separation to see if it correlates with $|\beta\mu|$. For this purpose, we have calculated the sum of Mulliken charges on substituents groups and on the TTFAQ unit. q_1 is the charge on the end cap groups (R), q_2 is the charge on the acetylenic bridge, and q_3 is the charge on the core including the dithiole group substituents (R1). We calculated charge separation in two ways: by dividing q_1 by q_3 and by subtracting q_3 from q_1:

$$q_1/q_3 = q_1/-(q_1+q_2) \tag{7.1}$$

$$q_1 - q_3 = q_1 - (-q_1 - q_2) \tag{7.2}$$

Both of these charge separation measures are good for $|\beta\mu|$, as seen in Figure 7.7. The same cannot be said for γ, where electronic size as seen in Figure 7.5c is much more appropriate than either measure of charge separation seen in Figure 7.8. Quadratic fits are shown in the plots in Figure 7.7 as they capture most simply and

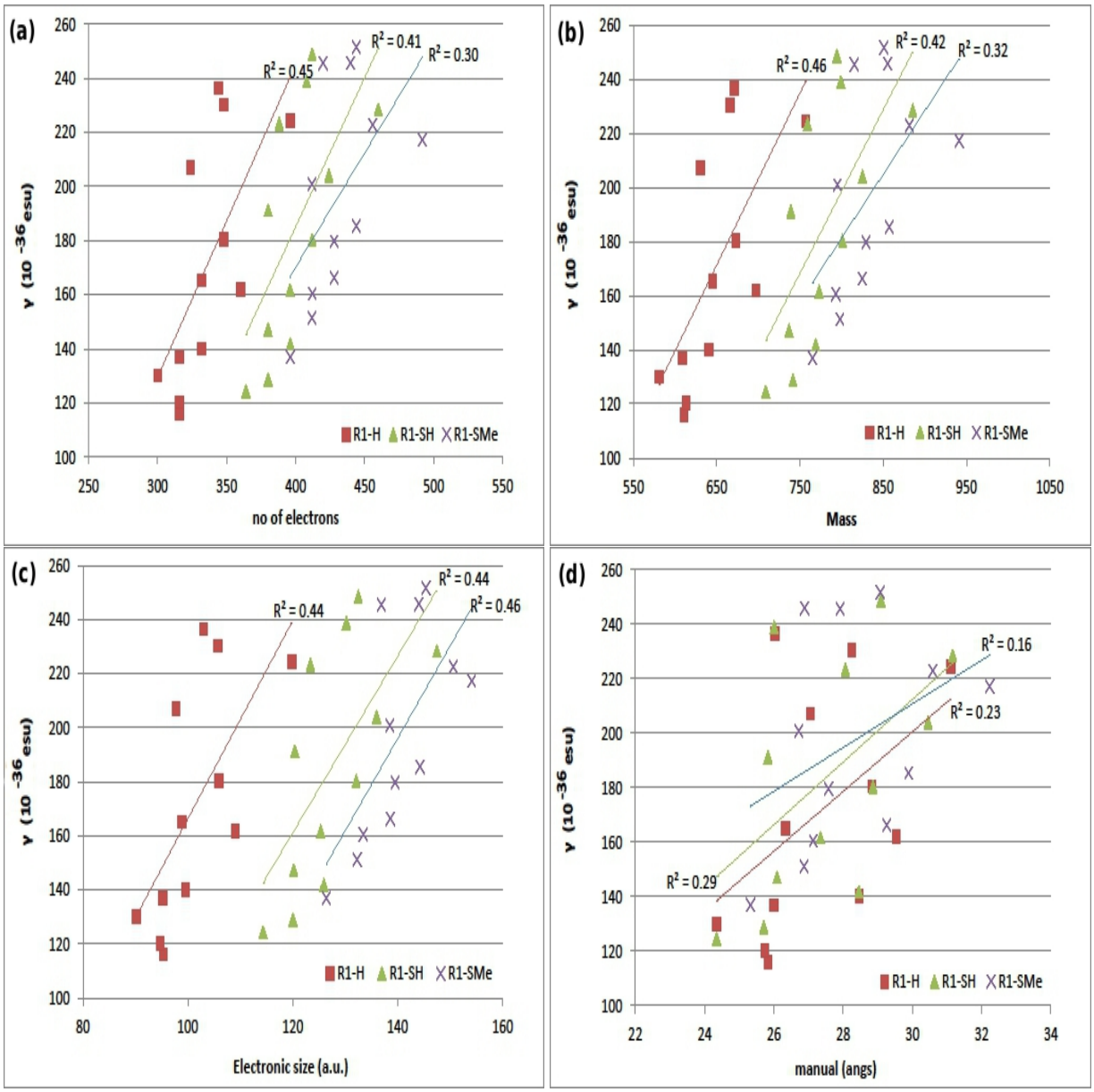

Figure 7.5: γ *vs* (a) number of electrons, (b) mass, (c) electronic size, and (d) manual internuclear size for TTFAQ-IV, where R = H, OH, CH_3, SH, OMe, SMe, CN, OCOMe, ^{t}BuS, NMe_2, NO_2, NH_2 (all values in a.u.).

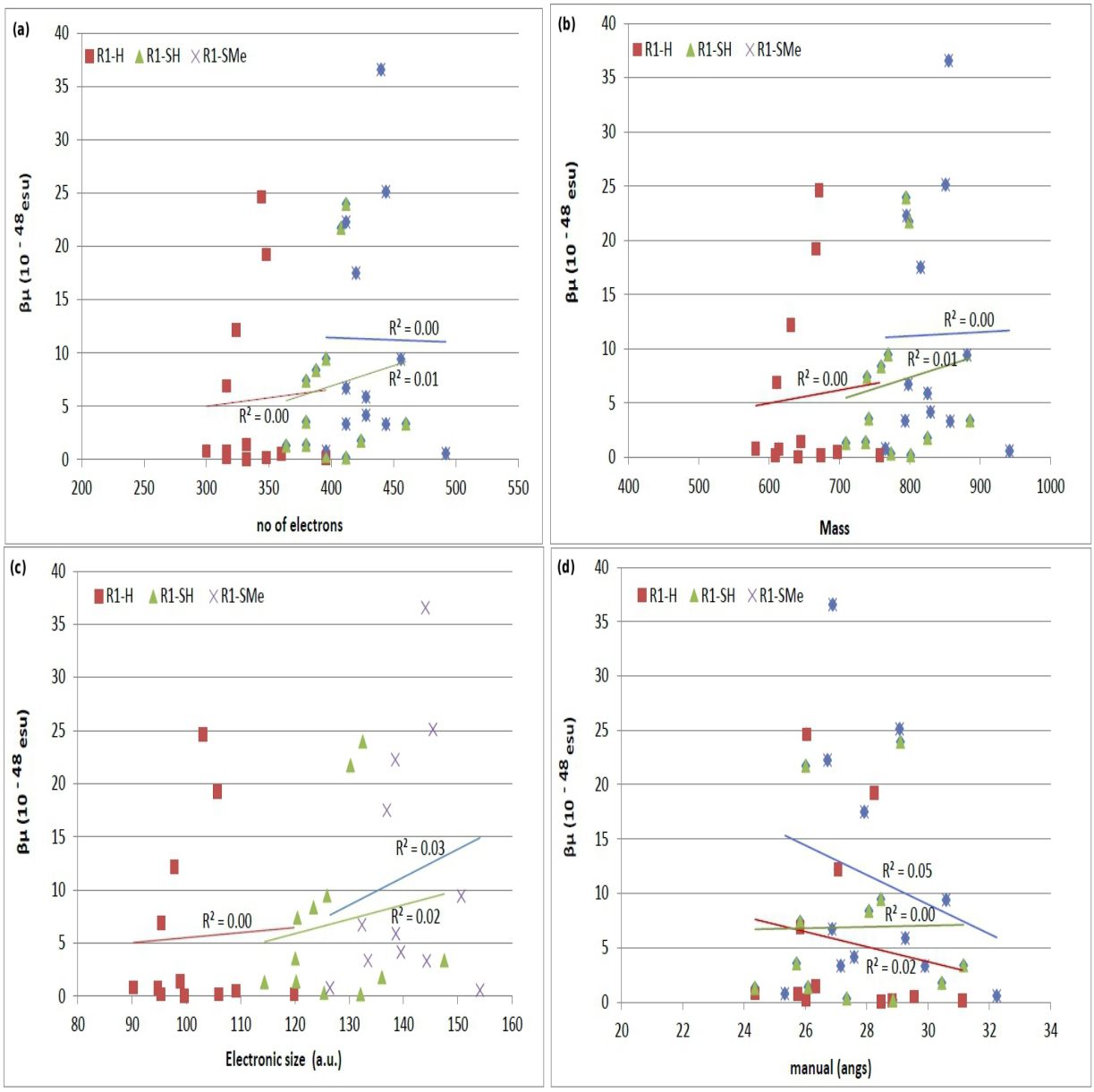

Figure 7.6: $|\beta\mu|$ *vs* (a) number of electrons, (b) mass, (c) electronic size, and (d) manual internuclear size for TTFAQ-IV, where R = H, OH, CH_3, SH, OMe, SMe, CN, OCOMe, ^{t}BuS, NMe_2, NO_2, NH_2 (all values in a.u.).

