A Theoretical Investigation of Roles of Backbone Amide Resonance in Protein Structure

John Neville Sharley

Discipline of Chemistry, School of Physical Sciences, University of Adelaide

Submitted to the University of Adelaide in April 2016 for the degree of Doctor of Philosophy

Table of Contents

Abstract 5
Declaration 7
Acknowledgements 8
1 Introduction 9
- 1.1 *Classical view 9*
- 1.2 *Current view 10*
- 1.3 *Backbone-based view 11*

2 Natural Bond Orbital analysis 14
- 2.1 *Suitability for bonding analysis 14*
- 2.2 *Output in standard format 14*
- 2.3 *Querying XML 16*
- 2.4 *Automated construction of experiments 18*

3 Established DFT methods calculation of conjugation disturbed in the presence of torsional hyperconjugation 20
- 3.1 *Statement of authorship 20*
- 3.2 *Author contact 20*
- 3.3 *Abstract 21*
- 3.4 *Notation 21*
- 3.5 *Overview 21*
- 3.6 *Introduction 23*
- 3.7 *Methods 24*
- 3.8 *Results and discussion 24*
  - 3.8.1 Vinylamine 24
  - 3.8.2 Ethanamide 31
    - 3.8.2.1 Carbonyl orbital disturbances 31
    - 3.8.2.2 Pyramidalization at nitrogen 34
  - 3.8.3 Antiparallel beta sheet 39
- 3.9 *Conclusion 40*
- 3.10 *Acknowledgements 42*
- 3.11 *Appendix 1 43*
- 3.12 *Appendix 2 72*

4 Variation of protein backbone amide resonance by electrostatic field 75
- 4.1 *Statement of authorship 75*
- 4.2 *Author contact 75*
- 4.3 *Abstract 76*
- 4.4 *Key phrases 76*
- 4.5 *Notation 77*
- 4.6 *Overview 77*
- 4.7 *Introduction 78*
- 4.8 *Software 79*
- 4.9 *Results and discussion 79*
  - 4.9.1 Constrained to O-C-N plane 79
  - 4.9.2 Partial sp3 hybridization at N 85
  - 4.9.3 Other geometry variations 87
  - 4.9.4 Torsional steering 87
  - 4.9.5 Backbone amide nitrogen as hydrogen bond acceptor 89
  - 4.9.6 RAHB chains in electrostatic field 90
  - 4.9.7 Electrostatic field vectors at backbone amides 93
  - 4.9.8 Protein beta sheet 94
  - 4.9.9 RAHB protein helices 95

4.9.10 Proline 96
4.9.11 Permittivity 96
4.9.12 Protein folding, conformational and allosteric change 97
4.9.13 Unstructured regions 97
4.9.14 Sources of electrostatic field variation in proteins 98
4.9.15 Absence of charged residues 98
4.9.16 Development of methods 98
4.10 Hypotheses 99
4.10.1 Beta sheet 99
4.10.2 Amyloid fibril 100
4.10.3 Polyproline helices types I and II 101
4.10.4 Protein folding 102
4.10.4.1 Overview 102
4.10.4.2 Peptide group resonance drives protein folding 103
4.10.4.3 Specification of fold separated from how to fold 104
4.10.4.4 Comparison with earlier backbone-based theory 104
4.10.4.5 Observation 106
4.10.4.6 Summary 107
4.10.5 Molecular chaperones and protein complexes 107
4.10.6 Nitrogenous base pairing 108
4.11 Conclusion 110
4.12 Acknowledgements 111
4.13 Appendix 1 112
4.14 Appendix 2 132
5 Do cooperative cycles of hydrogen bonding exist in proteins? 133
5.1 Statement of authorship 133
5.2 Author contact 133
5.3 Abstract 134
5.4 Introduction 134
5.4.1 Review of Resonance-Assisted HB 134
5.4.2 Phenomena anticipated if cyclic cooperative HB exists in proteins 135
5.5 Notation 136
5.6 Methods 136
5.6.1 Counterpoise correction 136
5.6.2 Dispersion correction 137
5.6.3 Software packages 137
5.6.4 Haskell 137
5.7 Results and Discussion 138
5.7.1 HB angle 138
5.7.1.1 Experiments involving hydrogen bonding between an amide group and hydrogen fluoride 138
5.7.1.2 Use of hydrogen bonding to carbonyl oxygen where C-O-N is far from linear 144
5.7.1.3 Experiments involving hydrogen bonding between two amide groups 145
5.7.2 Inter-backbone amide hydrogen bonding C-O-N angles in the Protein Data Bank 148
5.7.3 Linear chain of formamides 148
5.7.4 Formamide cycles 150
5.7.5 Three alpha helices 153
5.7.6 Artificial beta sheets 156
5.7.7 Cyclic HB in the Protein Data Bank 157
5.7.8 Joined HB chains of beta sheet 162
5.8 Conclusion 165
5.9 Acknowledgements 168
5.10 Appendix 1 169
5.11 Appendix 2. Polyvaline Parallel Beta Sheet 183
5.12 Appendix 3. Polyvaline Antiparallel Beta Sheet 187

6 Amino acid preference against beta sheet through allowing backbone hydration enabled by the presence of cation 191

- *6.1 Statement of authorship 191*
- *6.2 Author contact 191*
- *6.3 Abstract 192*
- *6.4 Introduction 192*
  - 6.4.1 Alpha helix preferring amino acid residues in a beta sheet 192
  - 6.4.2 Cation interactions with protein backbone oxygen 193
  - 6.4.3 Quantum molecular dynamics with quantum mechanical treatment of every water molecule 193
- *6.5 Methods 194*
- *6.6 Results 195*
  - 6.6.1 Preparation 195
  - 6.6.2 Experiment 1302 195
  - 6.6.3 Experiment 1303 197
- *6.7 Discussion 199*
  - 6.7.1 HB networks of water 199
  - 6.7.2 Subsequent to rupture of a transient beta sheet 199
  - 6.7.3 Hofmeister effects 200
- *6.8 Conclusion 201*
- *6.9 Future work 202*
- *6.10 Acknowledgements 203*
- *6.11 Appendix 1. Backbone hydration in experiment 1302 204*
- *6.12 Appendix 2. Backbone hydration in experiment 1303 205*

7 Conclusion 207

- *7.1 Significance and contribution 207*
- *7.2 Future work 210*

References 211

Abstract

Protein structure, including its dynamics, is of pervasive significance in biology. A protein's structure determines its bindings with other molecules [1], and from that its roles in signal transduction, enzymatic activity and mechanical structure. Few cellular processes have no protein involvement. The relationship between a protein's sequence of amino-acid residues and its three-dimensional structure, partial or otherwise, has long been of considerable interest [2].

A theory of protein folding is proposed in Section 4.10.4 (Hypotheses/Protein folding). This theory varies and extends the backbone-based theory proposed by Rose *et al*. [3]. This theory may prove to be the most significant offering of the thesis.

Overall, this thesis investigates the variation in peptide group resonance and its implications for Resonance-Assisted Hydrogen Bonding [4, 5], RAHB, such as exists in inter-peptide group hydrogen bonding. Natural Bond Orbital [6], NBO, analysis is used for this investigation, and Section 2.1 summarizes the virtues of NBO.

Chapter 3 is concerned with methods for computational investigation of protein structures, and finds that methods and basis sets most often used for these investigations are particularly unsuitable when any beta sheets are present due to poor modelling of amide resonance and hence of RAHB that features in the hydrogen-bonded chains of backbone amides of protein secondary structures such as beta sheets [7] and alpha helices [8].

Chapter 4 reports experiments quantifying the sensitivity of amide resonance to electrostatic field with component parallel or antiparallel to the amide C-N bond. This sensitivity allows electrostatic properties including permittivity of amino-acid residue sidechains to influence backbone amide resonance and hence secondary structure RAHB chains, giving a novel mechanism relating residue sequence to structure. Also, this variation in amide resonance is energetically significant even without considering a hydrogen bonding context. Variation of peptide group resonance is expected to vary the barrier height of prolyl isomerization [9]. Subsequent to quantifying this effect in isolated amides and in a RAHB chain, hypotheses are offered concerning the stability of beta sheets, amyloid fibrils [10] and polyproline helices [11]. An analogous sensitivity in nitrogenous base pairing [12] is conjectured. A hypothesis is offered concerning protein complexation and molecular chaperone [13] action.

Chapter 5 is motivated by the observed increase of stabilization when cooperative hydrogen bonding chains are cyclized in non-protein contexts [6] and by the phenomena anticipated if these cycles exist in proteins as described in the chapter. The question of the optimal geometry for amide-amide hydrogen bonding is revisited with emphasis on the inequivalence of amide oxygen lone pairs. A possible avenue for the design of HB-chain polymers with improved stability is discussed.

Chapter 6 studies a dependency of amino acid residue preference against beta sheet secondary structure by backbone hydration in the presence of cation, doing so by Quantum Molecular Dynamic simulation of a beta sheet with a full quantum mechanical treatment of each solvent molecule.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Name: John Neville Sharley

Signed:

Dated: 2016-09-26

Acknowledgements

Prof. John A. Carver is acknowledged for reading this thesis and offering editing suggestions. I am profoundly grateful for the enlightened research environment he provided.

Prof. Peter M. W. Gill is acknowledged for reading the earliest draft of Chapter 3 and pointing out the resemblance of the second mentioned problem, pyramidalization at amide nitrogen at wavefunction method/Pople basis set, to the benzene non-planarity problem [14].

I claim that there is ample evidence here that pre-candidature mentoring by Iain Murchland and Cvetan Stojkoski resulted in public good.

eResearch South Australia is acknowledged for hosting and administering machines provided under Australian Government Linkage, Infrastructure, Equipment and Facilities grants for Supercomputing in South Australia, directing funds to the acquisition of Nvidia Tesla GPU nodes and allocating 64 CPU cores and 256 GB RAM of the NeCTAR Research Cloud (a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy) to the present work.

1 Introduction

Proteins are ubiquitous in biology and most biological processes can be expected to involve proteins in at least one regard, either as a signal transducer, enzyme, transporter or element of mechanical structure. Proteins are nature's most structurally capable molecules, assuming a diverse range of three-dimensional structures depending on the protein's amino acid sequence. The relationship between a protein's function and its biological structure has long been emphasized [1]. In a given biological environment, a linear sequence of amino acid residues often spontaneously assumes a specific three-dimensional arrangement. How it does so has long attracted considerable interest [2] and is referred to as the protein folding problem. Understanding the folding process would remove a major obstacle to achieving a molecular account of biology.

The present work addresses the protein folding problem. Four manuscripts are presented as Chapters 3 to 6, and the common theme is the relationship of peptide group resonance to protein structure. Chapter 3 assesses the suitability of quantum chemical methods usually applied to proteins for calculating peptide group resonance. Chapter 4 demonstrates a new factor in protein folding using suitable quantum chemistry methods. Additionally, a theory of protein folding is proposed in Section 4.10.4 (Hypotheses/Protein folding). This theory extends the backbone-based theory of protein folding proposed by Rose *et al*. [3] and posits a mechanism by which amino acid residue sequence is evaluated as specification of protein fold. Chapter 5 investigates the existence of cooperative cycles of Resonance-Assisted Hydrogen Bonds [4, 5], RAHBs, in proteins. If such cycles exist, they would be stabilizing of proteins by means other than recognised secondary structure types such as peptide group hydrogen bonded helices [8] and beta sheets [7] and thereby introduce complexity to folding. Chapter 6 demonstrates a cation dependency in the sidechain blocking of backbone hydration associated with amino acid preference for beta sheets [15].

Context for this thesis is offered below, starting with a brief history of the classical era of protein folding, proceeding to review the failings of the current view of protein folding and then onto a backbone-based theory of protein folding proposed by Rose *et al*. [3].

1.1 Classical view

Independent proposals that protein was composed of polypeptides by Franz Hofmeister and Emil Fischer coincided on the same day of 1902 [16]. Fischer gave the name peptide to the bonding of the carboxyl group of one amino acid to the amino group of another. The first amino acid to be discovered was asparagine in 1806 [17], and the last of the 20 common acid was threonine in 1935 [18].

Hydrogen bonding was first mentioned by Moore and Winmill in 1912 [19]. The search for regular patterns of hydrogen bonding between peptide groups culminated in the discovery of alpha helices in

1950 [8] and beta sheets in 1951 [7]. In 1935, Linus Pauling proposed that H-bonds were partially covalent rather than classically electrostatic [20], which was experimentally confirmed in 1999 [21].

The reversible denaturation of proteins was a matter of interest from the 1930s onward [22]. The investigation of the relationship between polypeptide amino acid sequence and biologically active conformation of the polypeptide reached a milestone in an experiment performed by Anfinsen *et al*. [23] in 1961. Taking a denatured polypeptide, ribonuclease, which is enzymatic only in its biological conformation in which one specific set of pairings of cysteine residues are disulfide bonded, they showed that the biologically active conformation was restored in the presence of a disulfide bond interchange agent which was not specific to the protein. A "thermodynamic hypothesis" was proposed in which the biologically active state is uniquely defined as having the global minimum Gibbs free energy.

The dihedral angle of the polypeptide backbone at the peptide bond, C-N, is referred to as omega, the backbone N-CA bond dihedral is referred to as phi and the backbone CA-C bond dihedral is referred to as psi. If the N in the peptide bond is not part of the imino acid proline, the backbone amide group H-N-C-O approximates planarity [24, 25], and so the omega dihedral may then be omitted from consideration of backbone conformation. If the N in the peptide bond is part of proline, omega is less well defined, but tends to one of two conformations, cis or trans [26]. The "Levinthal paradox" [27] is a rough estimate of the time for a protein to find its global minimum free energy by the exhaustive search of conformations based on the assumption that the phi and psi dihedrals of a backbone, though sterically restricted, are essentially independent. This assumption leads to estimates of folding times that are absurdly long, and is taken to imply that the number of pathways that folding may take between the folded and unfolded state is distinctly limited, which is the essence of the classical view of protein folding. Much thinking about protein folding has been guided by conclusions prompted by this estimate.

1.2 Current view

The key question in protein folding has been: is the folding pathway clearly and perhaps uniquely defined and this pathway responsible for the production of specific structure, or are there arbitrarily many conformational paths a protein could take to fold to the same specific structure? This second view was referred to as the new view [28] but might now be referred to as the current view. An objection has been raised to the interpretation of experiments that support the current view [29]. Rose *et al*. [3] offers different objections to the current view, and a summary of these is given below.

Spin glasses were applied to the study of behaviour of folding with arbitrary pathways to a unique solution [30]. A spin glass consists of roughly equal numbers of ferromagnets and antiferromagnets of fixed position that may take on irregular orientations. The orientations of the magnetic spins change as the spin glass descends into an energy minimum. Such a system is metastable, that is, it is stable in local minima rather than only in the global minimum. This metastability is referred to as frustration. Beneath a transition temperature, the spin glass commits to the “basin of attraction” [31] of a minimum. The only remedy for descent into a local minimum is to retry by “simulated annealing” [31] in which the temperature is raised above the transition temperature and lowered again. Determination of the global minimum is a nondeterministic polynomial time-complete problem [32], in which exhaustive search of all possibilities is required to be certain of having found the optimal solution.

The principle of minimal frustration assumes that sidechain-sidechain contacts that are not native in biological structure do not cause metastable states during folding, giving a funnel-shaped folding landscape having more conformational restriction down the funnel [33, 34]. There is a historical context for this assumption, as it was used in the folding models of Go [35]. This assumption is contrary to the essence and purpose of spin glasses, for frustration is the essence of the spin glass problem [36]. With this assumption, the folding has more in common with unfrustrated ferromagnets than spin glasses. Frustration is what Levinthal sought to address in introducing fold-defining pathways. The current view is flawed.

In the current view, there are no general principles for amino acid sequence satisfaction of the principle of minimum frustration and each sequence meets this principle in its own way. Individual sequences carry the burden of stabilizing a fold relative to other folds and minimizing frustration on the way to its fold. The space of possible sequences is large, and in the current view, each biological sequence has been evolved to solve the protein folding problem anew.

The classical and current views might be described as sidechain-based, a plausible emphasis given the chemical structural uniformity of the backbone and Anfinsen’s premise that information for folding is given by amino acid sequence. A pre-occupation with sidechain-sidechain contacts and formation of native sidechain-sidechain contacts follows within this view. However, Rose *et al*. [3] describe a differing view of protein folding being backbone-based folding.

1.3 Backbone-based view

A brief summary preparatory to this thesis is given of Rose *et al*.’s [3] argument in favour of a backbone basis for protein folding.

Organic osmolytes interact primarily with the polypeptide backbone [37] in the unfolded state [38], with some osmolytes such as glycerol being protective of protein structure and others such as urea being destabilizing of protein structure. The uniform influence of these osmolytes on all proteins is consistent with the existence of a universal mechanism which may be taken to be related to the common component of amino acid residues, the backbone [3].

The principal inter-peptide group hydrogen bonded secondary structures, alpha helices and beta sheets, may be extended indefinitely. If folding is backbone-based and sidechains are secondary, provided that inter-peptide hydrogen bonds are more favourable than peptide-water hydrogen bonds, then these extensible secondary structures will tend to spontaneously form. Delimited by dipeptides less favourable to a particular secondary structure type and assisted by non-extensible inter-peptide hydrogen bonding in the form of beta turns [39], combinations of extensible secondary structures form. The number of secondary structure instances in an independently foldable module or domain [40] is low, often single digit, so the number of ways they may be composed to form tertiary structures is limited, a point discussed further in Section 4.10.4.4 of the present work. The secondary structures diffuse, collide [41] and anneal [42] to arrive at a tertiary structure. Secondary structures are locally determined, and these determine protein topology [43]. This hierarchical assembly gives a funnel-like folding landscape. Metastable traps are possible [44], but maximization of backbone hydrogen bonding is posited to be the mechanism by which metastable traps are largely avoided. The mechanism of folding does not primarily rest on evolution of folding-capable residue sequences.

In two-state folding, folding conditions may be varied to change the ratio of individual molecules in the two states. Such a change to folding conditions does not cause the assumption of a different fold. What determines native fold is present throughout the varying folding conditions. Stability is independent of conformation [3].

Rose *et al*. [3] define a protein's fold as being given by its backbone hydrogen bonding scaffold and structure as given by coordinates of all atoms. With this definition, fold is experimentally determined to have formed before close packing of sidechains, a state referred to as a molten globule [45, 46]. That hydrogen bonded backbone scaffolding is observed to form before stable sidechain interactions [47] is significant. Inter alia, it suggests that close packing and exclusion of water are not closely associated with fold [47]. The conformational entropic cost of forming the backbone scaffold is reduced because the motion of sidechains is not restricted as in a close packed arrangement. The entropic cost of backbone scaffold formation is further reduced by restriction of backbone conformation, but not sidechain conformation, in polyproline helix type II secondary structure [48] which frequently occurs in unfolded proteins.