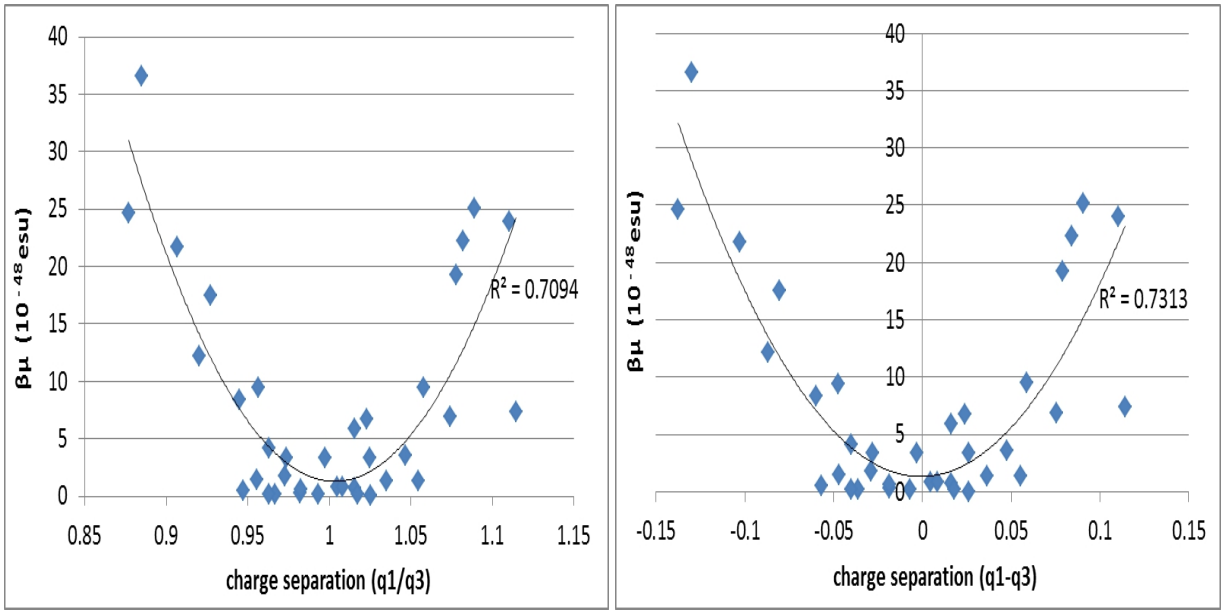

Figure 7.7: (a) $|\beta\mu|$ *vs* charge separation (q_1/q_3) values; (b) $|\beta\mu|$ *vs* charge separation ($q_1 - q_3$) values, graphs for TTFAQ-IV with different R and R1, where R = H, OH, CH_3, SH, OMe, SMe, CN, OCOMe, ^{t}BuS, NMe_2, NO_2, NH_2 (all values in a.u.), The lines are quadratic fits, which capture most simply and directly the observed trends.

directly the empirical relation between the two metrics; a slightly poorer relationship is observed with separate linear fits on either side of 1 (in the case of q_1/q_3) or 0 (in the case of $q_1 - q_3$).

Another interesting feature of these charge separation plots is their indication of whether an end group is acting as a donor or an acceptor *with respect to the TTFAQ core*. For example, the TTFAQ unit behaves as a strong donor in the presence of NO_2 where it has a large positive charge, contributing to the much larger $|\beta\mu|$ values for that R group. In other words, this charge separation measure accounts for the active participation of what might simply be considered a backbone or support group in the electro-optical activity of the molecule.

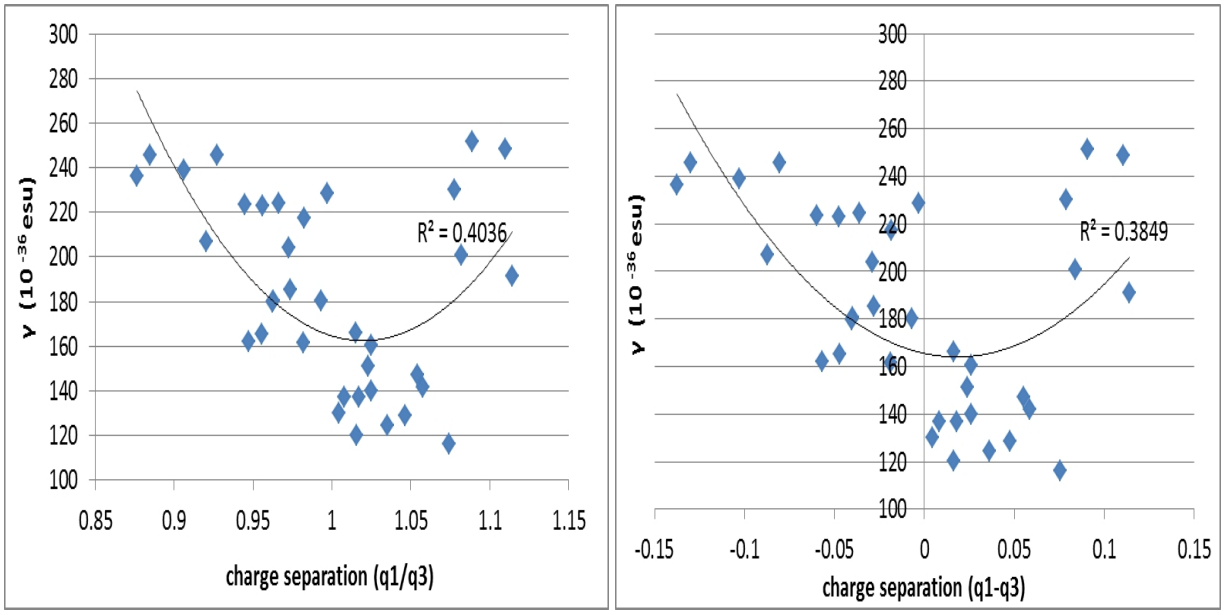

Figure 7.8: γ *vs* charge separation values (a) q_1/q_3 and (b) $q_1 - q_3$ for TTFAQ-IV with different R and R1, where R = H, OH, CH_3, SH, OMe, SMe, CN, OCOMe, ^{t}BuS, NMe_2, NO_2, NH_2 (all values in a.u.). The quadratic fit lines are included simply to allow direct comparison to Figure 7.7.

7.5 Conclusions and Future directions

We have identified simple predictors for strong NLO response, identifying geometric and electronic structure and NLO property relationships that could be used to synthesize organic opto-electronic materials with large β and γ values.

Our calculations indicate the highest $|\beta\mu|$ for R1=SMe and R=NO_2 ($36.56 \times 10^{-48} esu$) and the highest γ for R1=SMe and R=$N(Me)_2$ in all the classes of molecules considered here. More generally, we conclude that the strongest second order NLO response is for the strongest acceptor R group and the strongest third order NLO response is for the strongest donor group.

Electronic size, which incorporates elements of number of electrons, number of atoms/mass, and overall internuclear size, shows good correlation with γ. In other words, the electronic size captures the effective size of molecules when it comes to

electro-optical properties. For example, adding SH as an R group does not change the internuclear distances very much but it definitely changes electronic size. Conversely, growing a molecule by adding atoms does not necessarily mean growing electronically and increasing NLO response.

Charge separation is an easy proxy for $|\beta\mu|$ and indicates the effect of a substituent within the context of the particular molecule, allowing for a finer screening tool for potential new NLO materials. TTFAQ-III may be a good candidate for further expansion along the same lines as TTFAQ-IV.

7.6 Supporting Information

Additional Supporting Information may be found in Appendix C.

Bibliography

[1] Dalton, L. *Adv. Polym. Sci.* **2002,** *158,* 1-86.

[2] Dalton, L. R.; Steier, W. H.; Robinson, B. H.; Zhang, C.; Ren, A.; Garner, S.; Chen, A.; Londergan, T.; Irwin, L.; Carlson, B.; Fifield, L.; Phelan, G.; Kincaid, C.; Amend, J.; Jen, A. *J. Mater. Chem* **1999,** *9,* 1905-1920.

[3] Marder, S. R.; Kippelen, B.; Jen, A. K.; Peyghambarian, N. *Nature* **1997,** *388,* 845-851.

[4] Marder, S. R.; Torruellas, W. E.; BlanchardDesce, M.; Ricci, V.; Stegeman, G. I.; Gilmour, S.; Bredas, J. L.; Li, J.; Bublitz, G. U.; Boxer, S. G. *Science* **1997,** *276,* 1233-1236.

[5] Champagne, B.; Plaquet, A.; Pozzo, J.-L.; Rodriguez, V.; Castet, F. *J. Am. Chem. Soc.* **2012,** *134,* 8101–8103.