Rate of folding has been shown to increase with increase in total content of the local inter-peptide hydrogen bonds that exist in alpha helices, beta hairpins and beta turns [49]. Whereas the thermodynamic hypothesis views folding as a continuous variation of free energy, in a backbone hydrogen bonded account, it is discontinuous, for backbone hydrogen bonds are made and broken.

Inter-peptide group hydrogen bonding is enthalpically favoured over peptide group-water hydrogen bonds by 1-2 kcal/mol [50, 51] and so inter-peptide group hydrogen bonding promotes chain collapse, though this was opposed by Kauzmann who posited that hydrophobia largely determined protein stability [52]. There is still not universal agreement on this most important point. Examples of natively unfolded proteins folding in the presence of organic osmolytes are known [53], and the free energies of transfer of sidechains and backbone from water to osmolyte can be obtained and the dominance of the backbone contribution shown [54]. Since the peptide group is without hydrophobic components, this dominance does not involve increase in hydrophobic interactions. The folding cooperativity of whole proteins can be predicted, with the backbone being the primary contributor [54].

Rose *et al*.'s [3] proposal of a backbone-based theory of protein folding was published a decade ago in 2006, enough time for the field to have provided an initial response. No more convincing proposal has been published in that time.

2 Natural Bond Orbital analysis

2.1 Suitability for bonding analysis

Natural Bond Orbital [6], NBO, analysis was used for experiments appearing in Chapters 3, 4 and 5. Chemical bonding analyses additional to NBO were not used, as justified in a very recent review [55], a reading of which is recommended and indeed required if it is believed that comparison of NBO results with those of other methods might be informative or revealing. The review describes the unique features of NBO methods before undertaking specific comparison of NBO with QTAIM and EDA methods.

Atomic orbitals are those of an approximation to idealised free atoms rather than molecular atoms and are an unsuitable basis for bonding analysis. The shapes and energies of occupied molecular orbitals are not uniquely defined under arbitrary unitary transformations as per Fock's theorem [56]. Natural Atomic Orbitals, NAOs, capture effective atomic orbitals in the molecular context and are uniquely defined. NAOs are then used in the orthonormality-preserving construction of NBOs.

In maintaining strict orthonormality, NBO preserves Pauli exclusion and Hermitian energetics and so NBO is free of the overlapping attributions arising from orbital or wavefunction non-orthogonality. NBO is based only on the *eigen* or intrinsic properties of the interactions of the electronic first-order reduced density matrix (1-RDM), rather than any assumptions concerning geometry or symmetry and is independent of the form of the original wavefunction. This basis in the *eigen* properties of the 1-RDM yields consistency and predictive capacity, uniquely qualifying NBO for chemical bonding analysis.

2.2 Output in standard format

Manually transcribing numbers from output to a spreadsheet or other form of display becomes most unsatisfactory as the amount of data to be transcribed rises, for this activity is time consuming, has a non-zero error rate and must be repeated should the experiment conditions be varied in any way and the experiment performed again. Where chemical bonding analysis outputs large amounts of data, transfer of this output to other programs must be automated.

At NBO 6.0 [57] the reports are still designed for human reading. A Haskell [58] program was written to read this human-oriented output. This is an unreliable approach in any language including text-oriented programming languages, for any enhancement or bug fix to NBO may change the number, spelling or capitalization of words on a line, a field's start and end position on an output line if fixed positions are assumed, or vary labels or formatting used to determine the start, subsections, line type and end of a report section within output. For example, to maintain human readability, line widths are kept low, and one way this is done is to minimize the width of integer fields such as atom and orbitals identities. When these widths need to be increased, every output report must be changed, and a program reading the output must be changed if it is dependent on positions in a line and tested even if it is not believed to have such a dependency. Each NBO enhancement or bug fix may change output formatting, so any change to NBO must be accompanied by a check of the format of all output and change to the program which reads NBO output. Also, if NBO output is saved for later use, the version of the extraction program suitable for that version of output must be saved as well. The reliable maintenance of the extraction program is too burdensome for the approach to be viable for more than a short term. This issue is not confined to NBO, and it is widespread practice to write programs to extract data from reports designed for human reading, a practice which places the burden of (unsatisfactorily) dealing with data extraction by non-standard means on each user, which repeats the effort, rather than having it dealt with once by each analysis package. Further, this effort by one user is repeated for each analysis package used, which does not encourage use of several analysis packages such as when combining the output of multiple packages into an integrated report. Reliable data extraction from output is necessary for robust automation of presentation and discovery.

The alternative to extraction of data from output formatted solely for human reading is to change the output generating program to produce output formatted in accordance for standards for reliable reading by program. A format commonly used for this purpose is Extensible Markup Language [59], XML, standardized by the World Wide Web Consortium [60], W3C. Each data element in an XML file is described by a hierarchical path of element names, and the width of each data element is given by the unambiguous tags enclosing the data of each element. Standardized facilities exist for checking whether an XML file is formatted in accordance with the standard and for querying the contents of an XML file. An XML file might be seen as a portable, transmissible and standard database, supporting standards-based extraction, querying and reformatting to suit a given end use.

The source code of NBO is available, and modification of NBO to output XML for the data needed for the experiments of the present work was undertaken. The salient features this modification are use of the FORTRAN [61] I0 format descriptor for integer fields, for this gives an output field of the minimum width to describe a particular integer value which is then enclosed in XML tags, essentially implementing a variable-width field. Maximum width double precision floating values in scientific format were output to XML, so again there was no restriction to field width given in FORTRAN or XML other than by the range of integer and double precision floating values intrinsic to the primitive data types of FORTRAN.

In writing XML, the Document Object Model [62], DOM, might be used so that an XML document under construction can be navigated by the creating program and can have elements inserted where appropriate in the document rather than only appended. This navigation of the XML document under construction would have increased the extent of modification of the NBO program, and it was chosen instead to use FORTRAN formatting to sequentially write XML to an output file. This means that combination of XML elements to form structures other than those occurring sequentially in the course of processing cannot be done in the NBO program. However, the XSLT [63] language is suitable for such transformations of the XML document after production by NBO, though no need to do so has yet been encountered and the sequentially produced XML is presently canonical for querying. An argument favouring the use of DOM for document creation is that the XML standard may change and so it is better to have formatting external to the program creating the XML document. This argument is thin, for changes to the XML standard are most infrequent and older versions of the standard remain readable by query languages. The parts of the NBO program creating XML formatting have been localized to callable modules, which is useful if the formatting were ever to change.

The problem then remaining is that changes for XML writing must be reapplied to NBO when any enhancement or bug fix is applied to NBO that changes a module also changed for XML writing, which means scrutinizing every NBO enhancement or bug fix to see if there is any intersection of modules being changed with those changed for XML writing, and if so merge the changes for the fix and for XML. This merging must be tested. The maintenance cost of this approach is again too high to be borne for more than the short term. The solution was to contribute the changes for output of XML to the NBO project, and XML output will be supported in a future NBO release after the data appearing in XML output is extended from that supporting the present work to all possible investigations. Upon delivery of that release, NBO output can be queried in a facile manner by all users.

It is proposed that analysis programs of any kind produce only output compliant with standardized formatting, and that the standard be XML. Where reformatting to ease human reading is advantageous, it can be had by supplying a query of the standards compliant output. In this way, the analysis program does no formatting for human readable output, and is only concerned with writing output in the standardized format. Where a user wishes to customise the human readable format, they modify the supplied query and do not have to modify the analysis program.

2.3 Querying XML

Having obtained NBO output in XML format, the question is then how to best get desired subsets of that data into a form suitable for spreadsheets or visualization tools. Visualization was used extensively in the discovery phase of the present work, and its more widespread use for scientific purposes has long been advocated [64]. Use of XSLT is conventional for document to document transformation, and this was used early in project whereas XQuery 3.0 [65] was used later. Except where the aim is to change the structure of an XML document, XQuery is to be preferred over XSLT, for its processing model is more general purpose than that of XSLT. Both XSLT and XQuery are of the functional programming paradigm [66], as is the yet more general purpose programming language Haskell referred to Chapter 5. The advantages of the functional programming paradigm are well documented [66].

XQuery was used to query the output of multiple NBO analyses to produce a single small XML file that was then imported to a spreadsheet for presentation of tables and charts. The import was not automated since this would require a commitment to the application programming interface of a particular spreadsheet package. An XQuery program was written for each type of molecule, with lesser variation for each kind of experiment. These programs will be offered as contribution to the NBO project to appear in an example program directory, along with an 'nbolib' library which is a shared module of reusable XQuery functions.

XSLT programs that write X3D [67] definitions of interactive three dimensional molecular visualizations from NBO XML were written. These visualizations are partially abstract [68] in that atoms are shown by nuclear coordinates and atomic number as is conventional, but representation of orbitals is abstract rather than being at all representative of spatial extent. The motivation for these visualizations was that the number of orbitals and orbital interactions appearing in the NBO output of a molecule the size of a protein is large, and threading together networks of donor-acceptor interactions by inspection of textual output is quite inconvenient. It is a great advantage to have a supplied visualization of networks of interactions between NBOs. Two-center NBOs were depicted as a narrow cylinder between the two atoms, with polarization marked on the cylinder. Charge transfer is shown as a cylinder between the polarization mark of the donor NBO to the polarization mark of the acceptor NBO. The polarization marks of an NBO and its antibonding NBO usually differ, so the representations of charge transfers between NBO 1 and NBO 2* and NBO 2 and NBO 1* do not overlay except where both NBOs are completely unpolarized. The colour of the charge transfer cylinder is graded with the kcal/mol value of the interaction. The labels and attributes of the atoms, NBOs and donor/acceptor interactions are shown by clicking on the object of interest. For interactions between 1-center and 2-center NBOs, the 1-center endpoint of the interaction is given as a point on the van der Waals-proportional atom surface. 1-center to 1-center interactions are not shown on this display, though a separate display could readily be written for these. In one display, both donor-acceptor and steric interactions were shown, taking advantage of the fact that the steric interactions are not shown for acceptor orbitals so that charge transfer interactions do not occlude steric interactions or vice versa.

A similar X3D writing XSLT program was written to visualize NBO and Natural Localized Molecular Orbital [6] dipole moments. This visualization is preferable to appreciating spatial orientation given in textual form. In general, relationships and irregularity is more readily seen in graphical visualization than in text form [64].

This approach, using XSLT or XQuery to write X3D from XML produced by an analysis program, is highly recommended. While writing X3D from XML is a conventional application for XSLT and XSLT was used, in future the suitability of XQuery for this purpose will be explored. Using X3D to define displays and interactions and using a third-party X3D implementation means that a 3D visualization program does not have to be written and a language defining visualization has already been specified.

All of the XSLT and XQuery programs developed in the present work will be offered as contributions to the NBO project, in addition to the modifications for XML writing already contributed.

For an XQuery/XSLT implementation, Saxonica's [69] products are suggested for their up-to-date standards compliance and extensive testing. The present work used Saxon PE rather than Saxon HE, as at the time HE did not have support for mathematics functions such as the trigonometric functions or square root. The most recent versions of Saxon HE include mathematics functions for XQuery and XPath [70], which is another consideration in favour of XQuery rather than XSLT. Use of the most recent versions of XQuery and XPath is encouraged, and presently the most recent version of both is 3.1.

For an X3D implementation, Instantreality [71] is suggested, being free, fast on modern hardware and quite reliable with problems limited to a few bugs introduced in recent releases and a long standing limitation on the length of time that an object's text description stays in the browser description line when the pointer device is moved over the object.

2.4 Automated construction of experiments

The foregoing discussion outlines the automation of use of NBO results up to the point of importation into a spreadsheet or visualization program. Construction of atomic coordinates and electronic structure system, ESS, and NBO parameters for jobs leading to NBO analysis was also automated. Thousands of quantum chemical experiments were conducted in the present work, and the number of jobs in each experiment was up to a thousand, and manually preparing the input for these jobs would have been prohibitive. The automation gave consistency and made it readily possible to vary an experiment and rerun the experiment's jobs.

Each experiment-level parameter file, which is suffixed by the name of the script that interprets it to produce many job-level parameter files, was kept somewhat general by parameterizing the model coordinate and constraint sets and the identity of atoms used as reference points for relative placement of coordinate sets, rotations and translations. Each experiment-level parameter file referred to a file giving a list of methods each with selected parameters and a file giving a list of basis sets, and the corresponding experiment-level script generated job parameters for each combination of method, basis set and ESS-input coordinate and constraint files. The directory scheme used was first level for the experiment parameter file, second level to group jobs of a common starting molecular conformation and third level for all combinations of ESS job parameters within that starting conformation. Two ESSs were supported for full automation including job submission, being Orca [72-74] and Gaussian [75]. Utilities were written to check for normal job termination of all jobs in an experiment and to bring all input files, output XML and coordinates back to a workstation. The XML was automatically post-processed to incorporate job parameters and where relevant, PDB atom name and residue sequence numbers, so querying by protein naming conventions was possible.

Central to all coordinate set generation programs was the shared use of a cluster of C++ [76] classes that were written to build and manipulate coordinate sets. This coordinate set manipulation library used the Stanford Network Analysis Platform [77] to determine on which side of a bond an atom was connected in the graph of bonds, so torsion about a bond could apply only to atoms connected to one side of the bond. Programs using this coordinate manipulation library created chains or rings of molecules of differing molecule count at definable relative geometry.

3 Established DFT methods calculation of conjugation disturbed in the presence of torsional hyperconjugation

3.1 Statement of authorship

Title of Paper	Established DFT methods calculation of conjugation disturbed in the presence of torsional hyperconjugation	
Publication Status	☐ Published ☐ Submitted for Publication	☐ Accepted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	arXiv:1512.04171 First version appeared on arXiv (arxiv.org) on 2015-12-14.	

Name of Principal Author (Candidate)	John Neville Sharley		
Contribution to the Paper	all		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	2016-09-26

3.2 Author contact

john.sharley@pobox.com

3.3 Abstract

Accurate treatment of amide resonance is important in electronic structure calculation of protein, for Resonance-Assisted Hydrogen Bonding [5, 78, 79], RAHB, in the hydrogen bonded chains of backbone amides of protein secondary structures such as beta sheets [7] and alpha helices [8] is determined by amide resonance. Variation in amide resonance is the means by which the hydrogen bonding in these chains is cooperative.

Amide carbonyl orbitals are revealed by Natural Bond Orbital [6, 57, 80], NBO, analysis to substantially maintain sigma/pi separation in the presence of torsional hyperconjugative [6] interactions with wavefunction methods but not with established Density Functional Theory [81, 82], DFT, methods. This DFT error is most pronounced with small basis sets such as are used with DFT for proteins to reduce the basis function count. This error disturbs calculation of a range of amide donor-acceptor and steric interactions.

This finding has important implications for the selection of electronic structure methods and basis sets for protein calculations. For example, great caution is needed in interpreting the results of applying established DFT methods to proteins containing any beta sheets. We recommend that every protein DFT calculation be accompanied by NBO assessment of maintenance of amide carbonyl sigma/pi separation and absence of carbonyl bond bending. Further, we propose that these metrics be standard benchmarks of electronic structure methods and basis sets.

3.4 Notation

"->" denotes NBO resonance-type charge transfer and "|" denotes NBO steric exchange repulsion. "(" and ")" enclose specification of an orbital type and follow an atom name for single-center NBOs and a pair of atom names separated by "-" for two-center NBOs.

Examples: N(lp) for the amide nitrogen lone pair NBO, O(lp-p) for the oxygen p-type lone pair NBO, O(lp-s) for the s-rich lone pair NBO, C-O(p)* for the pi carbonyl antibonding orbital NBO and N(lp)->C-O(p)* for the primary amide resonance type charge transfer.

3.5 Overview

NBO analysis provides an optimal account of correlated electron density which is useful for determining differences in electron density. Differences in NBO-derived quantities are not arbitrary with respect to electron density. NBOs are not unitarily equivalent to molecular orbitals [83].

NBO is not committed to maintenance of sigma/pi separation of orbitals of multiple bonds. The NBO account of sigma/pi separation varies according to electron density. NBO analysis may reveal loss of sigma/pi separation in the presence of angular strain or chemical bonding. In the case of double bonds having hyperconjugative interactions, there exists a significant difference between the electron density calculated by established DFT methods versus MP2 [84] and in the extent of sigma/pi separation. This context is not particularly multireference. MP2 largely maintains sigma/pi separation and absence of bond bending, while DFT methods show large variation in the extent of sigma/pi separation and bond bending. This DFT error is sensitive to basis set and in vinylamine is reduced at the correlation consistent basis set [85] level aug-cc-pV5Z though remains significant at aug-cc-pV6Z.

Calculated amide electronic interactions disturbed by this error are not limited to loss of sigma/pi separation and carbonyl bond bending, and include variation of charge transfer from the nitrogen lone pair to carbonyl antibonding orbitals and steric interactions C-O(s)|C-O(p), O(lp-s)|C-O(s) and O(lp-s)|C-O(p).

Vinylamine, ethanamide and polyvaline and polyalanine antiparallel beta sheets are used for these investigations. Ethanamide is used as a minimal example of the amide group having hyperconjugative interactions with the carbonyl group. Results of applying 43 non-double hybrid DFT [86, 87] methods in combination with 9 basis sets are reported. That these DFT issues also apply to protein beta sheets is shown with LC-wPBE [88], though these issues are not limited to that method.

Reduction of sigma/pi separation and increase in bond bending in the amide carbonyl group gives a reduction in calculated amide resonance. This necessarily means that RAHB of hydrogen-bonded chains of protein backbone amide groups is inaccurately calculated. RAHB is cooperative, and errors in calculated amide resonance may also be regarded as cooperative.

These findings have important implications for the selection of electronic structure methods for protein structure. Depending on method/basis set, any calculation of a range of internal amide electronic interactions may have a surprisingly large error. Results of calculation of cooperativity in backbone amide hydrogen bond, HB, chaining within a beta sheet using DFT will be significantly in error even when medium size basis sets are used, and in practice small basis sets have been used due to the secondary structure atom and basis function count. Hyperconjugative interactions with the backbone amide carbonyl group occur in other hydrogen bonded secondary structure types, so this warning applies to proteins generally. Also, we have observed but do not otherwise report on variability in the amino-acid residue-specific extent of loss of sigma/pi separation and increase in bond bending. We recommend that every DFT calculation of protein be accompanied by NBO assessment of the extent of amide carbonyl sigma/pi separation and bond bending.

These findings lead to the proposal that these molecular situations be regarded as standard benchmarks of electronic structure methods and basis sets, and be used for refining parameter values where a method takes parameters.

A second kind of error arising from double bonds having hyperconjugative interactions is also reported here. This second error applies to the amide group, and results in pyramidalization at the amide nitrogen. It occurs depending on basis set when using wavefunction methods. The result of this error may be seen in coordinate geometry, though NBO may be used to quantitate the extent of the shift of nitrogen from sp2 to sp3 hybridization. This error is similar to the benzene non-planarity failures arising from negative out-of-plane bending frequencies reported in [14]. This error also reduces amide resonance since the nitrogen sp3 hybridization is necessarily in competition with amide resonance. The above remarks about accuracy in calculations of RAHB also apply to this error. This second error is corrected by geometry optimization on the Counterpoise Correction [89] potential energy surface with fragment boundary defined at the C-N bond of ethanamide, but not at the C-C bond.

3.6 Introduction

The Natural Bond Orbital analysis procedure has properties that particularly recommend it for the analysis of electron density. Natural orbitals are natural in the sense of being the best possible account of correlated electron density [6], which suggests their use as the standard form for comparison of electron density. The localization of natural orbitals without loss of information as is done by NBO increases suitability for this purpose. NBO is usually directed to analysis of bonding similarity across molecular contexts, but here we use NBO to analyse the bonding differences resulting from different methods applied in the same molecular context.

The present study focuses in particular on the Natural Hybrid Orbitals [6], NHOs, for these describe bond bending and sigma, pi and higher angular momentum character of bonds. There is no built-in bias or geometric preconception embedded in NHO description of bonds, and all aspects of hybridization are converged to an optimal description of electron density [6]. The NBO account of electron density can report reduction in sigma/pi separation and increase in bond bending, depending on the electron density. As mentioned above, differences in NBO-derived quantities are not arbitrary with respect to electron density.

Density Difference Representation [90] is not oriented to chemical bonding and so the chemical significance of density change is not available in that representation.

NBO is the only chemical bonding analysis used in the present study, and justification for this is left to a recent review [55] which is very briefly summarized in the present paragraph. NBO preserves Pauli exclusion and Hermitian energetics and is thus free of the overlapping attributions arising from orbital or wavefunction non-orthogonality. NBO is based only on the *eigen* or intrinsic properties of the interactions of the electronic first-order reduced density matrix (1-RDM), rather than any assumptions concerning geometry or symmetry and is independent of the form of the original wavefunction. This basis in the *eigen* properties of the 1-RDM yields consistency and predictive capacity.

We used NBO to quantify the extent of sigma/pi separation and bond bending in double bonds having hyperconjugative interactions in vinylamine, ethanamide and an antiparallel beta sheet. In vinylamine and ethanamide, torsion is constrained across the single bond central to hyperconjugation during geometry optimization. In beta sheets, no geometry optimization constraints are necessary to maintain the torsion for non-planar hyperconjugation. The surveys reported here use constrained ethanamide as a model for protein backbone amide groups, and the relevance of this model is supported by NBO analysis of fully geometry optimized polyalanine and polyvaline protein secondary structures.

Backbone amide resonance is central to RAHB in protein backbone amide hydrogen bond chaining of certain protein secondary structures including beta sheets. Accurate calculation of RAHB cooperativity is important in calculating protein electronic and geometric structure.

3.7 Methods

Methods used in experiments are as implemented by Gaussian 09 D.01 [75], Orca 3.0.3 [72-74] and TeraChem 1.5K [91-94].

A pre-release version of NBO [95] was used for its XML [59] output option. The XML was queried with XQuery 3.0 [65] or XSLT 3.0 [63] as implemented by Saxon-PE 9.6.0.4 [69], and the results imported into Excel 2013 [96].

Jmol 14.2.2_2014.06.29 [97] was used for visualization of orbitals.

Except where otherwise stated, the default integration grid of the respective quantum chemistry package is used. Except where otherwise stated, the SCF convergence default was used for Gaussian, and VeryTightSCF was specified for Orca. The differing default SCF procedures were used for Gaussian and Orca. Where results differ between packages, the possibility that these differences are due to differences in default parameters is explored. In the case of the second problem mentioned in the overview, the results of Gaussian and Orca differ, and non-default values for integration grids, SCF convergence values and SCF procedures are explored.

3.8 Results and discussion

3.8.1 Vinylamine

As a preliminary note, for charts with many methods named on the horizontal axis, the first of which is Ap1:Figure 3.25, Gaussian's implementation of standalone functionals are grouped according to the classifications of pure, hybrid, range separated hybrid and double hybrid [98]. VSXC [99] to N12 [100] are pure functionals, B3LYP [101] to MN12-SX [102] are hybrid functionals, wB97 [103] to LC-wPBE are range separated hybrids and B2-PLYP [86] to mPW2-PLYPD [104] are double hybrids. The Gaussian implementation of MP2 and HF appear to the right of mPW2-PLYPD. Orca implementations of methods appear to the right of HF. MP2 appears twice, once as a Gaussian implementation and once as an Orca implementation.

The results appearing in [6] of loss of sigma/pi symmetry in vinylamine at 10 degree C-C-N-H Torsion with B3LYP/6-311++G** are in accord with our results (Figure 3.1 and Ap1:Figure 3.25). However, the ~20% loss of p character with B3LYP becomes less than 1% with MP2 (Ap1:Figure 3.23 and Ap1:Figure 3.25). Only with the smallest basis sets used, 6-31G** and def2-SVP, does the loss of p character with MP2 approach 1%, and with other basis sets is ~0.5%. Of the basis sets used, 6-311++G** gives the least loss of p character with MP2.

The C-C NBO morphologies for B3LYP/6-311++G** and MP2/61-311++G** can be seen in Figure 3.6 to Figure 3.13. These graphics are for 10 degree torsion, which we use for commonality with [6], though greatest loss of sigma/pi symmetry is found at 15 to 30 degrees hyperconjugative torsion depending on basis set except for aug-cc-pVTZ (Figure 3.1 and Ap1:Figure 3.23).

NHO deviation from line of centers, bond bending, also differs significantly between DFT methods and wavefunction methods. These deviations are charted as Figure 3.2, Ap1:Figure 3.24 and Ap1:Figure 3.26. For the C-O(pi) NHOs, 90 degrees is no deviation from expectation based on line of centers. For some basis sets with B3LYP the bond bending is as large as 45 degrees from the reference 90, contrasting with the 7 degree maximum bond bending with MP2.

Differences in the polarization of the C-O(sigma) and C-O(pi) NBOs is shown in Ap1:Figure 3.27 and Ap1:Figure 3.28. The variable performance of different DFT methods with aug-cc-pVTZ can be seen. Broadly, aug-cc-pVTZ is most associated with greatest difference between DFT methods, and between DFT methods and wavefunction methods. aug-cc-pVDZ offers markedly less DFT divergence, which is frequently observed in this work. As is usual in this work, there is greatly less divergence between properties calculated with correlated wavefunction methods than with DFT methods.

The B3LYP and MP2 results are dichotomous, and nomination of which is in such large error is most desirable. It is not a given at the outset that the error is with B3LYP, for DFT methods are in principle capable of incorporating static correlation and uncorrected MP2 is not. In answering this question we refer to the results of other DFT methods including double hybrid methods [86], corrected MP2 methods [105, 106], DLPNO-CCSD(T) [74] and the multireference MRCI+Q [107], and note that B3LYP loss of symmetry begins to collapse at very large basis sets starting at aug-cc-pV5Z (Figure 3.3 to Figure 3.5).

Ap1:Figure 3.25 to Ap1:Figure 3.28 show sigma/pi symmetry, bond bending and polarization at a range of DFT methods including hybrid methods, long-range corrected methods [108] and double hybrid methods, and some corrected MP2 methods, the invariably worst-performing method HF [34] and single point calculation at DLPNO-CCSD(T). The double hybrid [86] methods tested have only twice the small error of MP2, but computational cost scales similarly.

The poor performance of HF and single hybrid methods and better performance of double hybrid and MP2 variants yields suggests an underlying mechanism of these errors, that being different treatment of multicenter exchange delocalization. DFT single hybrid and GGA exchange does not have multicenter delocalization, HF's multicenter exchange delocalization is uncorrected, MP2's correlation largely corrects HF's excessive multicenter delocalization [109] and double hybrid DFT's correction by correlation is not quite as good as MP2's.

We anticipate that with optimization at CCSD [110] or above, the loss of p character will be further reduced from the already small MP2 figure. Where DLPNO-CCSD(T) has been used in this work, it has necessarily been used for single point energy calculation, with geometry optimization at MP2 or SCS-MP2 and aug-cc-pVTZ, aug-cc-pVQZ or def2-QZVPP.

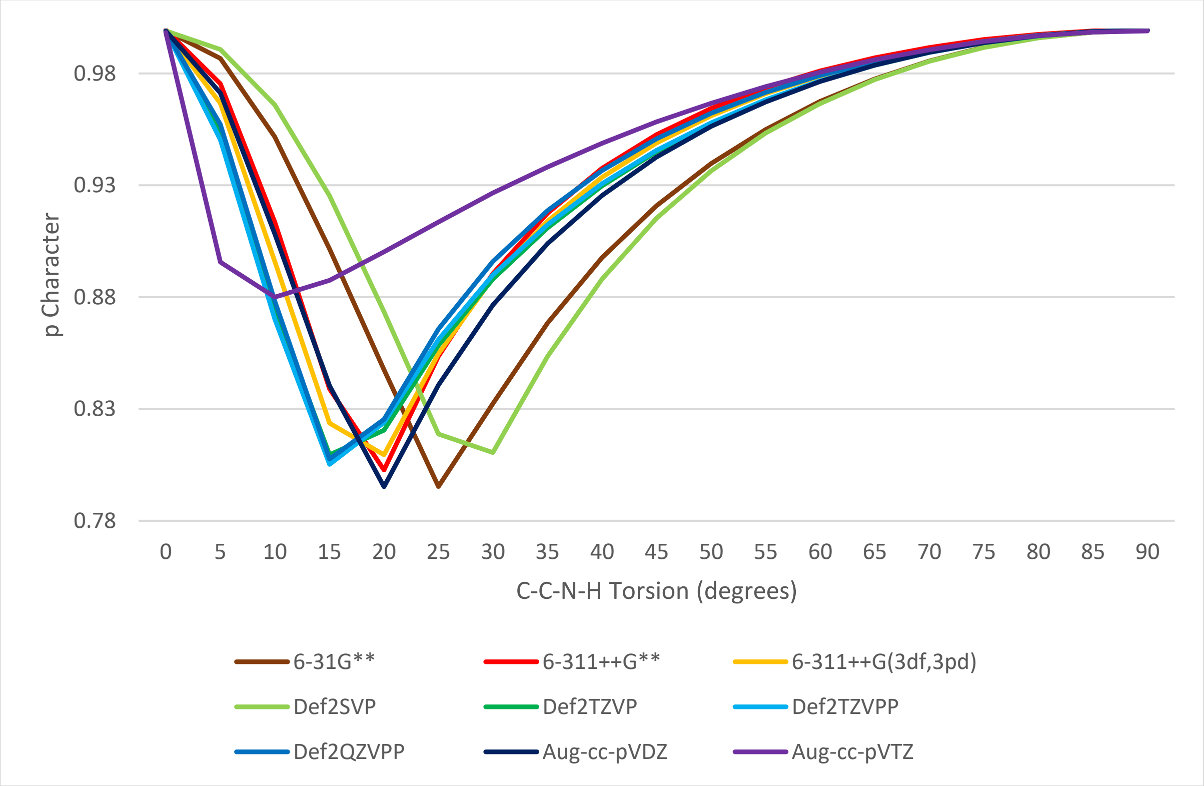


Figure 3.1. C-C(pi) Central Carbon NHO p Character in Vinylamine at Varying C-C-N-H Torsion and Basis Set with B3LYP

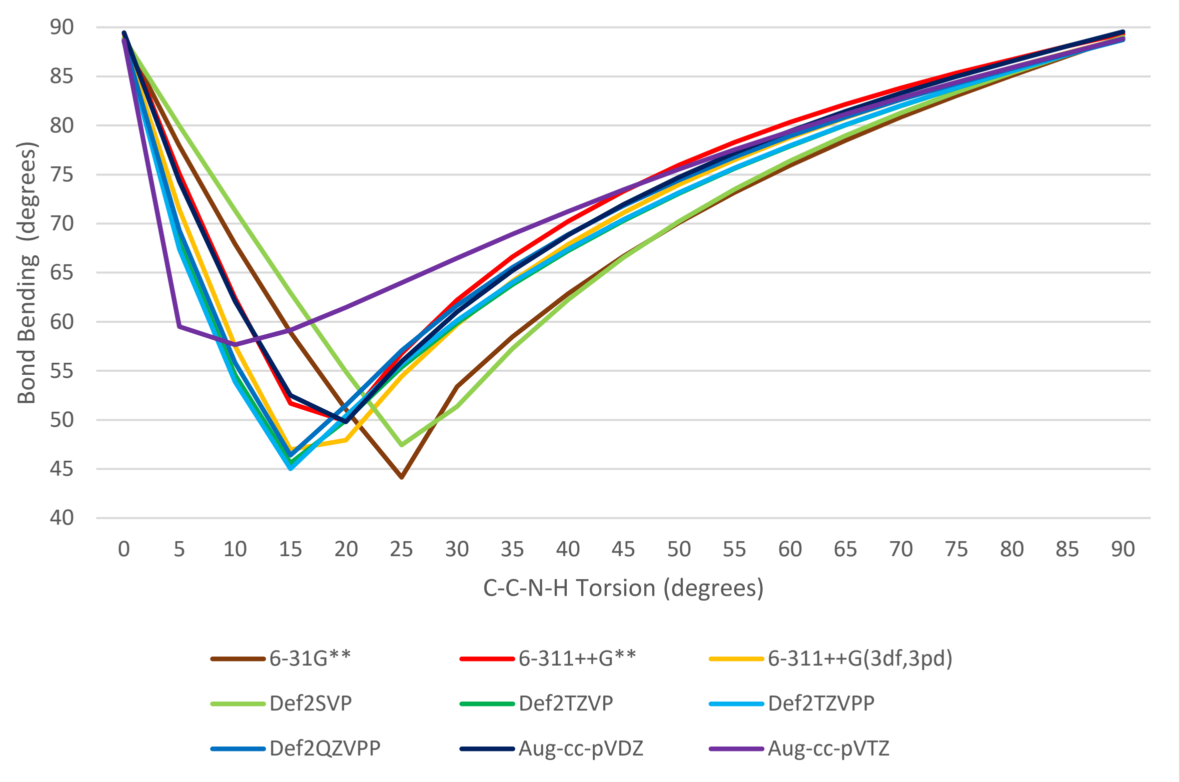


Figure 3.2. C-C(pi) Central Carbon NHO Bond Bending in Vinylamine at Varying C-C-N-H Torsion and Basis Set with B3LYP

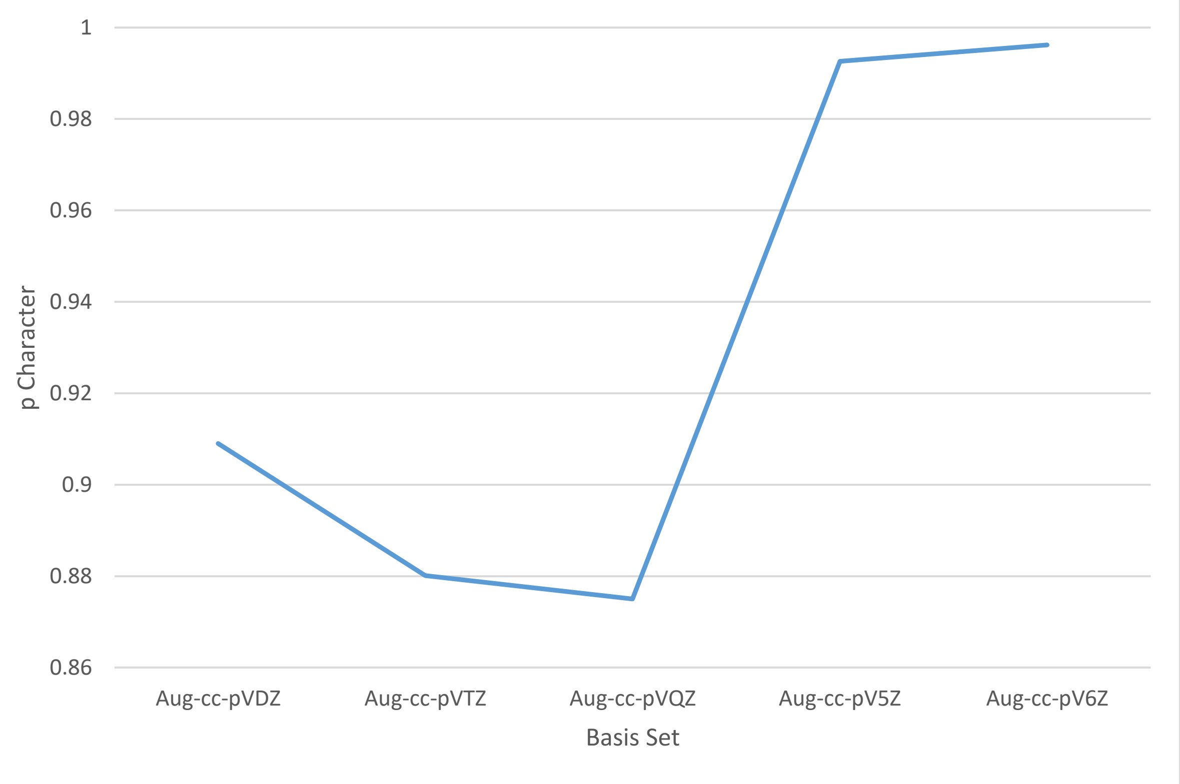


Figure 3.3. C-C(pi) Central Carbon NHO p Character in Vinylamine at 10 Degree C-C-N-H Torsion with B3LYP and Varying Basis Set