[6] Williams, D. J. *Angew. Chem. Int. Ed.* **1984,** *23,* 690-703.

[7] Dalton, L. R. *J. Phys. Condens. Matter* **2003,** *15,* R897-R934.

[8] Charney, E. *Chem. Phys. Lett.* **1980,** *75,* 599-601.

[9] Cohen, A. *Chem. Phys. Lett.* **1979,** *68,* 166-169.

[10] Chandrakumar, K. R. S.; Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem. A* **2004,** *108,* 6661-6666.

[11] Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem.* **1996,** *100,* 17429-17433.

[12] Ghanty, T. K.; Ghosh, S. K. *J. Phys. Chem.* **1993,** *97,* 4951-4953.

[13] Varandas, A. J. C.; Rodrigues, S. *Chem. Phys. Lett.* **1995,** *245,* 66-74.

[14] Lan, Y.-Z.; Kang, H.-L.; Niu, T. *Int. J. Quantum Chem.* **2013,** *113,* 949–958.

[15] Galán, E.; Andreu, R.; GarÃŋn, J.; Mosteo, L.; Orduna, J.; Villacampa, B.; Diosdado, B. E. *Tetrahedron* **2012,** *68,* 6427-6437.

[16] Hatua, K.; Nandi, P. K. *Computational and Theoretical Chemistry* **2012,** *996,* 82-90.

[17] Iyoda, M.; Yamakawa, J.; Rahman, M. J. *Angew. Chem. Int. Ed.* **2011,** *50,* 10522-10553.

[18] Shao, M.; Zhao, Y. *Tetrahedron Lett.* **2010,** *51,* 2892–2895.

[19] Chen, G.; Dawe, L.; Wang, L.; Zhao, Y. *Org. Lett.* **2009,** *11,* 2736-2739.

[20] Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. *J. Comput. Chem.* **1990,** *11,* 82-87.

[21] Clark, M.; Cramer, R. D.; Opdenbosch, N. V. *J Comput Chem* **1989,** *10,* 982-1012.

[22] Avogadro: an open-source molecular builder and visualization tool. Version 1.0.0 http://avogadro.openmolecules.net/.

[23] Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. *J. Cheminf.* **2012,** *4,* 17.

[24] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Mod.* **1999,** *16,* 133–138.

[25] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993,** *14,* 1347-1363.

[26] Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

[27] Champagne, B.; Perpète, E. A.; van Gisbergen, S. J. A.; Baerends, E.-J.; Snijders, J. G.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. *J. Chem. Phys.* **1998,** *109,* 10489-10498.

[28] Jacquemin, D.; Champagne, B.; Hättig, C. *Chem. Phys. Lett.* **2000,** *319,* 327-334.

[29] R.A. Poirier, MUNgauss (Fortran 90 version), Chemistry Department, Memorial University of Newfoundland, St. John's, NL, A1B 3X7. With contributions from

S.D. Bungay and A. El-Sherbiny and T. Gosse and J. Hollett and D. Keefe and A. Kelly and C.C. Pye and D. Reid and K. Saputantri and M.Shaw and M.S. Staveley and Y. Wang and J. Xidos.

[30] Hollett, J. W.; Kelly, A.; Poirier, R. A. *J. Phys. Chem. A* **2006,** *110,* 13884-13888.

Chapter 8

Conclusions and future directions

8.1 Conclusions

How to design an NLO material? To answer this question, one can consider the following points: a) backbone, to see if it can play an active role or act as a passive element in properties; b) shape in 3D, since one can get effective push-pull even for symmetrically substituted molecules; c) size; and d) charge distribution.

For the backbone, we found in Chapter 6 that the acetylenic bridge played a role in the NLO response of TTFAQ-IV with R=^{t}BuS (as defined in Figure 5.1) by coupling with the sulfur of the ^{t}BuS group, while it acts as a passive conduit for electrons between the R group and the central TTFAQ moiety for other R groups. In Chapter 6 we also observed that symmetrical substituted derivatives are not symmetrical in 3D and substituted TTFAQ-IV have non-zero $|\beta\mu|$ as these compounds have effective push-pull between TTFAQ core and the R groups.

After knowing the correlation between NLO response and end cap substituents of TTFAQ-IV derivatives, by taking advantage of computational techniques I designed more TTFAQ derivatives by varying the end cap groups on either sides of the

acetylenic bridges and on the dithiole ring.

Computational methods are useful in understanding and interpreting the trends of experimental results. Computational techniques are also useful in designing new materials and predicting properties. They are fast, modern and green. However, the calculations still require computational time. So we need a simpler method to predict the NLO response. That issue has been dealt with in Chapter 7. Charge separation and size are simple predictors for second-order (β) and third-order (γ) polarizabilities, respectively. Different size measures like number of electrons, number of atoms/mass, and overall nuclear size cannot fully explain the variations in γ, while electronic size does.

8.2 Future directions

π-extended TTFAQ can be used as a stable NLO material, where important stability factors include thermal and photophysical properties. However, there is a need for further studies to enhance their NLO properties. The following provides guidance for future investigations of NLO properties of this class of materials.

8.2.1 New substitutents

Other substituents like CH_3 on the dithiole ring could be added to see how this affects the donor ability of TTFAQ. Perepichka and co-workers have shown that CH_3 on the dithiole rings enhances intra-molecular charge transfer in 1,3-dithiole-nitrofluorene derivatives. CH_3 is more effective than SMe because of its stronger electron donating ability.[1] Therefore, it can be a future direction to look into NLO responses of TTFAQ with CH_3 dithiole rings.

The hetero-substituted R=SH, R1=NH_2/NO_2 TTFAQ has the higher $|\beta\mu|$ values

($93.79 \times 10^{-48} esu$) among the derivatives I considered. Borini and co-workers have used donor acceptor groups of different strength on polyacetylenes to study their effect on the NLO properties; they also found that combination of NH_2 and NO_2 augmented second order NLO properties.[2] In our case, since TTFAQ's donor ability alters with the presence of different donor and acceptor groups,[3] it would be interesting to see how different combinations of donor-acceptor end caps affect NLO properties of these compounds. We have tried the NH_2/NO_2 combination but not other combinations. CN is nearly as strong an electron withdrawing group as NO_2, and NMe_2 is a stronger electron donating group than the NH_2, so one could try CN/NMe_2, and so on, to see how this affects the second order NLO properties of TTFAQ.

8.2.2 Longer conjugated spacer

In future, one could try increasing the length of conjugation spacer for second order NLO enhancement. Increasing conjugation length may enhance NLO properties. Ramzan and co-workers have shown that a polyenic π-spacer (-C≡C-) in a donor-acceptor polyoxometalate cluster enhanced second order NLO properties of that cluster.[4] Herranz and co-workers' results indicate that ex-TTF derivatives have higher first hyperpolarizability with a more conjugated chain.[5] Hence one could test longer conjugating chains separating donor and acceptor moieties.

8.2.3 Radical cations and dications

Another promising direction is the study of γ for radical cations and dications of TTFAQ derivatives. The TTF radical cation has a large negative γ.[6, 7] Since TTFAQ also form radical cation and dication,[8, 9] it would be interesting to study their NLO properties as well.

8.2.4 Solid state and solution

For further studies one can look into the crystalline arrangement of these derivatives as well as solvent effects for these molecules in solution. Gas-phase calculations might be fine if you are looking at single-molecule effects, but many applications will not be using single molecules but rather crystals or thin films or solutions (particularly in chemical sensing).

In some cases, the molecular species may present similar electronic properties as single molecules, in solution, and in solid state. For example, Öhrn and co-worker observed that the electronic excitations of formaldehyde in aqueous solution are similar to those in the gas phase.[10] However, there are other cases where intermolecular interactions in the solid state or in solution will significantly affect molecular properties. Öhrn and co-worker showed in a contrasting article that indole is sensitive to solvent polarity and its optical properties are influenced by the solvent.[11] Perepichka and co-workers have found that the electrochemical properties of the TTFAQ-σ-fluorene system differ in the solid state and in solution.[12]

Frère and co-worker in their review paper showed how molecular structure, electrochemical properties and solid state properties of ex-TTF derivatives are related.[13] They found that the solid state environment affects properties primarily through two mechanisms. Geometrical constraints arising from packing effects can stabilize or destabilize an excited state, or a charge separated state, thus leading to changes in excitation energies or electron transport properties. The second main difference for a condensed phase system relative to a gas-phase or single-molecule one for these types of molecules is the $\pi - \pi$ interactions between molecules as they are stacked in a crystal or film.

Therefore, calculations on systems in the solid state or in solution are worth investigating, to better model the materials as they would be implemented in devices.

Bibliography

[1] Perepichka, D. F.; Ivasenko, O.; Moore, A. J.; Bryce, M. R. *Chem. Eur. J.* **2008,** *14,* 2757-2770.

[2] Borini, S.; Limacher, P. A.; Luthi, H. P. *J. Chem. Phys.* **2009,** *131,* 124105.