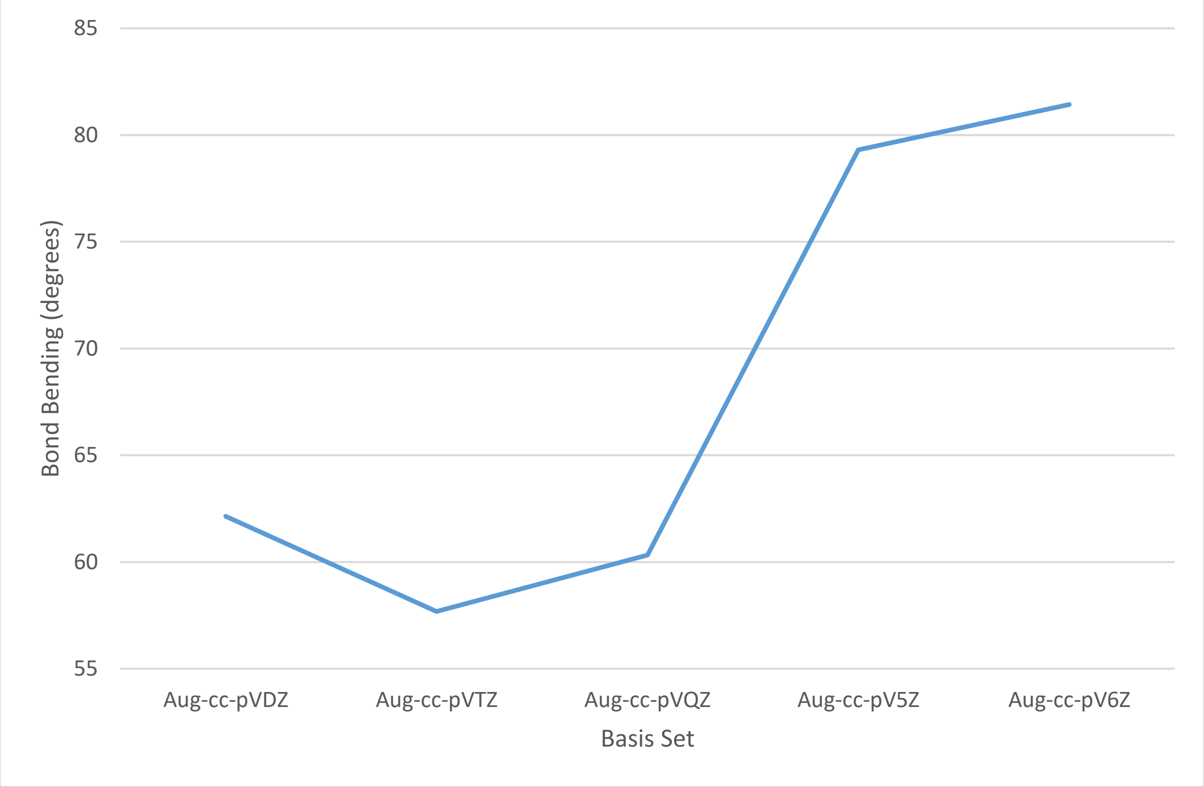


Figure 3.4. C-C(pi) Central Carbon NHO Bond Bending in Vinylamine at 10 Degree C-C-N-H Torsion with B3LYP and Varying Basis Set

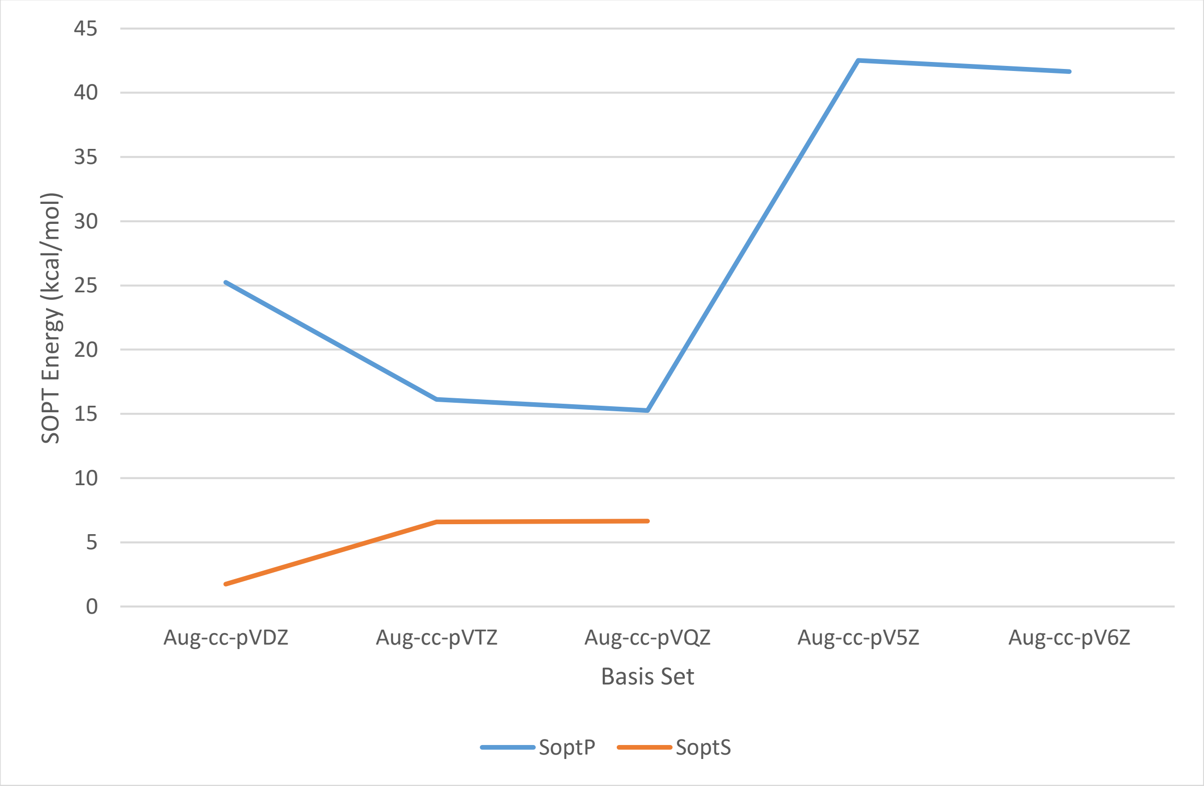


Figure 3.5. N(lp)->C-C(pi then sigma)* SOPT Energy in Vinylamine at 10 Degree Torsion with B3LYP and Varying Basis Set

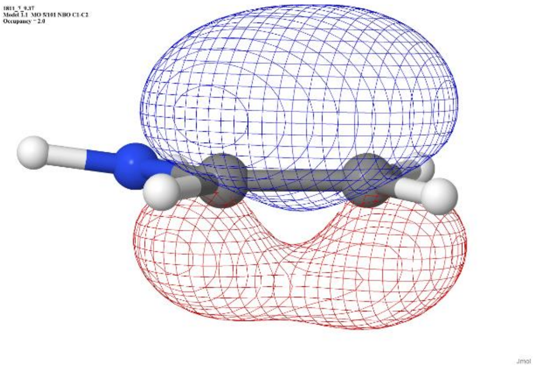


Figure 3.6. B3LYP/6-311++G** C-C(pi) NBO in Vinylamine at 10 Degree C-C-N-H Torsion

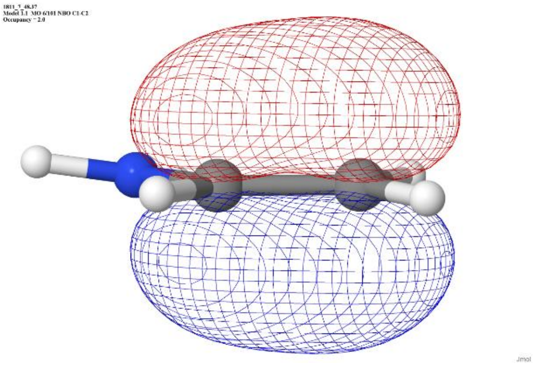


Figure 3.7. MP2/6-311++G** C-C(pi) NBO in Vinylamine at 10 Degree C-C-N-H Torsion

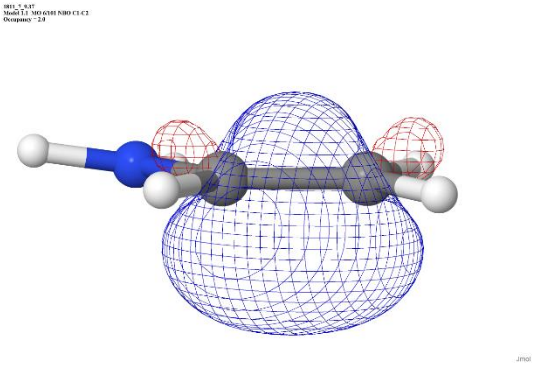


Figure 3.8. B3LYP/6-311++G** C-C(sigma) NBO in Vinylamine at 10 Degree C-C-N-H Torsion

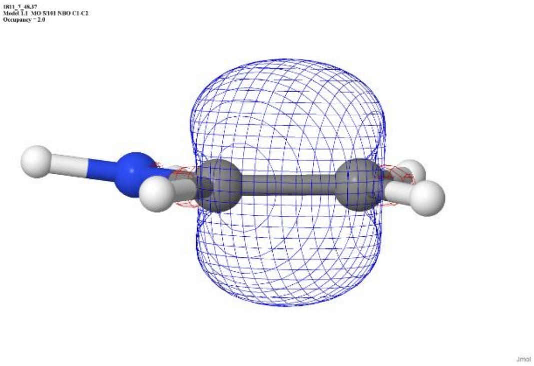


Figure 3.9. MP2/6-311++G** C-C(sigma) NBO in Vinylamine at 10 Degree C-C-N-H Torsion

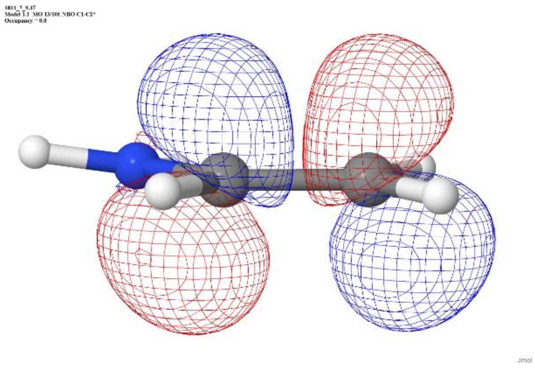


Figure 3.10. B3LYP/6-311++G** C-C(pi)* NBO in Vinylamine at 10 Degree C-C-N-H Torsion

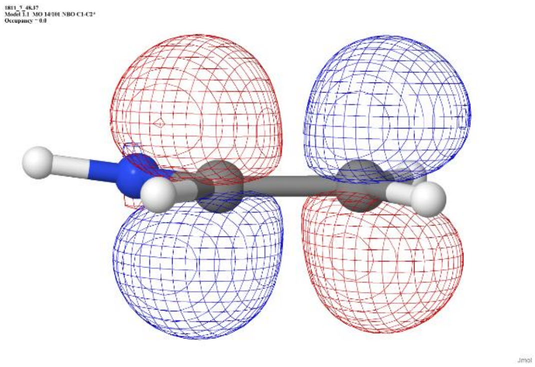


Figure 3.11. MP2/6-311++G** C-C(pi)* NBO in Vinylamine at 10 Degree C-C-N-H Torsion

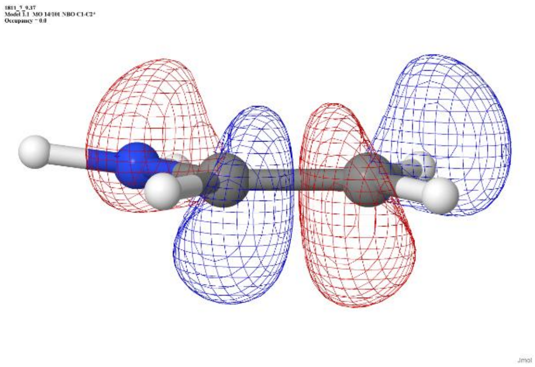


Figure 3.12. B3LYP/6-311++G** C-C(sigma)* NBO in Vinylamine at 10 Degree Torsion

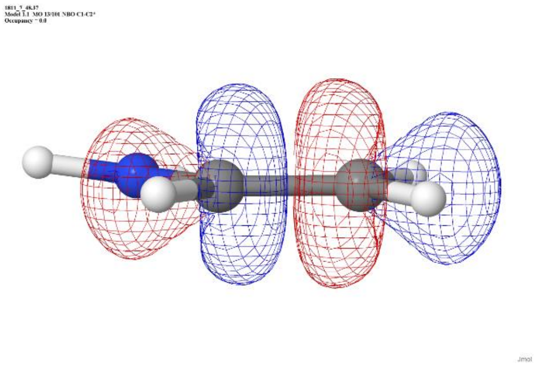


Figure 3.13. MP2/6-311++G** C-C(sigma)* NBO in Vinylamine at 10 Degree Torsion

3.8.2 Ethanamide

3.8.2.1 Carbonyl orbital disturbances

Ethanamide is the simplest molecule that can demonstrate the hyperconjugative interactions with the amide carbonyl group that may be seen in peptide backbones. In parallel beta strands, the O-C-CA-HA torsion may be taken as 173 degrees and in antiparallel beta strands as -165 degrees [111], being 7 and 15 degrees respectively from being perfectly antiperiplanar. We denote the ethanamide non-amide carbon as CA for commonality with peptides. Given the equivalence of the hydrogens connected to CA in ethanamide, rather than refer to them as HA1, HA2 and HA3, we refer to them generically as H, referring to them by O-C-CA-H dihedral angle.

Figure 3.14 and Ap1:Figure 3.29 contrast the difference in loss of sigma/pi symmetry of the carbonyl orbitals when B3LYP or MP2 is used. The y axis scales differ markedly, for the B3LYP loss of p character is ~5% for basis sets 6-31G** and def2-SVP at 175 degrees O-C-CA-H torsion. Of secondary significance, with basis sets def2-TZVP, def2-TZVPP and aug-cc-pVTZ the loss of p character approaches 1.5% at 165 degrees torsion. The data shows that a geometry away from perfectly antiperiplanar is necessary for hyperconjugation to disturb sigma/pi symmetry, and later figures bear out disturbances of other properties also require a geometry away from perfectly antiperiplanar. The 180 to 60 degree sweep of O-C-CA-H dihedrals does not quite complete the cycle of values, for only one H in CA-H is constrained, an arrangement that was selected to mimic peptide backbones in which only one of the substituents of CA is torsionally constrained. The unconstrained H that is moving into antiperiplanar dihedral angle does not remain at precisely 120 degree relative to the constrained H throughout the rotation.

Hyperconjugation between the sigma carbonyl orbitals and CA-H is centered at O-C-CA-H torsion of 180 degrees, and hyperconjugation between the pi carbonyl orbitals is centered at O-C-CA-H torsion of 90 degrees. Disturbances of properties at torsional angles associated with pi carbonyl hyperconjugation can be seen Ap1:Figure 3.31, Ap1:Figure 3.32 and Ap1:Figure 3.35.

Figure 3.16 shows a reduction in Second-Order Perturbative Analysis, SOPT, of donor-acceptor interactions [6] for the primary amide resonance delocalization of ~61 kcal/mol to ~41 kcal/mol in 5 degrees of torsion from antiperiplanar with B3LYP and basis set 6-31G**, and also large reduction with def2-SVP. Ap1:Figure 3.22 shows the N(lp) delocalization into C-O(s)* rather than C-O(p)*. However, with 6-311++G**, the variation in amide resonance delocalization throughout the range of torsion is only ~2 kcal/mol. This is a very large variation between basis sets.

Ap1:Figure 3.19 shows that the steric exchange energy of C-O(s)|C-O(p) increases from 0 to ~25 kcal/mol with 6-31G** in 5 degrees of O-C-CA-H torsion from antiperiplanar. Ap1:Figure 3.20 shows that in the same torsional range O(lp-s)|C-O(p) increases from 0 to ~7.5 kcal/mol with 6-31G**.

Ap1:Figure 3.21 shows that in this torsional range O(lp-s)|C-O(s) decreases by ~6 kcal/mol with 6-31G**.

N(lp)->C-O(p)* SOPT Energy is shown with varying torsional hyperconjugation for a selection of Minnesota density functional methods [112] in Ap1:Figure 3.37 and other DFT methods from the Truhlar group in Ap1:Figure 3.38. MN12-L [113], MN12-SX and M06-HF [114] are less afflicted of error at torsional hyperconjugation than other DFT methods tested here. Each of these three methods is tested with a range of basis sets, and results are shown in Ap1:Figure 3.58, Ap1:Figure 3.56 and Ap1:Figure 3.57 respectively. Of the three methods, MN12-L is the best performer, which is convenient from the point of view of protein calculations since it is a local and hence better scaling method [115]. MN12-SX is a range separated hybrid and M06-HF is a global hybrid [115], so local, range separated versus global nature of methods is not key to alleviation of the error at torsional hyperconjugation. They are all meta-GGA methods, with MN12-L being at only the third rung of Jacob's ladder [116, 117] and having zero percentage exact exchange.

While it is variation in rather than absolute value of N(lp)->C-O(p)* SOPT energy due to torsional hyperconjugation that is the focus of interest here, it must be remarked that the difference in this quantity between the methods at perfectly antiperiplanar geometry is the next concern in assessing DFT methods applied to proteins. At the highest quality basis set used, def2-QZVPP, the difference in N(lp)->C-O(p)* SOPT Energy between MN12-L and M06-HF is greater than 40 kcal/mol. This difference in calculating amide resonance will have large consequences for calculated RAHB. While MN12-L/def2-QZVPP or linear scaling variant thereof might be recommended for application to proteins on the basis of appropriate relative amide resonance response to torsional hyperconjugation, error in calculating the absolute value of amide resonance is a major obstacle. The occupancies of N(lp) and C-O(p)* in ethanamide calculated with wavefunction methods and DFT methods may be compared to gauge amide resonance. N(lp) occupancy, calculated at high quality basis sets, can be seen in Ap1:Figure 3.50, Ap1:Figure 3.51 and Ap1:Figure 3.52 and C-O(p)* occupancy can be seen in Ap1:Figure 3.53, Ap1:Figure 3.54 and Ap1:Figure 3.55. C-O(p)* occupancies are greater with DLPNO-CCSD(T) than with any of the DFT methods shown, with MN12-L having the largest C-O(p)* occupancies of the three favoured Minnesota density functionals. CASSCF(8,7)/MRCI+Q/def2-TZVPP in turn gives greater C-O(p)* occupancy than DLPNO-CCSD(T). While the second configuration weights of CASSCF/MRCI+Q are less than 0.02 (Table 3.2) indicating the calculation is not particularly multireference, it seems that improvement on even CCSD(T), the gold standard of single reference quantum chemistry, is needed for calculation of amide resonance. Multireference methods that scale to proteins may be necessary for calculation for protein RAHB.