[3] Fatima, S.; Xu, M.; Merschrod, E. *TBD* **in prep,** *00,* xx–xx.

[4] Janjua, M. R. S. A.; Khan, M. U.; Bashir, B.; Iqbal, M. A.; Song, Y.; Naqvi, S. A. R.; Khan, Z. A. *Comput. Theor. Chem.* **2012,** *994,* 34 - 40.

[5] Herranz, M.; Sánchez, L.; Martín, N. *Phosphorus, Sulfur, and Silicon and the Related Elements* **2005,** *180,* 1133-1148.

[6] Yamada, S.; Nakano, M.; Kishi, R.; Ohta, S.; Takahashi, H.; Furukawa, S. I.; Nitta, T.; Yamaguchi, K. *Synth. Met.* **2006,** *156,* 375-378.

[7] Nakano, M.; Yamada, S.; Yamaguchi, K. *Chem. Phys. Lett.* **1999,** *311,* 221-230.

[8] Christensen, C. A.; Bryce, M. R. *J. Org. Chem.* **2007,** *72,* 1301-1308.

[9] Jones, A.; Perepichka, D. F.; Batsanov, A. S.; Beeby, A. *Chem. Eur. J.* **2001,** *7,* 973-978.

[10] Öhrn, A.; Karlström, G. **2006,** .

[11] Öhrn, A.; Karlstrom, G. *J. Phys. Chem. A* **2007,** *111,* 10468-10477.

[12] Perepichka, D.; Bryce, M.; Perepichka, I.; Lyubchik, S.; Christensen, C.; Godbert, N.; Batsanov, A.; Levillain, E.; McInnes, E.; Zhao, J. *J. Am. Chem. Soc.* **2002,** *124,* 14227-14238.

[13] Frère, P.; Skabara, P. J. *Chem. Soc. Rev.* **2005,** *34,* 69–98.

Appendix A

Supporting information for "Practical models for ion-collagen binding"

A.1 Gas-phase vs PCM binding energies

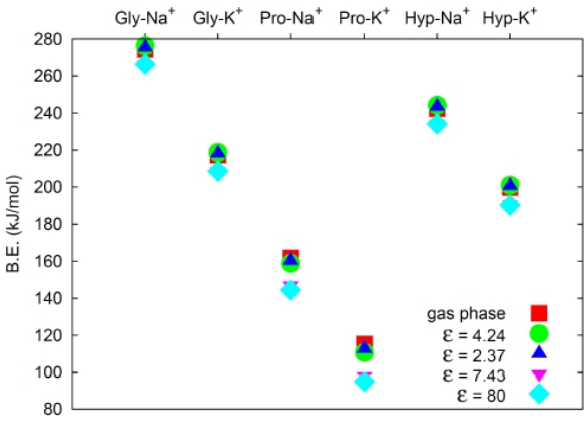

Figure A.1: Binding energies for triple-chain models with Na^+ and K^+ using different dielectric constants ϵ with the polarizable continuum model as well as the gas-phase binding energies, obtained at the B3LYP/6-31+G(d,p) level. (ϵ: ether=4.24, toluene=2.3741, tetrahydrofurane (THF)=7.4257, water=80)

A.2 Selecting the initial ion positions before optimization

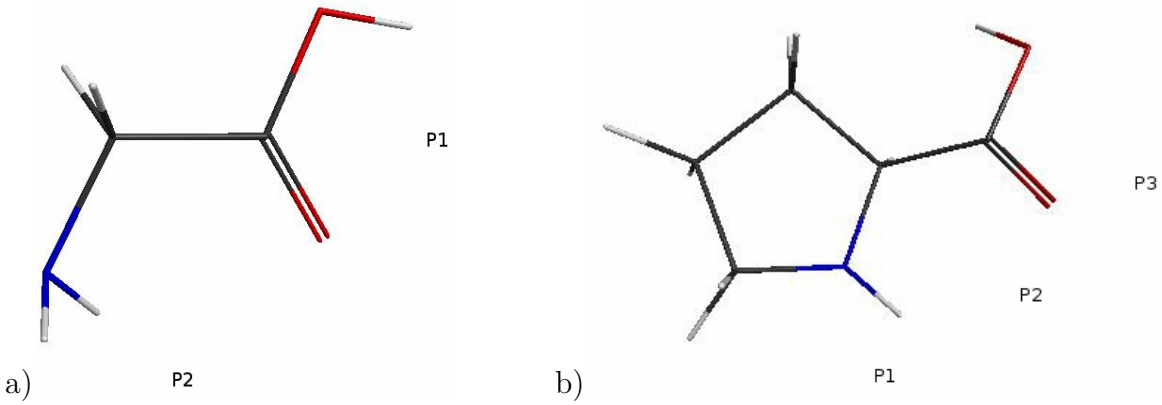

Figure A.2: Initial metal ion positions around amino acids a) glycine and b) proline. Equivalent starting positions were tested for the peptide and triple-chain models for all amino acids.

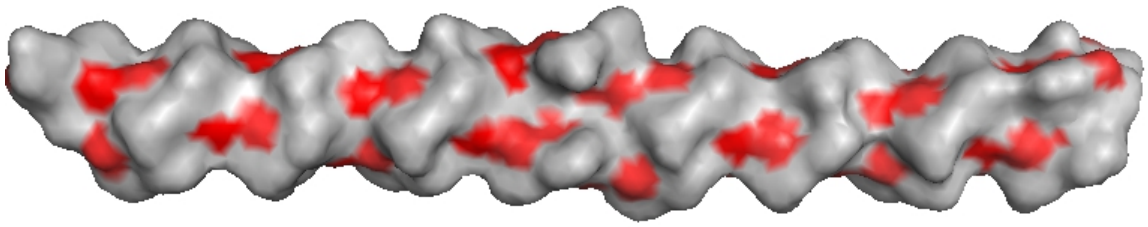

Figure A.3: Accessible carboxyl oxygens on the surface of 1CAG. This space-filling model emphasizes that much of the peptide functionality is buried. Although every third backbone carbonyl oxygen is within the triple helix, the other two are accessible to cations.

A.3 Optimized geometries, constrained and unconstrained models

Structures for models not shown in figures in the main manuscript. All ion-model distances are given in Å. K^+ ions are purple spheres; Na^+ ions are yellow spheres.

A.4 Amino acids

A.5 Peptide

A.6 Triple chain

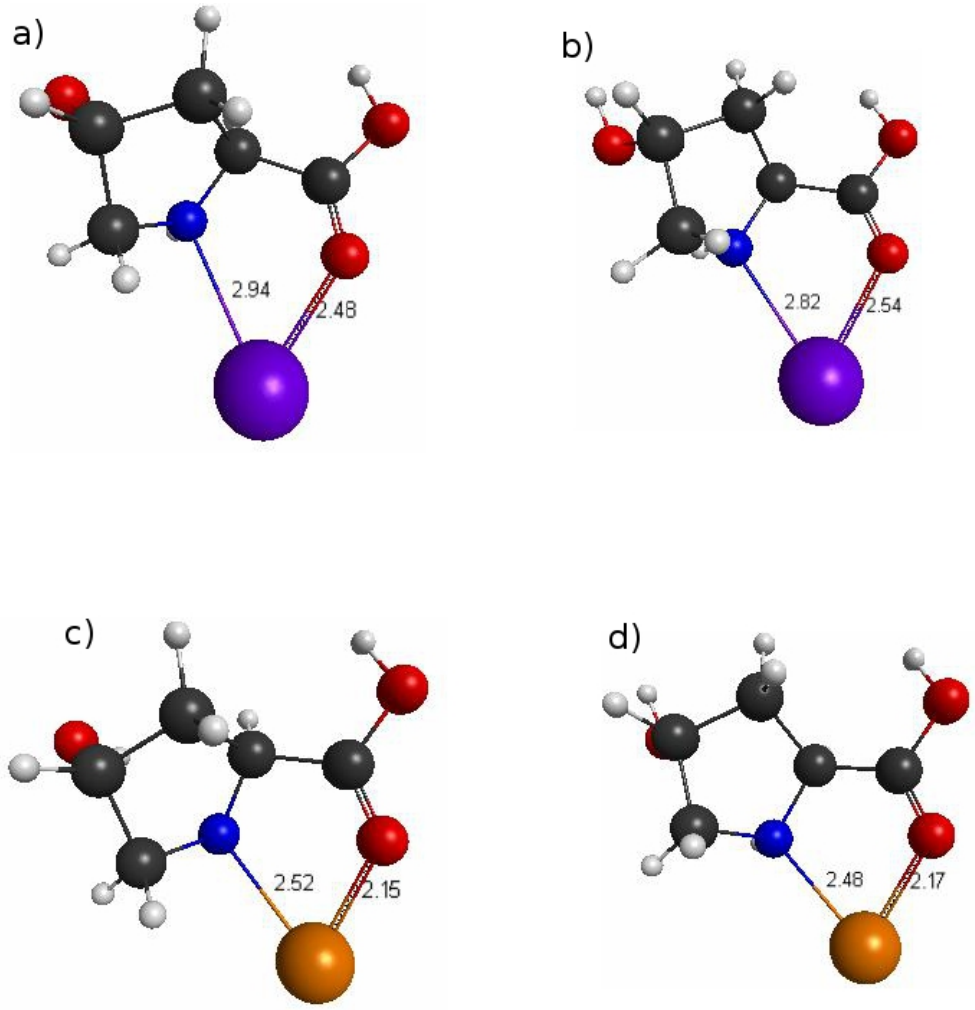

Figure A.4: Hydroxyproline-ion complexes for the amino acid models. K^+ in the a) constrained and b) unconstrained complexes; Na^+ in the c) constrained and d) unconstrained complexes, obtained at the B3LYP/6-31+G(d,p) level.