The influence of the D2 [118], D3 [93] and D3BJ [94] empirical dispersion corrections on the torsional hyperconjugative error reported here is shown where available for selected methods and the 6-31G** basis set in Ap1:Figure 3.42, Ap1:Figure 3.43 and Ap1:Figure 3.44 respectively. None of the empirical dispersion schemes tested alleviates the torsional error, though D3BJ alone can be recommended as not making the error any worse.

The influence of the omega parameter of LC-wPBE with 6-31G** on the torsional hyperconjugative error reported here is shown in Ap1:Figure 3.46. It can be seen that no listed value gives a result usable for proteins, and values less than 0.4 bohr^{-1} give the worst results. Gaussian 09 D.01 has no label for LC-wPBEh [119], so the equivalent LC-PBEhPBE is used. The influence of the HF exchange coefficient on the error is shown for selected values of omega in Ap1:Figure 3.46, Ap1:Figure 3.47 and Ap1:Figure 3.48. At omega=0.6 bohr^{-1}, small exchange coefficients are least disfavoured, and at omega=0.4 bohr^{-1} and omega=0.2 bohr^{-1}, large exchange coefficients are least disfavoured. It is likely that there exist no values for omega and the exchange coefficient that yield a method useful for proteins.

To explore whether the hyperconjugative error is due more to the exchange functional or to the correlation functional of pure DFT methods, all combinations of exchange functionals and correlation functionals listed in the Gaussian 09 User's Manual [98] tested for this error with 6-31G**. Results are shown for all listed exchange functionals with the PBE correlation functional [120] (Ap1:Figure 3.60), the VWN correlation functional [121] (Ap1:Figure 3.61) and the LYP correlation functional [122] (Ap1:Figure 3.62). With the exception when in combination with the exchange functionals which are quite erratic during torsional change, the choice of correlation functional is relatively unimportant in this test. The G96 exchange functional [123] has the shallowest N(lp)->C-O(p)* SOPT energy wells through O-C-CA-H torsion with all the listed correlation functionals. Results are shown for all listed correlation functionals with the XA exchange functional [124-126] (Ap1:Figure 3.63), the BRx exchange functional [127] (Figure 3.64) and the Slater exchange functional [124-126] (Ap1:Figure 3.65), and in these figures the curves are usually so little different that they partially overlay, again indicating that the choice of correlation functional is secondary in these tests. This is further shown in Ap1:Figure 3.59 by the use of the three listed exchange functionals that can standalone compared with results when combined with the VWN correlation functional. The torsional hyperconjugation error principally arises of errors in exchange functionals.

The variation in C-O(s)* and C-O(p)* energy levels for selected Minnesota density functionals with 6-31G** during change of O-C-CA-H torsion is shown in Ap1:Figure 3.39 and Ap1:Figure 3.40 respectively. The torsional hyperconjugation error is associated with the energy levels of C-O(s)* and C-O(p)* becoming closer. Decrease in the C-O(s)* energy level makes the orbital a more inviting charge transfer acceptor. The torsional hyperconjugative error is likely due to re-hybridization unduly favouring hyperconjugative charge transfer or sigma acceptor orbitals. The C-O(p)* energy level variation for MN12-L for selected basis sets is shown in Ap1:Figure 3.41. 6-311++G** behaves erratically, though 6-311++G(3df,3pd) performs well as determined by the basis set we suggest be taken as canonical, def2-QZVPP. There are two efficiency merits of MN12-L/def2-QZVPP, being that MN12-L is local and that def2-QZVPP does not have diffuse functions. The accuracy of MN15-L [128] in calculating amide resonance will be investigated in future, though the accuracy demands of calculating resonance in RAHB where errors are cooperative are likely forbidding.

The influence of basis set diffuse functions on the torsional hyperconjugation error is explored for B3LYP (Ap1:Figure 3.66) and MN12-L (Ap1:Figure 3.67). The introduction of diffuse functions to 6-31G** to form 6-31++G** is very beneficial. The torsional hyperconjugation error is very large for cc-pVDZ which has no diffuse functions. The calendar variations of correlation consistent basis sets are shown, though the lines are not distinct for apr-cc-pVDZ, may-cc-pVDZ and jun-cc-pVDZ for B3LYP and apr-cc-pVDZ and may-cc-pVDZ for MN12-L. The aug-cc-pVDZ curve has the deepest wells of any of the calendar variations of cc-pVDZ, and the removal of the diffuse functions on hydrogen to form jul-cc-pVDZ somewhat flattens the curve. Further removal of diffuse functions to form the other calendar correlation consistent basis sets makes little difference. A wider selection of Pople basis sets is shown in Ap1:Figure 3.68, and the introduction of polarization functions is associated with a worsening of the torsional hyperconjugation error. A wider selection of correlation consistent basis sets is shown in Ap1:Figure 3.69, and it can be seen that very high quality correlation consistent basis sets (aug-cc-pVQZ) give energy wells during O-C-CA-H for primary amide SOPT that are significantly deeper than the 2.7 kcal/mol given by MN12-L/def2-QZVPP. The good performance of MN12-L/def2-QZVPP (though absolute amide resonance irrespective of O-C-CA-H torsion must also be considered) on ethanamide and the poor performance of B3LYP/6-311++G** on vinylamine indicate that diffuse functions are not in general necessary or sufficient to accurately model torsional hyperconjugation in conjugative systems. Any method/basis combination for proteins must be benchmarked on the torsional hyperconjugative problem as described here. Further, the introduction of diffuse functions markedly increases the computational cost of electronic structure methods and increases convergence difficulties, most often prohibiting their use on even the smallest proteins.

3.8.2.2 Pyramidalization at nitrogen

Figure 3.17 shows how erratic 6-311++G** is with MP2. This accords with the modern practice of using correlation consistent rather than Pople basis sets [129] with wavefunction methods. MP2/6-311++G** is greatly disturbed by correlation associated with each of sigma and pi carbonyl hyperconjugation. Figure 3.18 shows the loss of O-C-N-H planarity associated with this disturbance. During the rotation, planarity is lost by 15 degrees in both directions from the amide plane. Correlation associated with hyperconjugation makes sp3 pyramidalization of N more favourable than planarity, weakening the amide resonance. At 95 degrees torsion, just 5 degrees from maximum pi hyperconjugation, the dihedral describing the non-planarity changes sign, breaking the symmetry centered at 120 degrees O-C-CA-H torsion. Non-planarity is not limited to Pople basis sets. With def2-TZVP, the non-planarity is quite significant at ~7 degrees, with def2-QZVPP at ~5 and aug-cc-pVTZ at ~3 degrees. At the hyperconjugative torsions at which amide non-planarity occurs, calculated amide resonance is impaired and necessarily so is the RAHB in which the amide might participate and hence cooperativity in hydrogen bonded secondary structures. The resonance of only one amide group in a hydrogen bonded chain need be erroneously calculated for RAHB throughout the chain to be in error. Since these errors all result in reduction of amide resonance, there are no compensating errors. Errors in the calculation of the resonance of multiple groups cooperate in producing error in calculation of the hydrogen bonding of the chain. Due to this cooperativity of errors, it is quite important that amide resonance be accurately calculated. def2-QZVPP and aug-cc-pVTZ are of reasonable size and repute, and this pyramidalization calls into question the use of MP2 for geometry optimization of peptides/proteins. This test is suitable for benchmarking during and after development of MP2 variants. Perhaps CCSD does not fail in this manner, but optimization with CCSD with large basis sets is not yet viable for extensive studies.

The findings of pyramidalization at amide N with MP2 are similar to a report of non-planarity of benzene [14]. Basis sets passing the benzene planarity test by having positive vibrational frequencies at planarity, such as aug-cc-pVTZ, do not fare altogether well in the geometry optimization tests used here although the correlation consistent basis sets are constructed to provide basis set incompleteness error, BSIE, balance. The report attributes the failures to elevated sigma-pi correlation. The report gives that one-electron theories such as DFT are immune to this two-electron BSIE. The errors with DFT described in the present work are then of different origin, but will still be a result of incorrect emphasis on certain correlations. The report points out that Atomic Natural Orbital [130, 131], ANO, basis sets are better again than correlation consistent basis sets, since ANO minimizes basis set superposition errors. We have not used ANO basis sets due to their computational expense. This pyramidalization at MP2/6-311++G** is correctable by geometry optimization on the Counterpoise Correction potential energy surface with fragment boundaries defined at the C-N bond of ethanamide, but not at the C-C bond (Ap2:Table 3.3).

Ap1:Figure 3.72 shows that the SCS correction to MP2 slightly attenuates resonance-type delocalization from the nitrogen lone pair NBO, but not in conjunction with Orbital Optimization. The result of DLPNO-CCSD(T) single point calculation over geometry optimized with MP2/aug-cc-pVQZ is in keeping with the geometry optimized SCS-corrected but not orbital-optimized results. This suggests suitability of SCS-MP2 optimization for DLPNO-CCSD(T) single point calculation for such studies. Since the SCS-MP2 and RI-SCS-MP2 results do not differ appreciably, RI-SCS-MP2 is useful for its reduced run-times. The DFT methods that produce results closest to those of SCS-MP2 are VSXC, tHCTH [132], B97D3 [94, 118] and N12. The furthest are by M11-L [133], M06 [134], wB97 and LC-wPBE, and these will underestimate amide resonance. As always, HF is far outlying and greatly underestimates amide resonance. Further light is cast on the irregular M11-L results seen in figures Ap1:Figure 3.70, Ap1:Figure 3.71, Ap1:Figure 3.73 and Ap1:Figure 3.74 by the O-C-N-H dihedral values seen in Ap1:Figure 3.75 in which variation with basis set is ~20 degrees. M11-L is subject to the pyramidalization at nitrogen seen with MP2. It can be seen in Figure 3.37 and Ap1:Figure 3.38 how low the N(lp)->C-O(p)* SOPT energy of M11-L, M06L, M06, N12 and SOGGA11 become with the noted standardized geometry. Since pyramidalization at amide N and N(lp)->C-O(p)* are necessarily in competition, it is to be expected that low N(lp)->C-O(p)* is associated with pyramidalization at N. It is anticipated that this is the connection between carbonyl orbital disturbances and pyramidalization at N, and how DFT and wavefunction methods can both manifest pyramidalization at N. The ~13 degree pyramidalization seen with Gaussian at MP2/6-311++G** is not seen with Orca even when the NoFrozenCore option is used for greater commonality with Gaussian defaults. The Gaussian option Stable=Opt reveals that the wavefunctions for both M11-L and MP2 are stable at the optimized geometry. Gaussian Stable=Opt also deems the wavefunction associated with the Orca-produced geometry to be stable. The vibrational frequencies check for geometric stability is not applicable since constraints are used. When Gaussian optimization proceeds from the Orca-optimized geometry, the highly pyramidalized geometry results. When Orca optimization proceeds from the Gaussian-optimized geometry, the highly pyramidalized geometry remains. The Orca stationary point is evidently not seen by Gaussian, though the Gaussian stationary point is seen by Orca. Developers and users of MP2 variant methods need to be aware of this variation in optimized geometry, for from it follows significant consequences for amide resonance and flexibility of protein backbones. A pyramidalization of ~13 degrees might be converted to such a pyramidalization in the other direction by backbone strain, giving a ~26 degree flexibility in each protein backbone omega (CA-C-N-CA') torsion. CASSCF [135] (8,7)/MRCI+Q[107] with def2-TZVPP and auxiliary def2-TZVPP/C over the MP2/6-311++G** geometry gives 0.8372 weight on a single configuration in ground state with next configuration weight in ground state being less than 0.02, so this pyramidalized geometry is only weakly associated with multireference character. Ap2:Table 3.1 gives the results of this multireference method for the non-pyramidalized geometries arising from beta strand psi torsions with MP2/aug-cc-pVQZ, showing good agreement with single reference wavefunction methods, and that the problem is not particularly multireference.

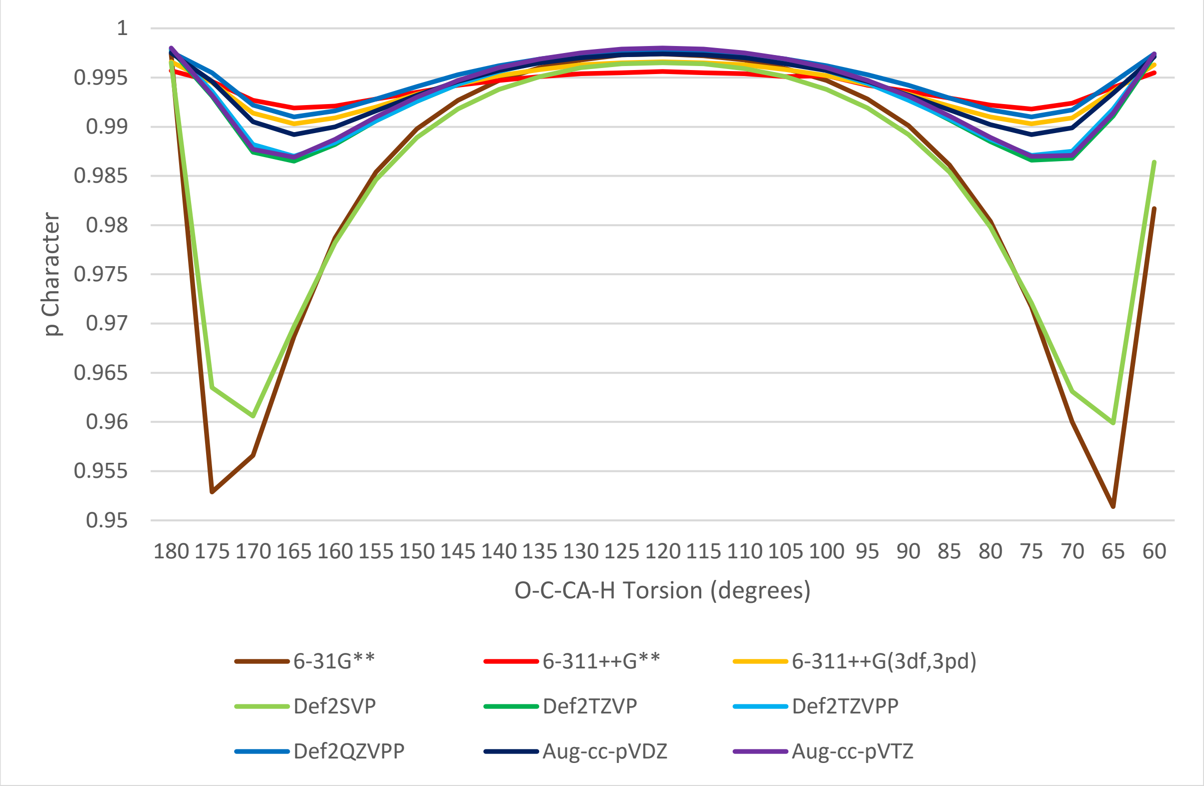


Figure 3.14. C-O(pi) Carbon NHO p Character in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at B3LYP

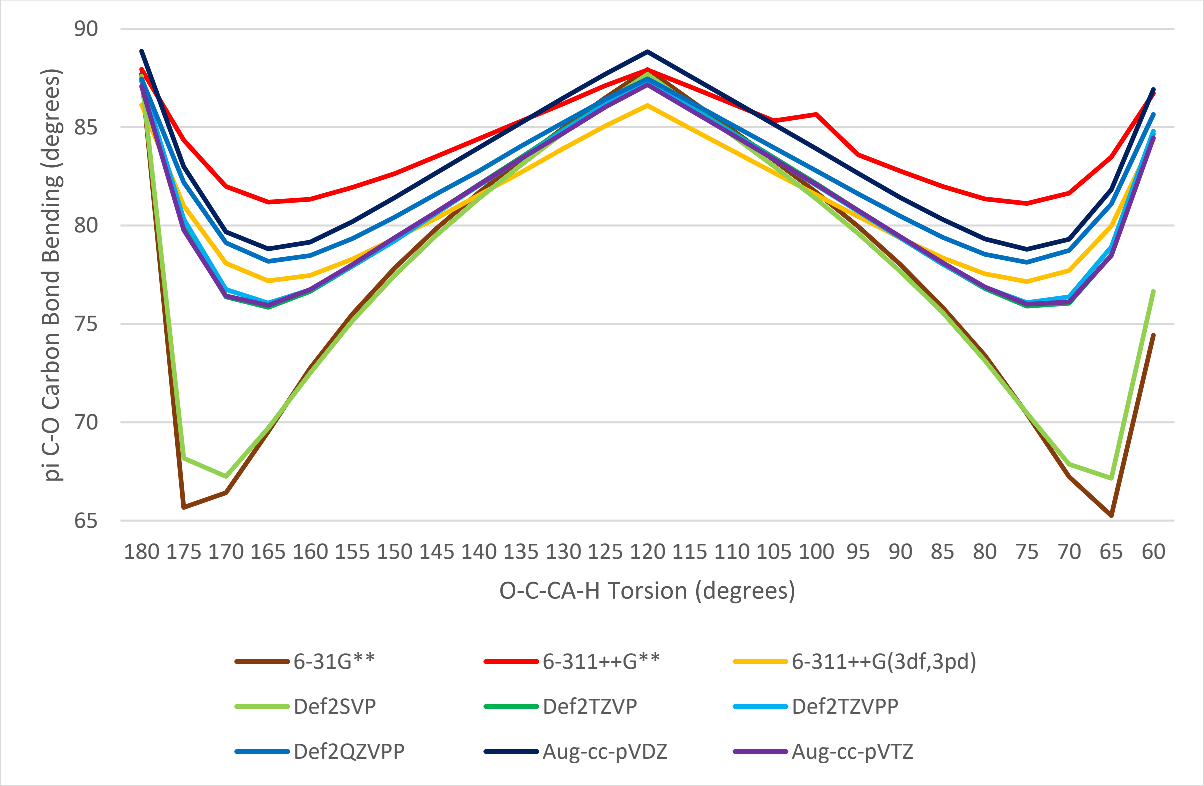


Figure 3.15. C-O(pi) Carbon NHO Bond Bending in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at B3LYP