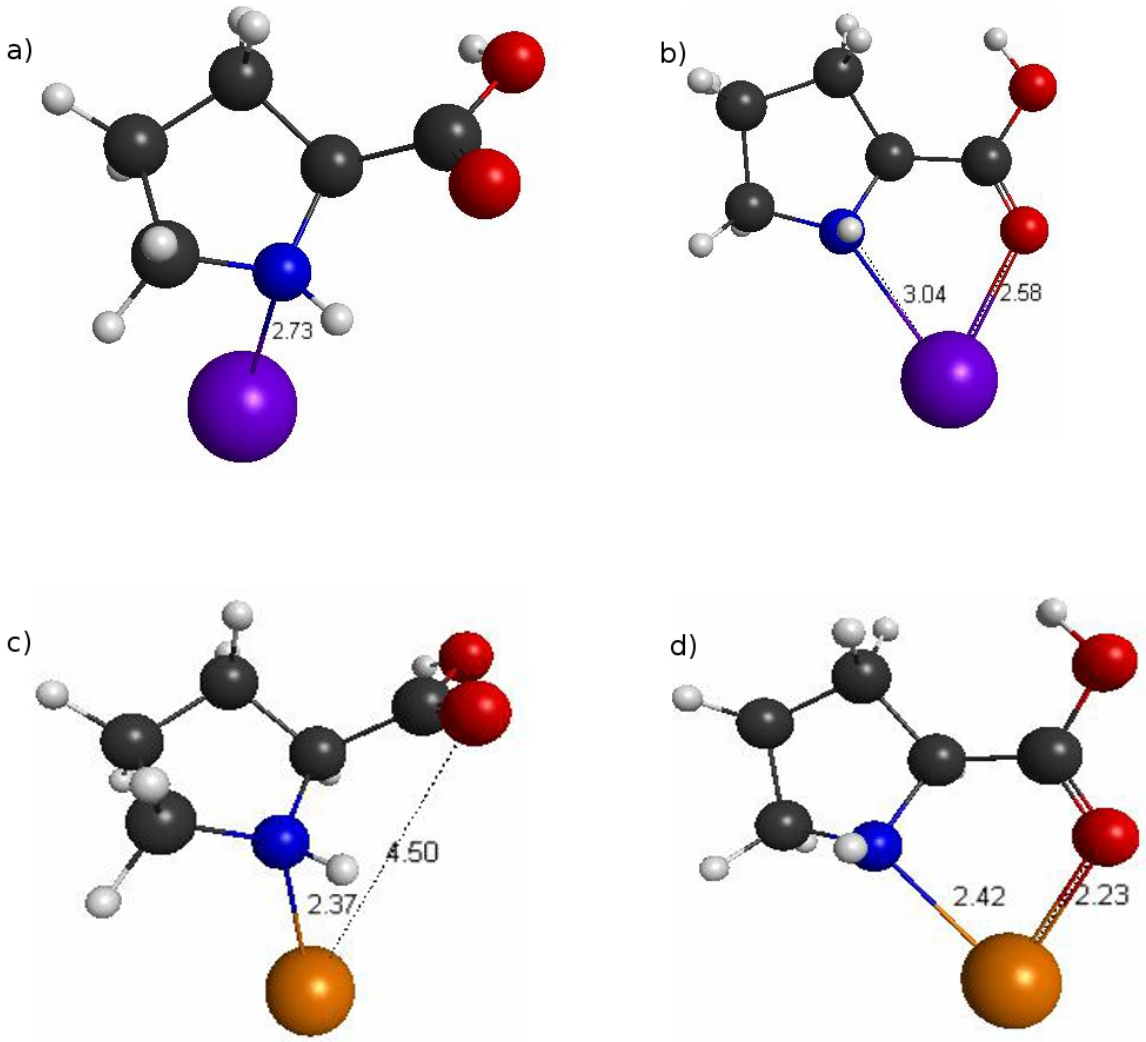

Figure A.5: Proline-ion complexes for the amino acid models. K^+ in the a) constrained and b) unconstrained complexes; Na^+ in the c) constrained and d) unconstrained complexes, obtained at the B3LYP/6-31+G(d,p) level.

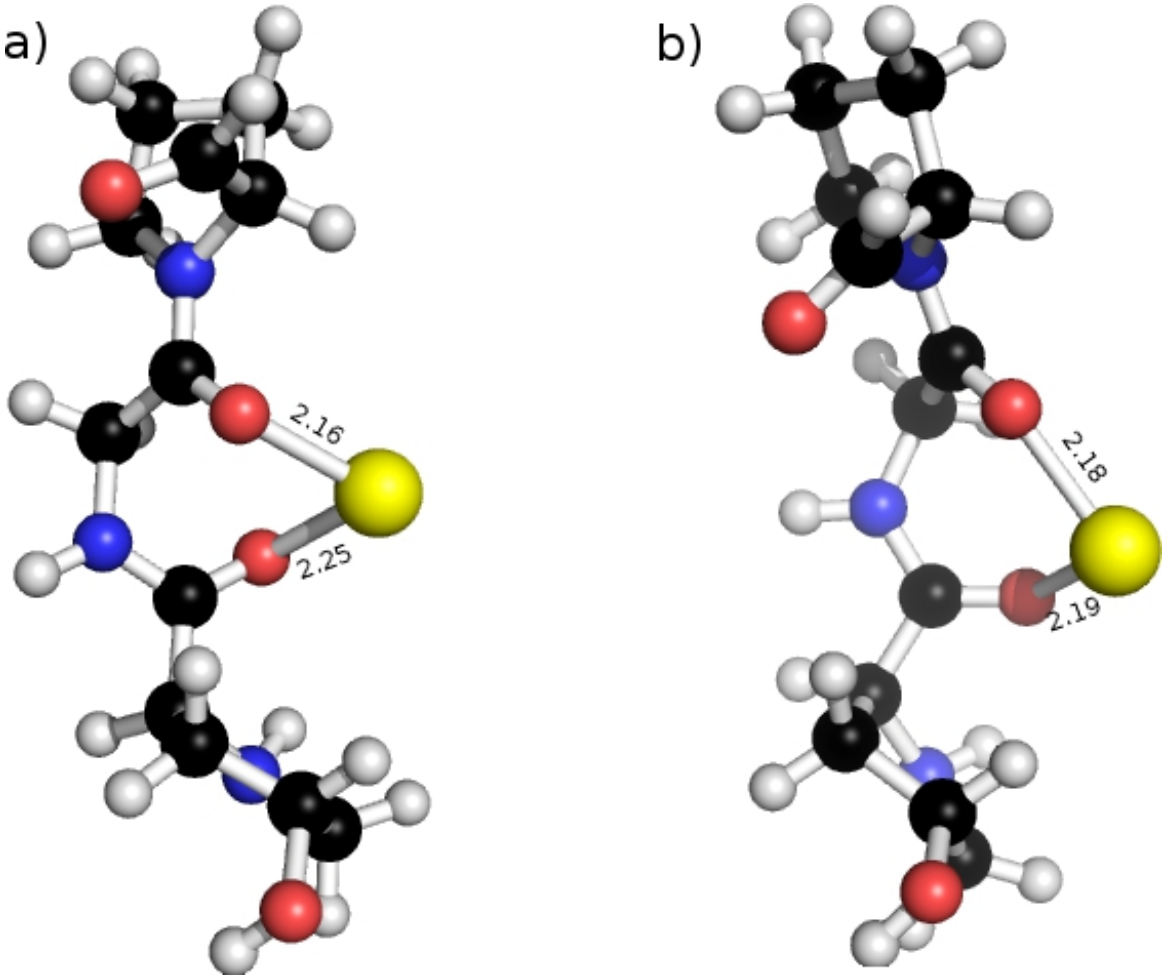

Figure A.6: Gly–Na^+ complexes for the peptide model comparing the a) constrained and b) unconstrained geometries, obtained at the B3LYP/6-31+G(d,p) level.

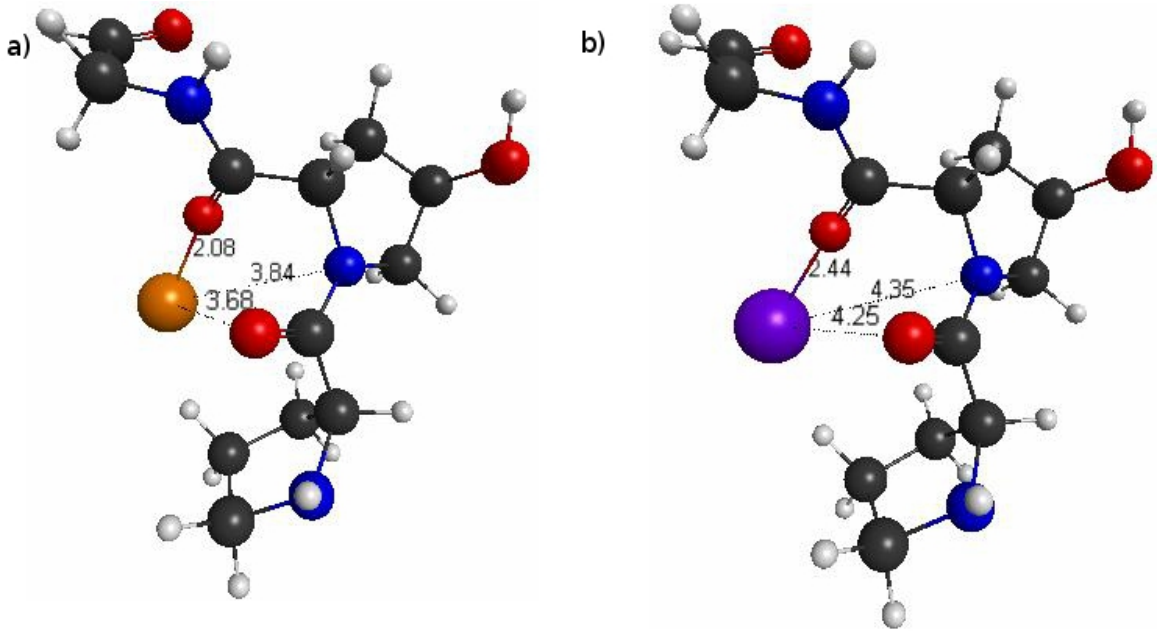

Figure A.7: Constrained a)Na^+ and b)K^+ hydroxyproline complexes for the peptide model, obtained at the B3LYP/6-31+G(d,p) level.

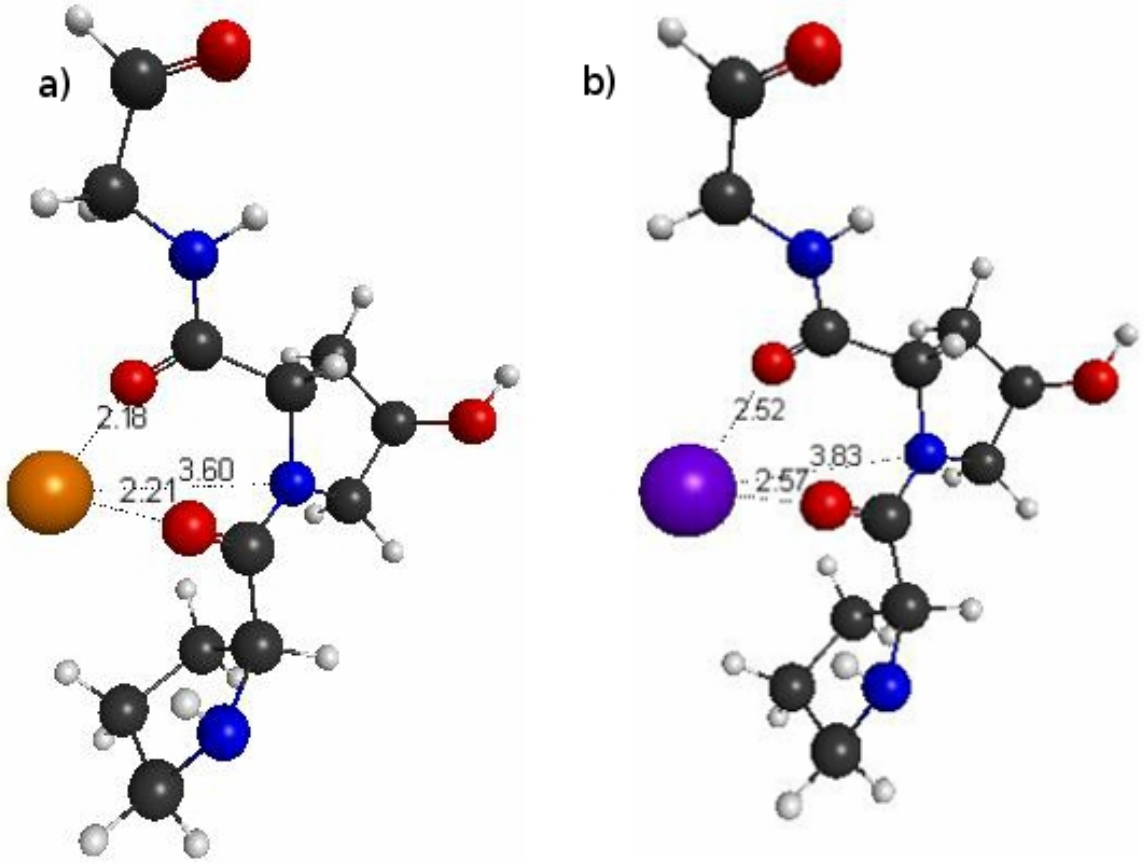

Figure A.8: Unconstrained a)Na^+ and b)K^+ hydroxyproline complexes for the peptide model, obtained at the B3LYP/6-31+G(d,p) level.

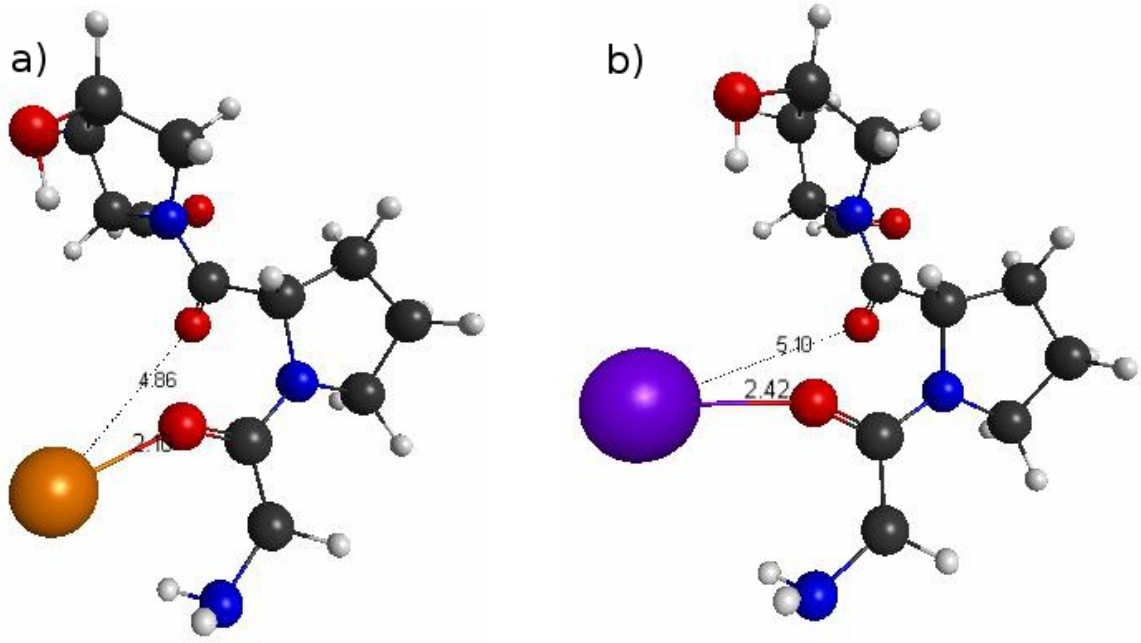

Figure A.9: Constrained a)Na^+ and b)K^+ proline complexes for the peptide model, obtained at the B3LYP/6-31+G(d,p) level.