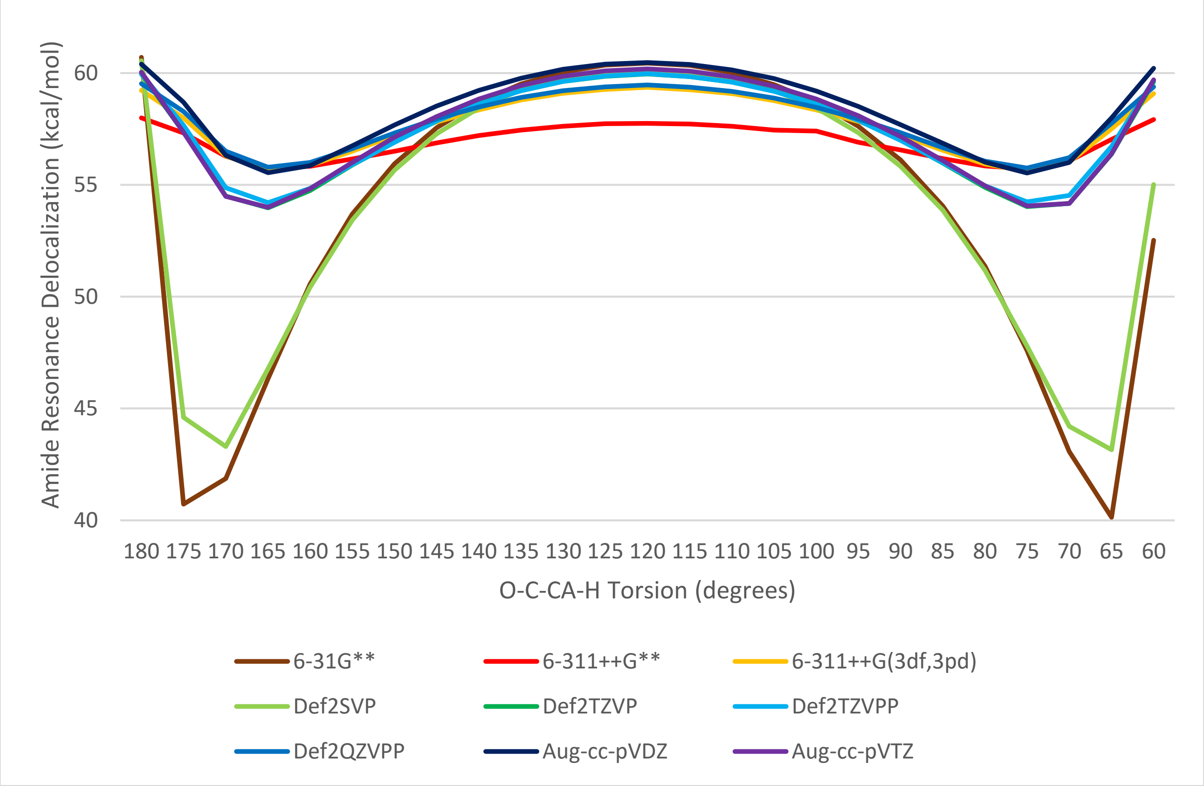


Figure 3.16. Second-Order Perturbation Theory Primary Amide Resonance Delocalization (N(lp)->C-O(pi)*) in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at B3LYP

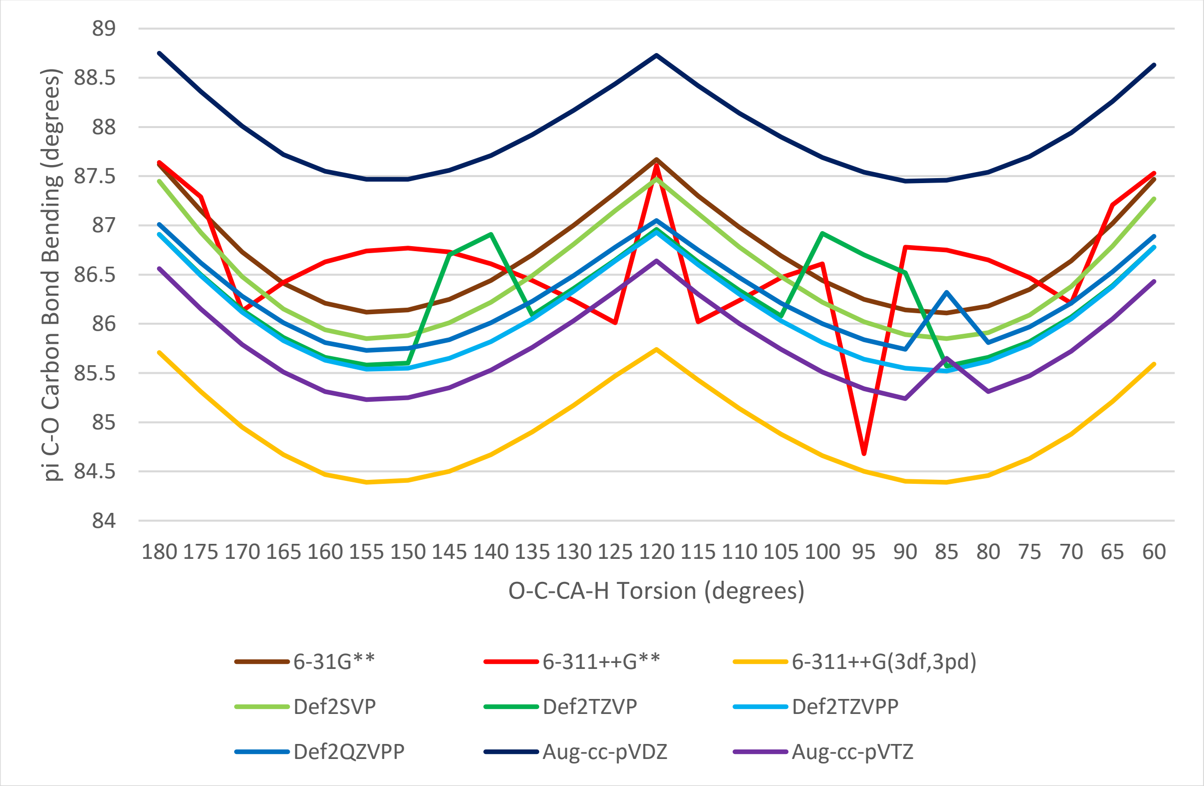


Figure 3.17. C-O(pi) Carbon NHO Bond Bending in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at MP2

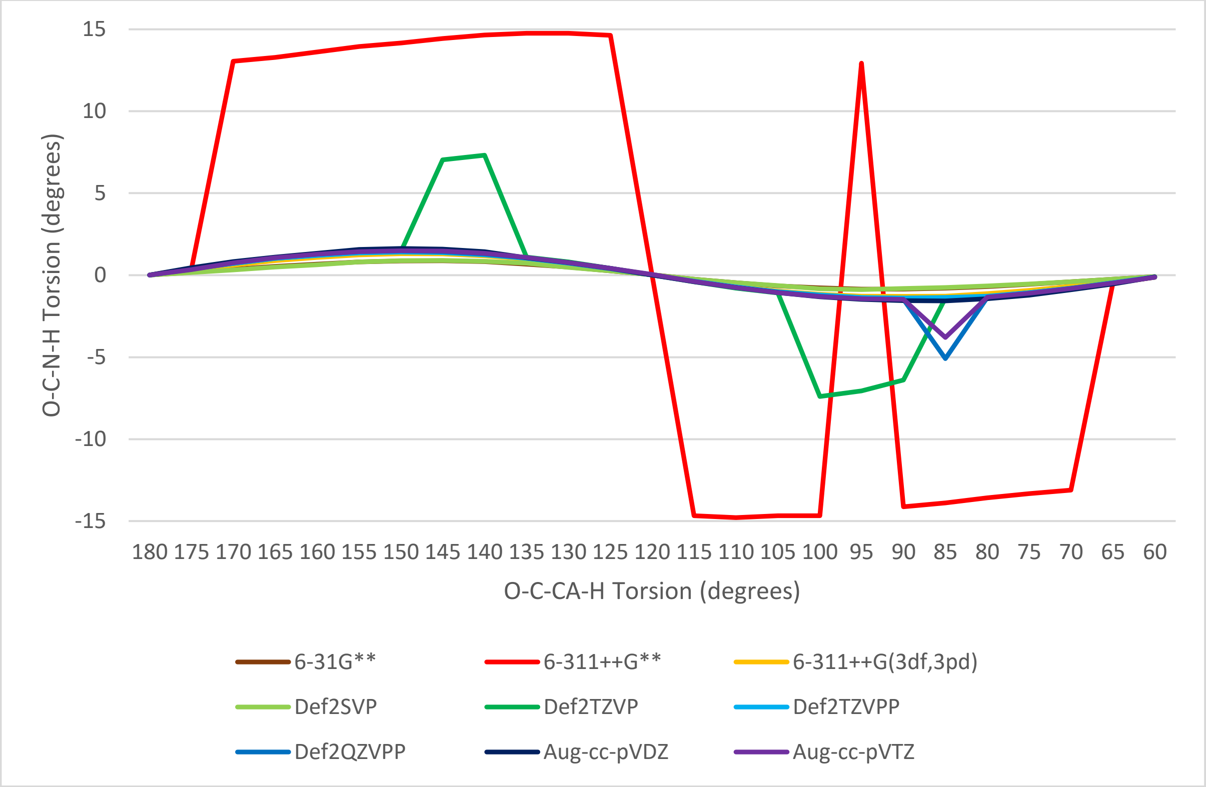


Figure 3.18. O-C-N-H Torsion in Ethanamide with Varying O-C-CA-H Torsion and Basis Set at MP2 FineGrid

3.8.3 Antiparallel beta sheet

Ap2:Table 3.4 and Ap2:Table 3.5 refer to gas phase antiparallel beta sheets which have been geometry optimized with TeraChem 1.5K with LC-wPBE (keyword wpbe in TeraChem), omega=0.4 bohr^{-1} and 6-31G**. While the ethanamide model requires optimization constraints to demonstrate torsional hyperconjugation, these beta sheets require no constraints, that is, they are fully optimized. Ap2:Table 3.4 refers to a polyvaline structure and Ap2:Table 3.5 to a polyalanine structure.

By extending ethanamide to species resembling alanine and valine, it can be shown that these backbone errors are very sensitive to sidechain even for uncharged and non-polar residues, so that the utility of methods afflicted of these errors is dubious in the case of heterogeneous amino acid residues. Also, it might be erroneously concluded that this modulation of backbone amide resonance by sidechain is a means by which sidechain determines protein fold.

The variation in the N(lp)->C-O(pi)* SOPT (“SoptP” column) has two sources, one being RAHB which tends to cause amide resonance in hydrogen bonded amide chain to peak in the middle of the chain and be lowest at the ends, and the effects being discussed here, bond bending and loss of local symmetry. The tables mentioned above make the point that this bond bending and loss of local symmetry exists in beta sheets rather than only in the ethanamide model.

These tables are sorted on bond bending ("Dev" column). While bond bending might be expected to be dependent of d orbital involvement to give non-cylindrically symmetric orbitals so the hybrids can point at each other, these tables, particularly Ap2:Table 3.4, do not strongly bear out this notion.

SOPT kcal/mol values under 1.0 do not appear in these tables. Noteworthy are the variations in N(lp)->C-O(p)* SOPT ("SoptP") and N(lp)->C-O(s)* SOPT ("SoptS"). That the latter is non-zero is evidence of carbonyl orbital disturbance.

There is a strong but imperfect association between the occupancy of C-O(s)* and bond bending and loss of carbonyl sigma/pi symmetry. Perhaps non-double hybrid DFT methods do not deal well with the situation in which there is occupancy of C-O(s)* in the context of large occupancy of C-O(p)*.

In Ap2:Table 3.5, the least bond bending is associated with O-C-CA-HA dihedral of ~150 degrees, which corresponds to the optimal angle (90 degrees) for hyperconjugation involving C-O(pi)* orbitals. This suggests that this hyperconjugation is not disruptive as hyperconjugation involving C-O(s)*.

Moderate SOPT kcal/mol values are proportional to resonance-type charge transfer [6]. The difference between hyperconjugative charge transfer from the methyl group into C-O(p)* and C-O(s)* has some association with bond bending and loss of sigma/pi symmetry. The variation in individual and collective hyperconjugative SOPT values is less than that of amide resonance, suggestive of a failure of method.

3.9 Conclusion

In contrast to single and multireference wavefunction methods, in the molecules studied, established DFT methods do not substantially maintain sigma/pi separation or bond straightness in double bonds that have hyperconjugative interactions at other than perfectly antiperiplanar geometry. Unlike ethene with C-C torsion, these failures are not due to the problem being strongly multireference. The double hybrid DFT methods are considerably better than other DFT methods, but do not quite attain the accuracy of MP2 methods though they incur similar computational costs. This DFT error is encountered in the study of proteins which include parallel and antiparallel beta sheets and attenuates amide resonance and hence necessarily RAHB in the hydrogen bonded chains of backbone amides. Since RAHB is cooperative, if all errors in calculations of amide resonance give values that are lower than accurate, then the errors themselves may be seen as cooperative. Since every backbone amide in a beta sheet is subject to roughly the same torsional hyperconjugation, hydrogen bonding in the beta sheet is cooperatively reduced. While the smaller basis sets such as D95** [136], 6-31G** and def2-SVP [137] are associated with the largest reductions in amide resonance, these basis sets have been popular because they are the largest that could be used at the atom counts of beta sheets. Great caution is needed in interpreting the results of applying established DFT methods to proteins.

Of the non-double hybrid DFT methods and basis set combinations tested, MN12-L/def2-QZVPP is least disturbed by torsional hyperconjugation. MN15-L and MN15 [138] have not yet been tested. MN12-L being a local functional and def2-QZVPP having no diffuse functions allows for efficiency of calculation, but no non-double hybrid DFT method tested yields occupancies comparable to those calculated by wavefunction methods and correlation consistent basis sets for the orbitals central to amide resonance, N(lp) and C-O(p)*. Though the multireference character of the problem is slight, CASSCF/MRCI+Q gives occupancies associated with greater resonance than does DLPNO-CCSD(T), leading to the concern that multireference methods may be necessary for accurate calculation of RAHB in proteins. Perhaps much improved scaling approximations to full configuration interaction [139, 140] will be useful. DLPNO-CCSD(T) in turn calculates occupancies associated with greater resonance than do non-double hybrid DFT methods, so non-double hybrid DFT methods then fail to accurately calculate absolute amide resonance regardless of O-C-CA-H torsion.

While this work is focused on amide resonance and RAHB, any non-double hybrid DFT calculated amide property is at risk of being significantly inaccurate due to this error. A range of quantities are disturbed by this error along with carbonyl sigma/pi separation and bond straightness, including charge transfers N(lp)->C-O(p)* and N(lp)->C-O(s)* and steric interactions C-O(p)|C-O(s), O(lp-s)|C-O(p) and O(lp-s)|C-O(s).

It can be shown with molecular species extending ethanamide to approximate alanine and valine that the loss of local symmetry with DFT is very sensitive to residue type. With a heterogeneous amino-acid beta sheet, DFT geometry optimization will be erratic rather than reliably wrong. Also, this sensitivity invites the erroneous conclusion that protein folding is influenced by this sidechain modulation of backbone amide resonance.

MP2 has a basis set and optimization implementation dependent propensity for large variation in pyramidalization at the amide nitrogen in ethanamide at C-CA torsion. This pyramidalization necessarily competes with amide resonance. When the erroneous tendency to pyramidalize becomes dominant, calculation will under-estimate RAHB for hydrogen bonded chains in which the amide is potentially involved. Also, the flexibility of a protein backbone is incorrectly increased since the pyramidalization may be switched from one side to another by backbone strain, giving large backbone amide omega torsion. This variation is not limited to Pople basis sets, and is seen to a lesser extent and in small ranges of C-CA torsion with def2-TZVP, def2-QZVPP and aug-cc-pVTZ. This is not exclusively an MP2 issue, for the quite recent DFT method M11-L also demonstrates this pyramidalization at antiparallel beta strand psi torsion with most basis sets tested. This pyramidalization error is not unambiguously present in geometry-optimized protein coordinates since this pyramidalization may arise by other means such as backbone strain and hydrogen bonding, so validation of methods with respect to this error is best done on small amides such as ethanamide or N-methylethanamide. In ethanamide, this second problem can be corrected by geometry optimization on the Counterpoise Corrected potential energy surface by definition of fragment boundaries at the C-N bond, but not at the C-C bond.

We recommend that non-wavefunction electronic structure calculation of protein be accompanied by NBO analysis of backbone amide carbonyl local symmetry and bond bending. Since hyperconjugations involving C-O(s)* and C-O(p)* are available at various C-CA torsions, this recommendation extends beyond the secondary structures tested here, antiparallel beta sheets, to other protein secondary structures that have hydrogen bonded chains of backbone amides. While the consequences of inaccurate calculation of amide resonances are not amplified in random coil as in secondary structure, they may be expected to be significant in determining random coil structure and hydrogen bonding networks.

NBO-based assessment of maintenance of local symmetry and bond bending is proposed for evaluation of existing or development of new electronic structure system methods. The tests used here involving vinylamine and ethanamide could serve as standard benchmarks. Measurement of pyramidalization and NBO assessment of sp hybridization at the amide nitrogen in ethanamide with carbonyl hyperconjugation at C-C torsion could similarly serve as a standard benchmark.

3.10 Acknowledgements

Prof. Peter M. W. Gill is acknowledged for reading the earliest draft of this manuscript and pointing out the resemblance of the second mentioned problem, pyramidalization at amide nitrogen at wavefunction method/Pople basis set, to the benzene non-planarity problem [14].

eResearch South Australia is acknowledged for hosting and administering machines provided under Australian Government Linkage, Infrastructure, Equipment and Facilities grants for Supercomputing in South Australia, directing funds to the acquisition of Nvidia Tesla GPU nodes and allocating 64 CPU cores and 256 GB RAM of the NeCTAR Research Cloud (a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy) to the present work.