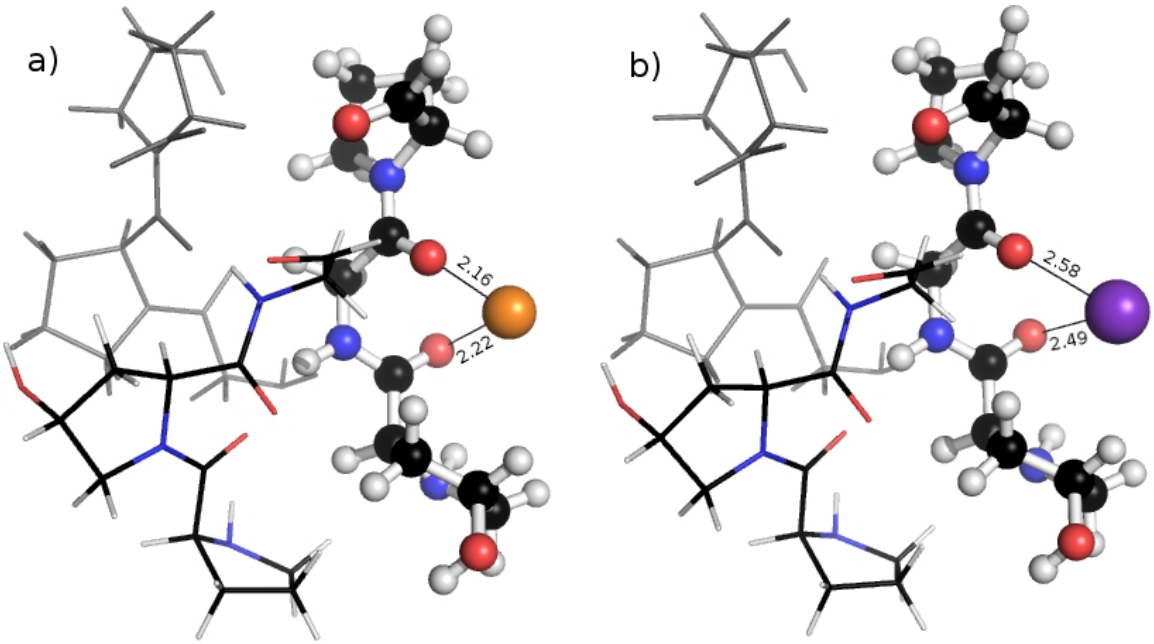

Figure A.10: Constrained a)Na^+ and b)K^+ glycine complexes in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

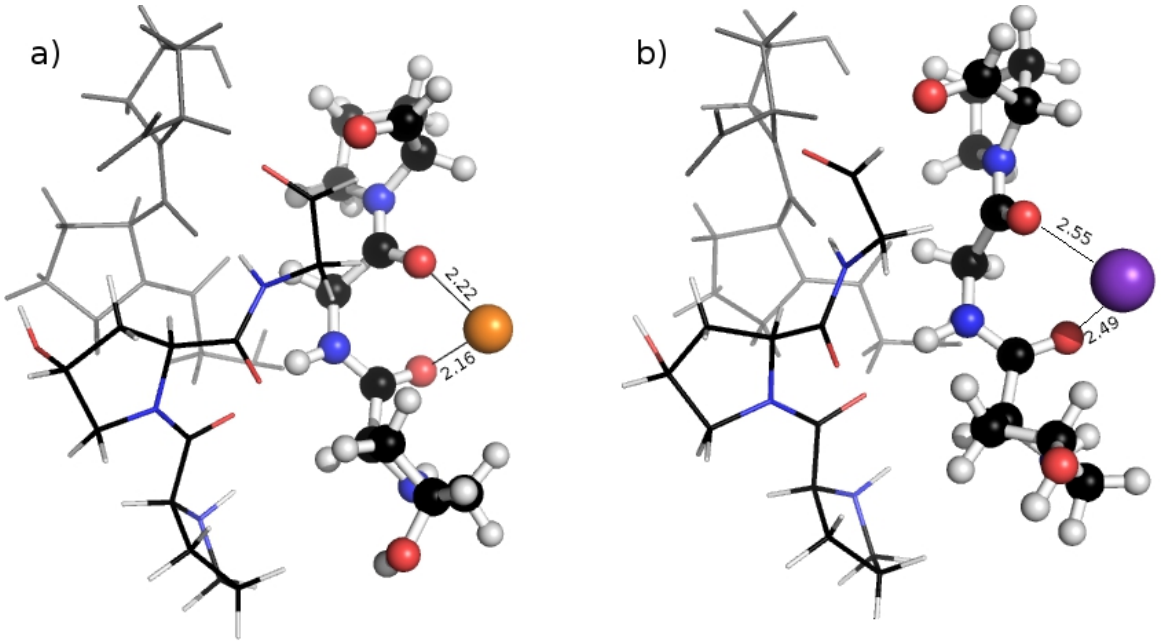

Figure A.11: Unconstrained a)Na^+ and b)K^+ glycine complexes in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

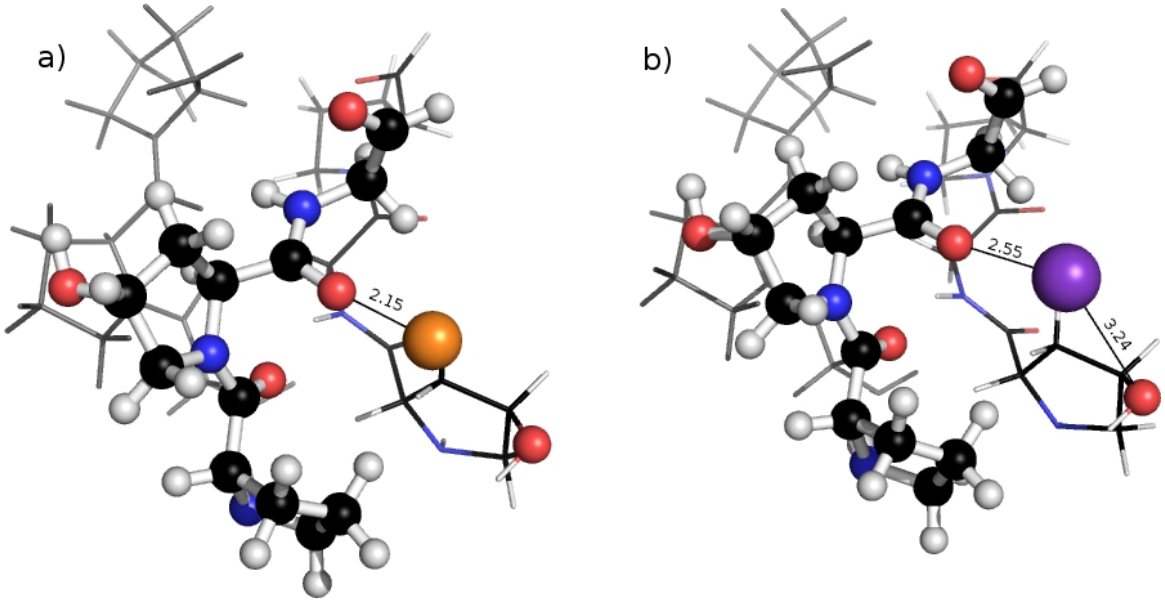

Figure A.12: Constrained a)Na^+ and b)K^+ hydroxyproline complexes in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

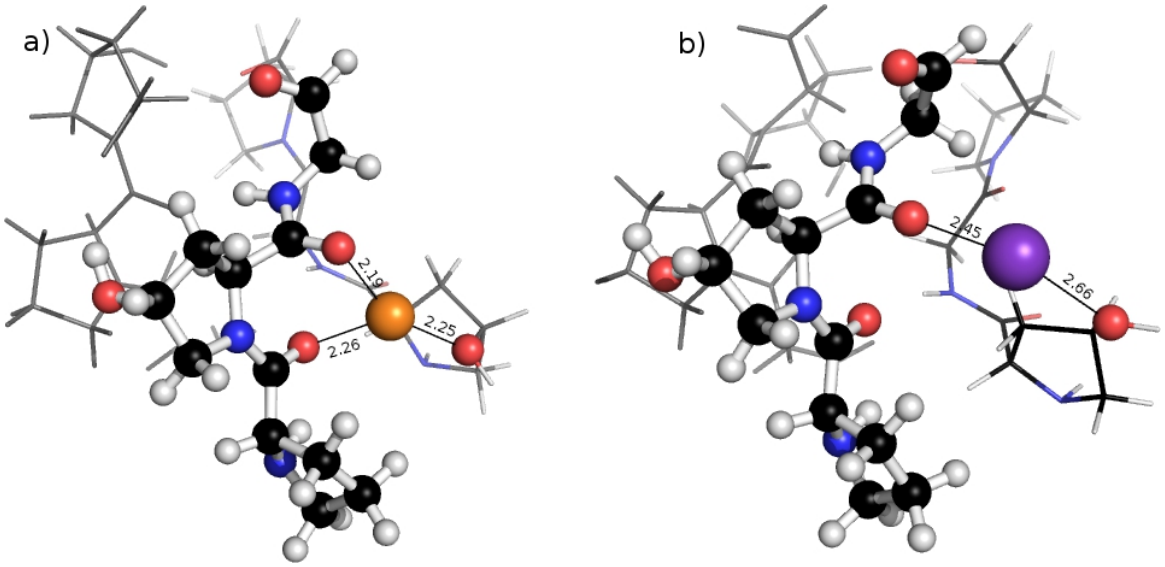

Figure A.13: Unconstrained a)Na^+ and b)K^+ hydroxproline complexes in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