3.11 Appendix 1

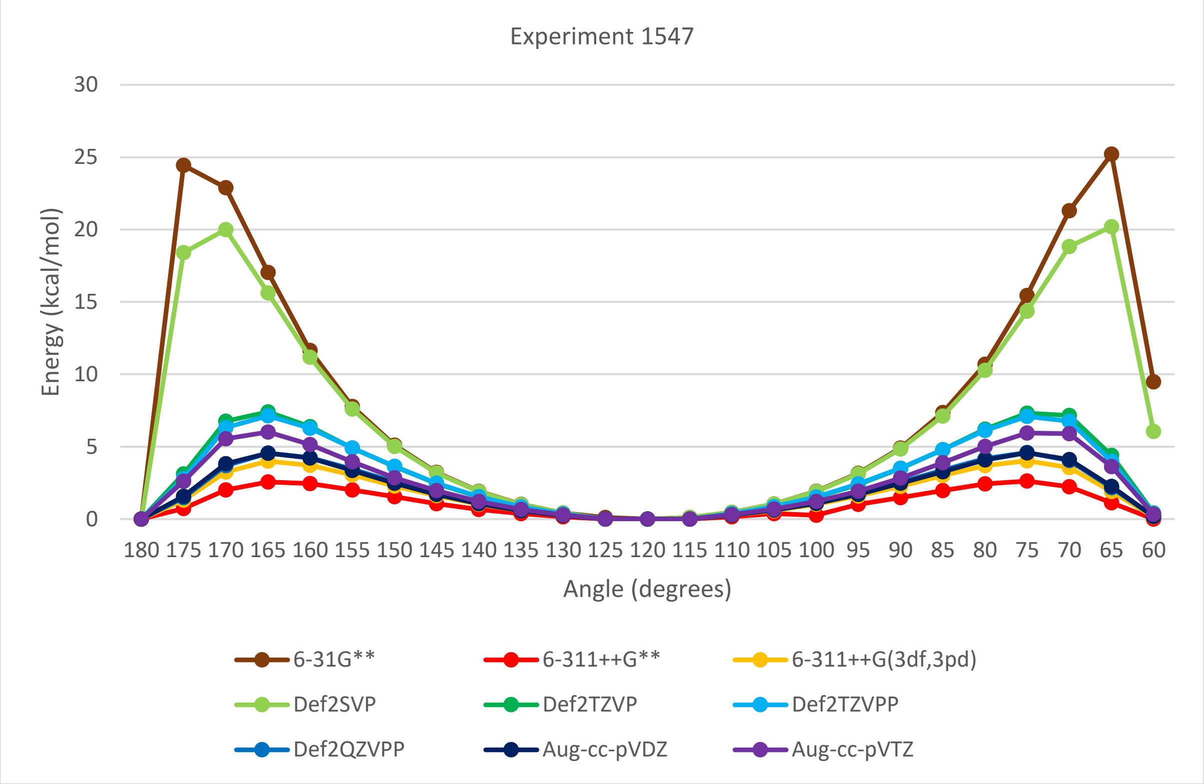


Figure 3.19. C-O(p)|C-O(s) Steric Exchange Energy in Ethanamide at O-C-CA-H Dihedral Angle with B3LYP

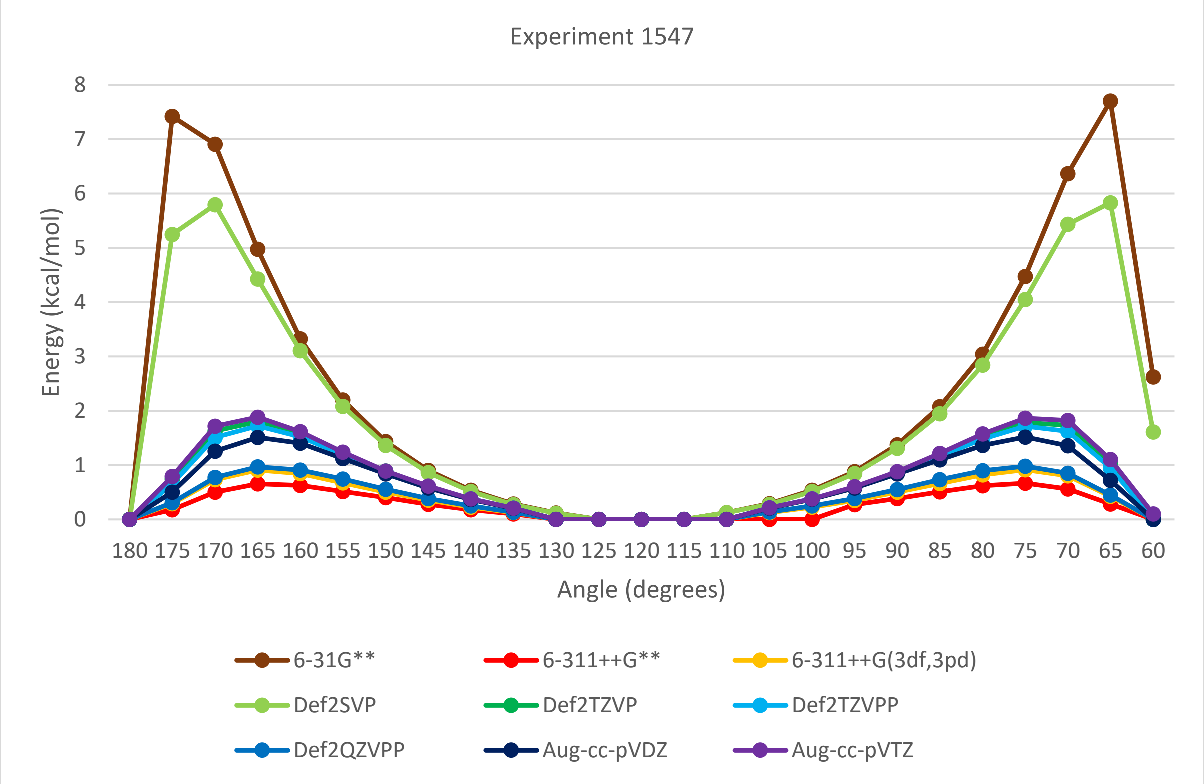


Figure 3.20. O(lp-s)|C-O(p) Steric Exchange Energy in Ethanamide at O-C-CA-H Dihedral Angle with B3LYP

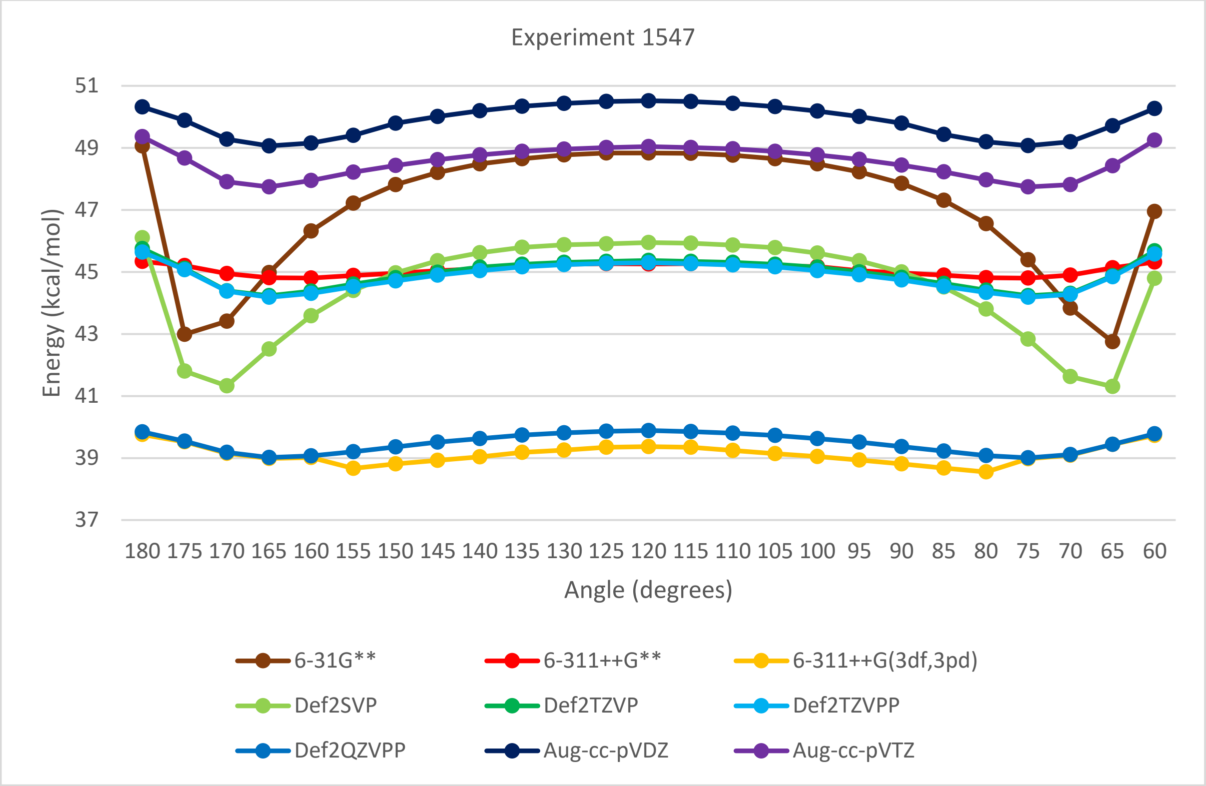


Figure 3.21. O(lp-s)|C-O(s) Steric Exchange Energy in Ethanamide at O-C-CA-H Dihedral Angle with B3LYP

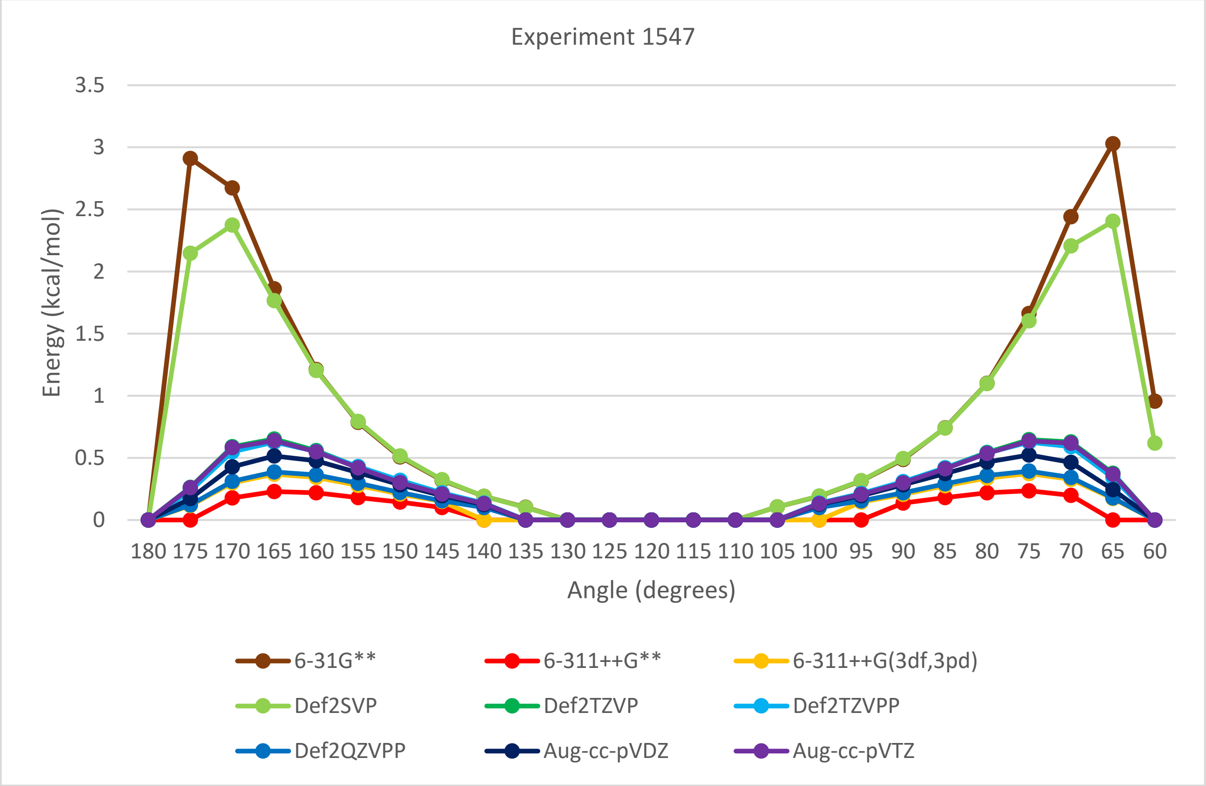


Figure 3.22. N(lp)->C-O(s)* SOPT Energy in Ethanamide at O-C-CA-H Dihedral Angle with B3LYP. Values less than 0.01 as zero.

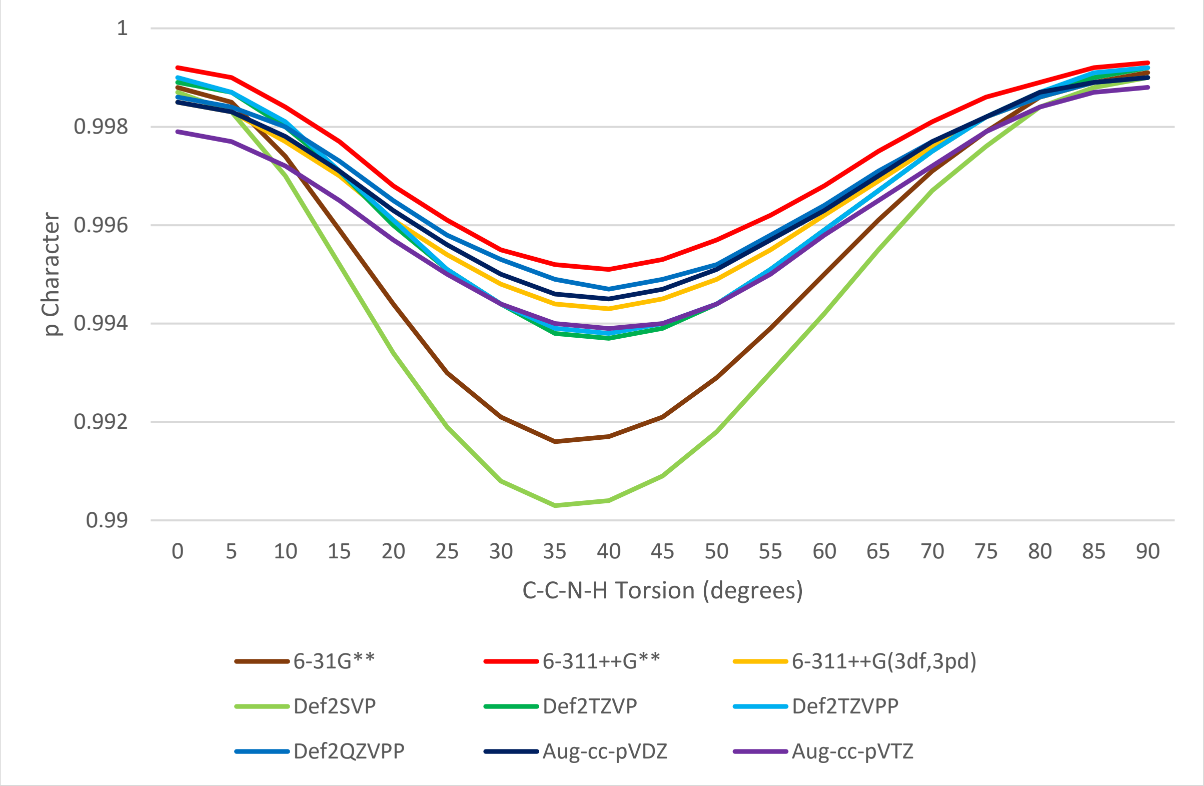


Figure 3.23. C-C(pi) Central Carbon NHO p Character in Vinylamine at Varying C-C-N-H Torsion and Basis Set at MP2

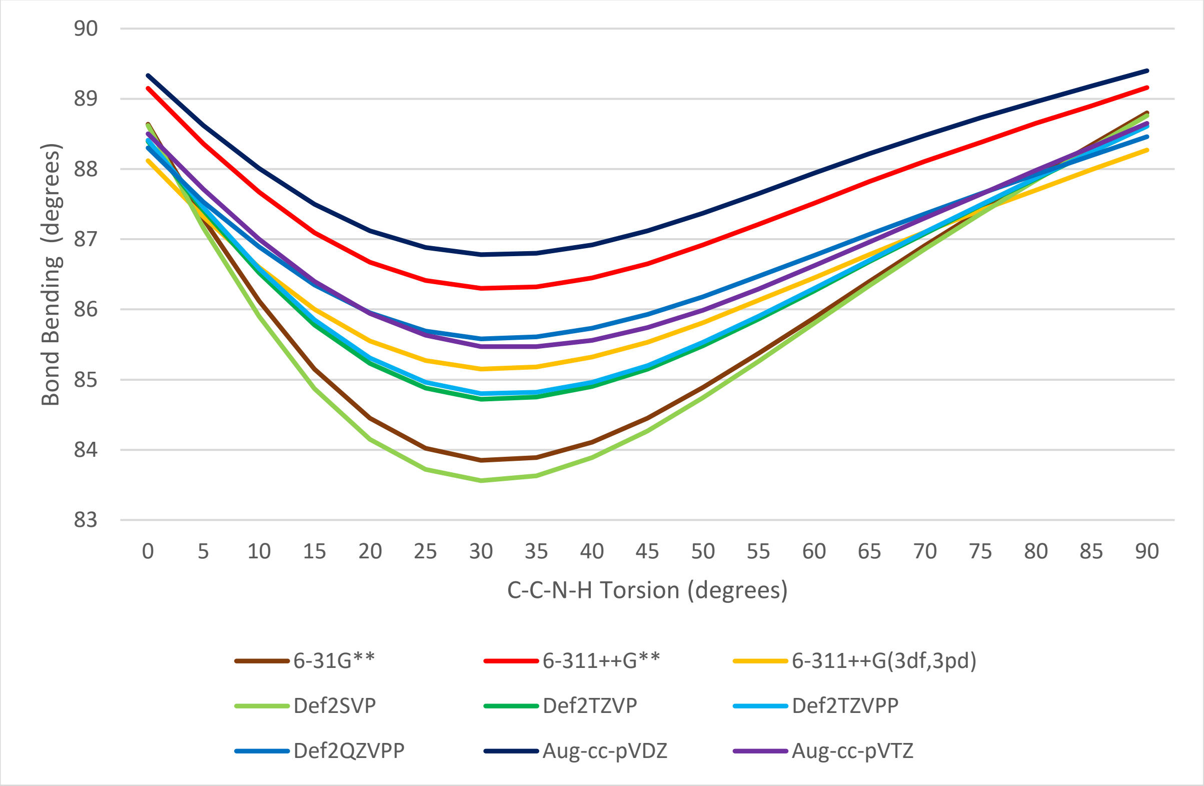


Figure 3.24. C-C(pi) Central Carbon NHO Bond Bending in Vinylamine at Varying C-C-N-H Torsion and Basis Set at MP2