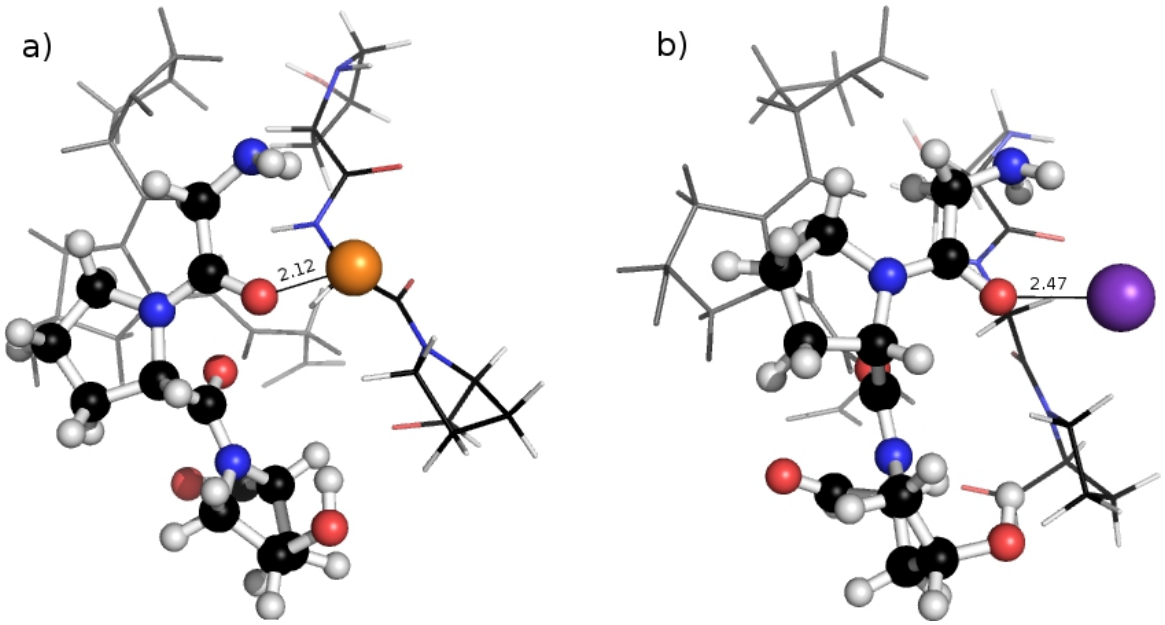

Figure A.14: Constrained a) Na^+ and b) K^+ proline complexes in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

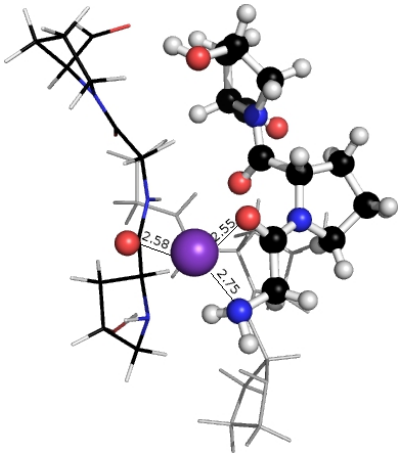

Figure A.15: Unconstrained K^+ proline complex in the triple chain model, obtained at the B3LYP/6-31+G(d,p) level.

Appendix B

Nonlinear optical properties of anthraquinone-type π-extended tetrathiafulvalene (TTFAQ)-derivatives

B.1 Calculated energies for the HOMO, LUMO, HOMO-LUMO gap for extended TTFAQ

R	E_{HOMO} (au)	E_{LUMO} (au)	ΔE_{HL} (eV)
NH_2	-0.255	0.070	8.844
CH_3	-0.265	0.063	8.925
OH	-0.266	0.063	8.953
OCH_3	-0.268	0.061	8.953
SCH_3	-0.272	0.056	8.925
SH	-0.273	0.054	8.898

Table B.1: Calculated energies for the HOMO, LUMO, HOMO-LUMO gap (ΔE_{HL}) for extended TTFAQ with R = NH_2, OCH_3, OH, SH, CH_3, SCH_3 (HF/6-31G).

Appendix C

Supporting information for "Assessing metrics of size and charge separation as predictors for non-linear optical response of organic materials"

C.1 β and γ values for TTFAQ with different R and R1 groups.

Table C.1: β and γ values for TTFAQ with different R and R1 groups (HF/6-31G).

R1-R	β ($\times 10^{-30} esu$)	$\lvert\beta\mu\rvert$ ($\times 10^{-48} esu$)	γ ($\times 10^{-36} esu$)
H-H	0.5454	0.8067	130.04
H-OH	0.7415	0.7371	120.18
H-CH_3	0.2045	0.2038	136.917
H-OMe	0.3452	0.0272	140.08
H-NH_2	6.8603	6.8842	116.092
H-NMe_2	0.1716	0.257	238.959
H-SMe	0.1385	0.171	180.487
H-CN	7.9707	12.174	207.115
H-SH	0.9425	1.4236	165.199
H-NO_2	16.1658	24.62	236.444
H-OCOMe	1.1217	0.5131	161.893
H-^{t}BuS	0.268	6.8842	224.298
NO_2-H-NH_2	58.1969	88.886	178.034
SH-SMe	0.1244	0.1853	180.179
SH-H	1.3434	1.3061	124.322
SH-CH_3	3.3832	1.3739	147.069
SH-OH	5.0855	3.5467	128.693
SH-OMe	6.2631	9.4486	141.71
SH-NH_2	13.9614	7.3854	191.149
SH-NMe_2	15.672	23.943	248.576

Continued on next page

Table C.1 – continued from previous page

R1-R	β ($\times 10^{-30} esu$)	$\lvert\beta\mu\rvert$ ($\times 10^{-48} esu$)	γ ($\times 10^{-36} esu$)
SH-^{t}BuS	2.2134	3.3618	228.37
SH-SH	2.3507	0.3167	161.59
SH-CN	7.7001	8.3704	223.285
SH-OCOMe	8.6641	1.7726	203.896
SH-NO_2	16.8886	21.72	238.775
NO_2-SH-NH_2	62.6564	93.792	226.744
SMe-H	0.513	0.774	136.98
SMe-SMe	2.17	3.31	185.37
SMe-CH_3	2.19	3.34	160.46
SMe-OH	4.4	6.71	151.27
SMe-OMe	5.43	5.87	166.2
SMe-NH_2	14.66	22.25	200.68
SMe-NMe_2	17.78	25.11	251.51
SMe-^{t}BuS	0.525	0.561	217.23
SMe-SH	2.72	4.15	179.56
SMe-CN	11.45	17.49	245.49
SMe-OCOMe	12.91	9.4	222.73
SMe-NO_2	23.93	36.56	245.61
NO_2-SMe-NH_2	61.03	91.95	213.64

Table C.2: Metrics of size for TTFAQ with different R and R1 groups (HF/6-31G).

R1-R	manual (Å)	electronic (a.u.)	no. of electrons	mass (amu)
H-H	24.35	90.21	300	580.80
H-Me	26.01	95.28	316	608.86
H-OMe	28.47	99.61	332	640.86
H-OH	25.74	94.74	316	613.44
H-NH_2	25.83	95.36	316	610.84
H-$N(Me)_2$	28.25	105.70	348	666.35
H-SMe	28.83	105.95	348	673.00
H-SH	26.34	98.88	332	644.94
H-^{t}BuS	31.12	119.84	396	757.16
H-OCOMe	29.54	109.10	360	696.88
H-CN	27.06	97.75	324	630.82
H-NO_2	26.04	103.06	344	670.81
SH-H	24.35	114.38	364	709.06
SH-Me	26.09	120.21	380	737.12
SH-OMe	28.46	125.93	396	769.12
SH-OH	25.71	120.08	380	741.70
SH-NH_2	25.83	120.45	380	739.10
SH-$N(Me)_2$	29.1	132.51	412	794.61
SH-SMe	28.86	132.10	412	801.26
SH-SH	27.35	125.37	396	773.20

Continued on next page

Table C.2 – continued from previous page

R1-R	manual (Å)	electronic (a.u.)	no. of electrons	mass (amu)
SH-^{t}BuS	31.16	147.51	460	885.42
SH-OCOMe	30.45	136.02	424	825.14
SH-CN	28.07	123.42	388	759.08
SH-NO_2	26.01	130.23	408	799.07
SMe-H	25.33	126.45	396	765.17
SMe-Me	27.15	133.45	412	793.23
SMe-OMe	29.26	138.60	428	825.23
SMe-OH	26.87	132.36	412	797.81
SMe-NH_2	26.72	138.51	412	795.21
SMe-$N(Me)_2$	29.07	145.44	444	850.72
SMe-SMe	29.9	144.28	444	857.36
SMe-SH	27.59	139.54	428	829.31
SMe-^{t}BuS	32.24	154.09	492	941.52
SMe-OCOMe	30.59	150.64	456	881.25
SMe-CN	27.92	136.93	420	815.19
SMe-NO_2	26.89	144.04	440	855.17