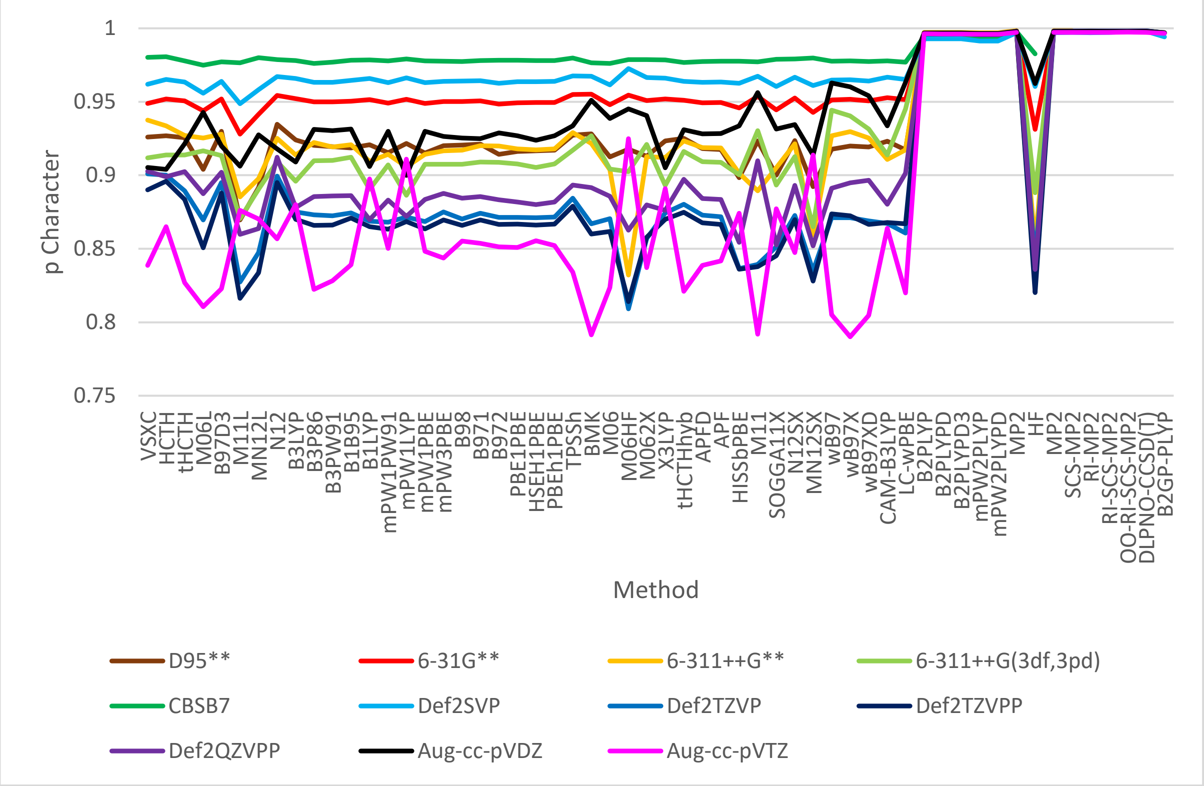


Figure 3.25. C-C(pi) Central Carbon NHO p Character in Vinylamine at 10 Degree C-C-N-H Torsion with Varying Method and Basis Set

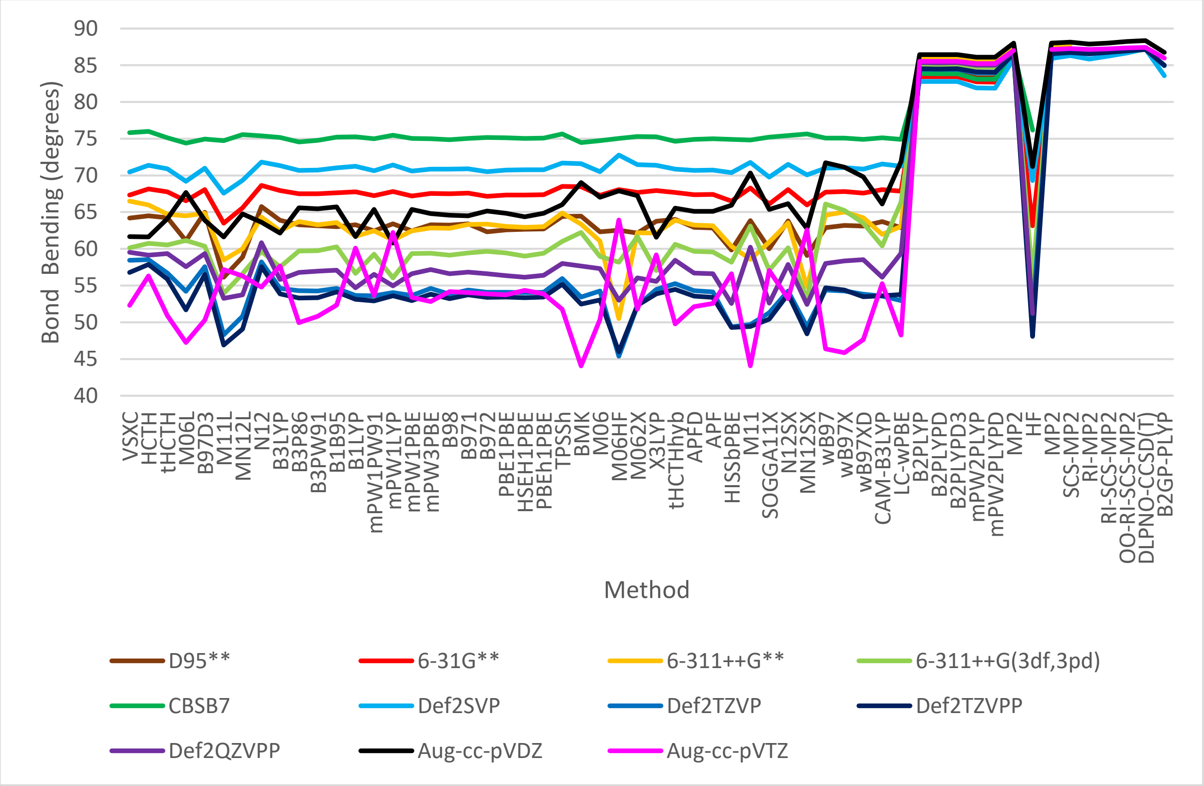


Figure 3.26. C-C(pi) Central Carbon NHO Bond Bending in Vinylamine at 10 degree C-C-N-H Torsion with Varying Method and Basis Set

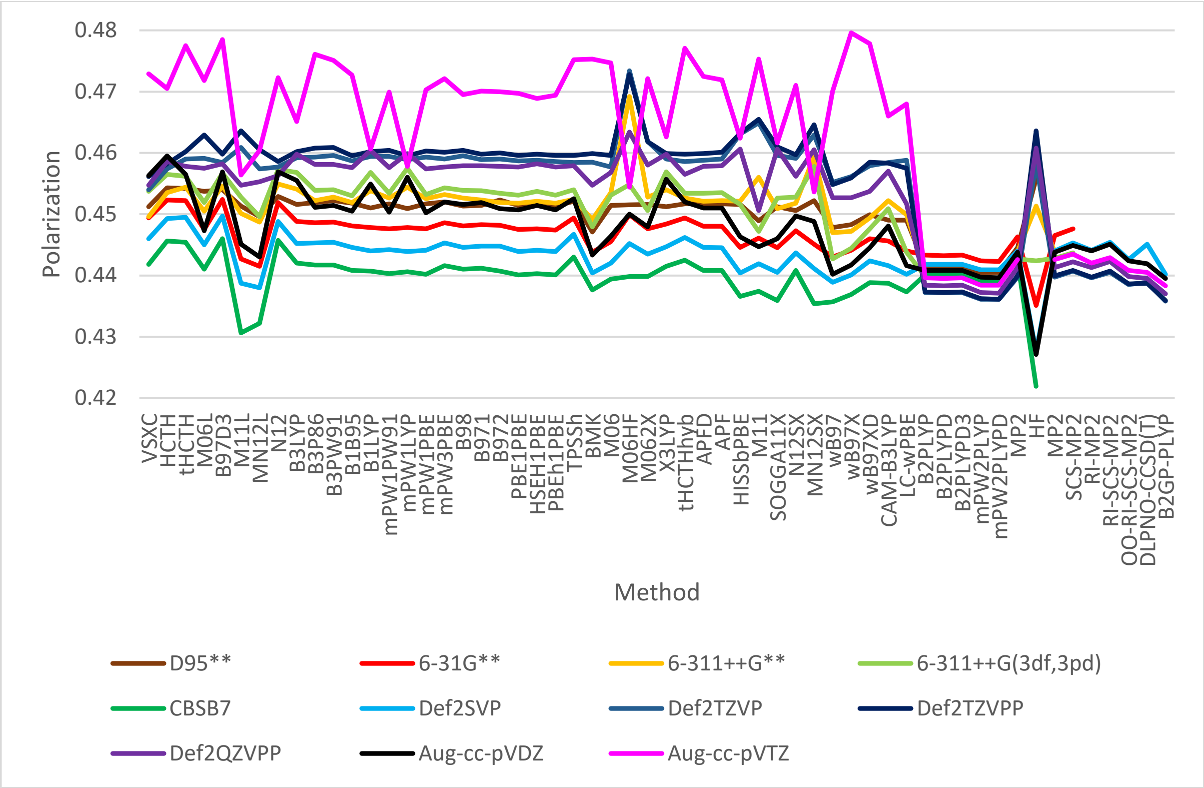


Figure 3.27. C-C(pi) Central Carbon NHO Polarization in Vinylamine at 10 degree C-C-N-H Torsion with Varying Method and Basis Set

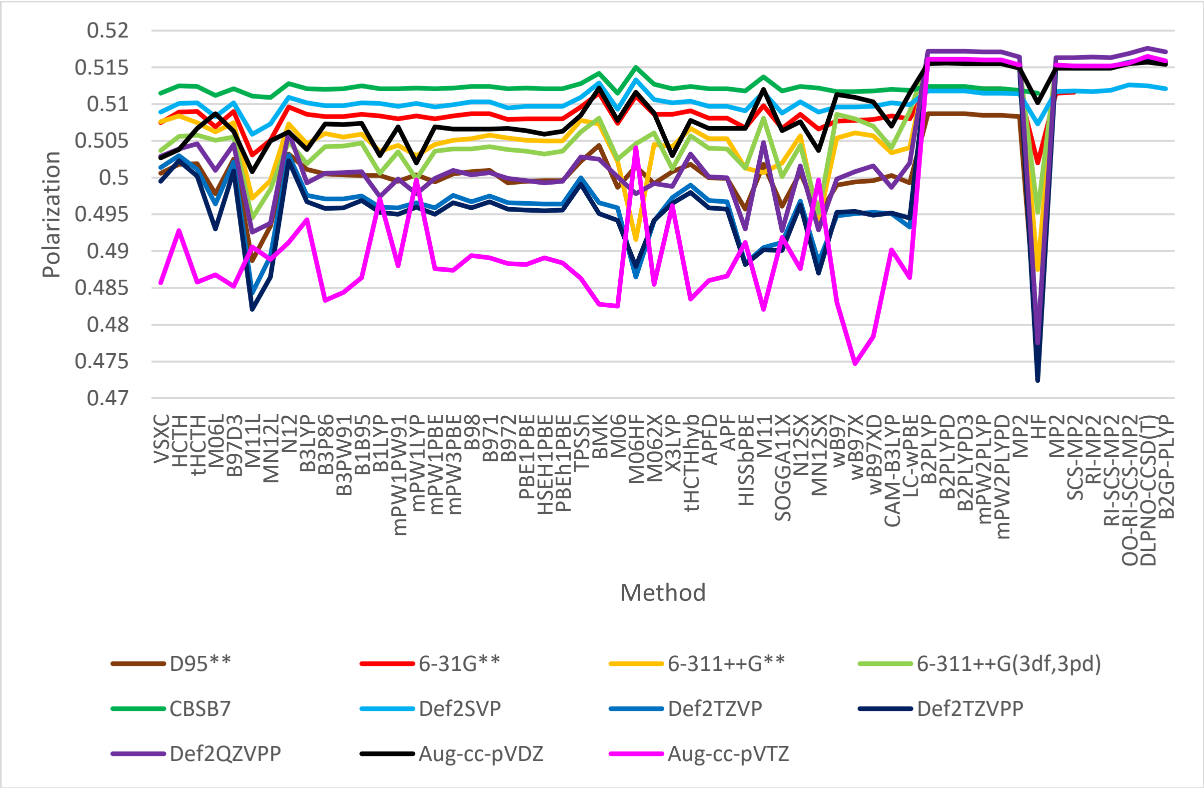


Figure 3.28. C-C(sigma) Central Carbon NHO Polarization in Vinylamine at 10 Degree C-C-N-H Torsion with Varying Method and Basis Set

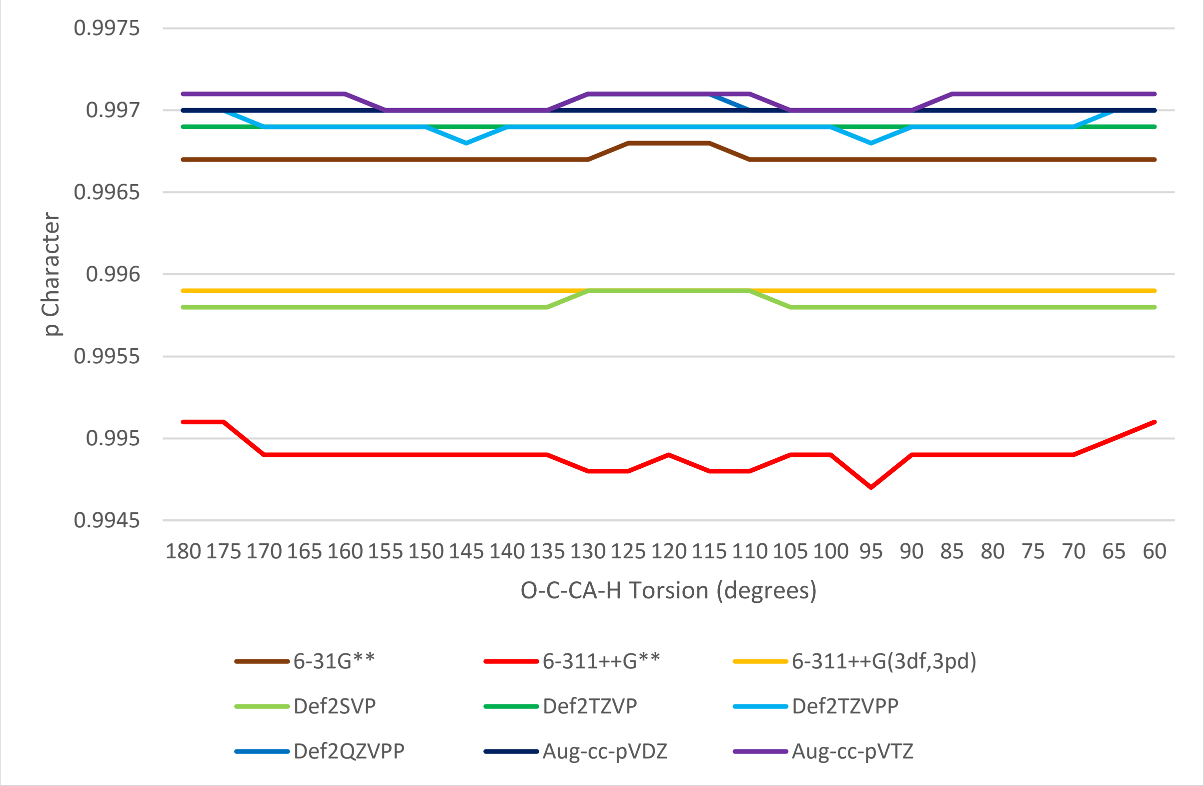


Figure 3.29. C-O(pi) Carbon NHO p Character in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at MP2

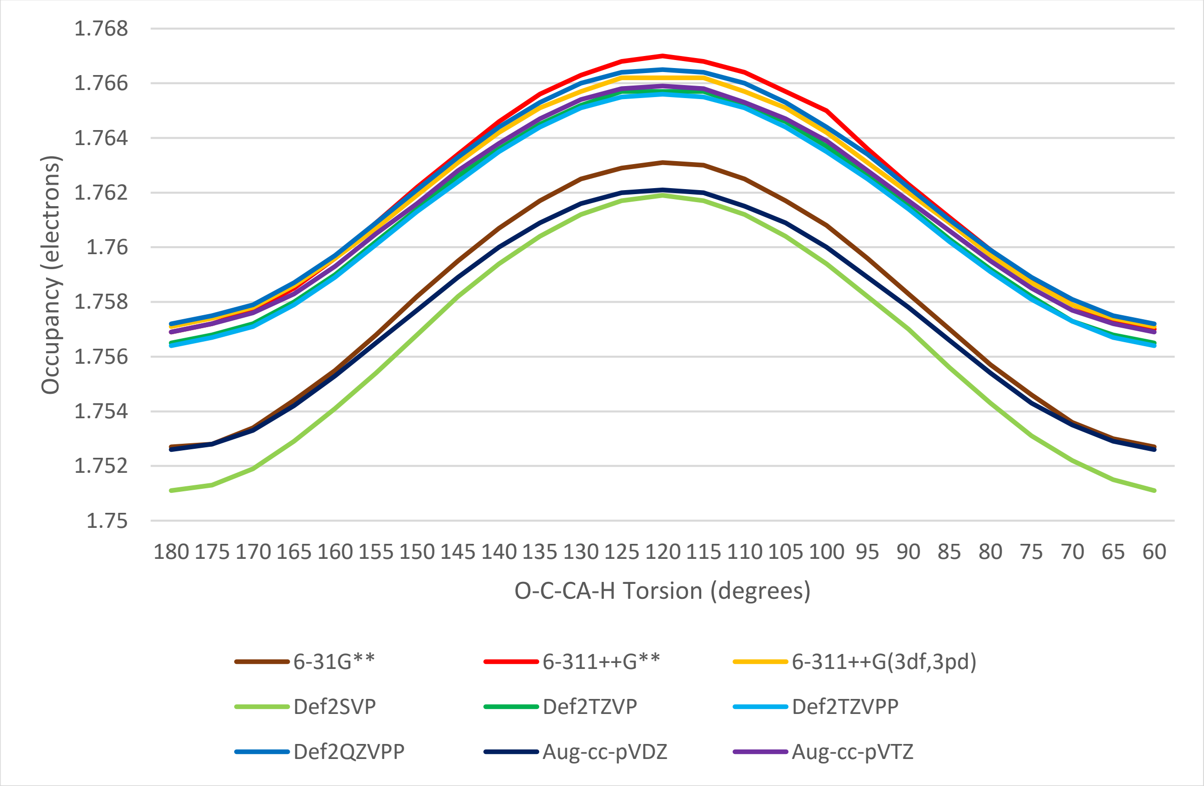


Figure 3.30. Nitrogen Lone Pair NBO Occupancy in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at B3LYP

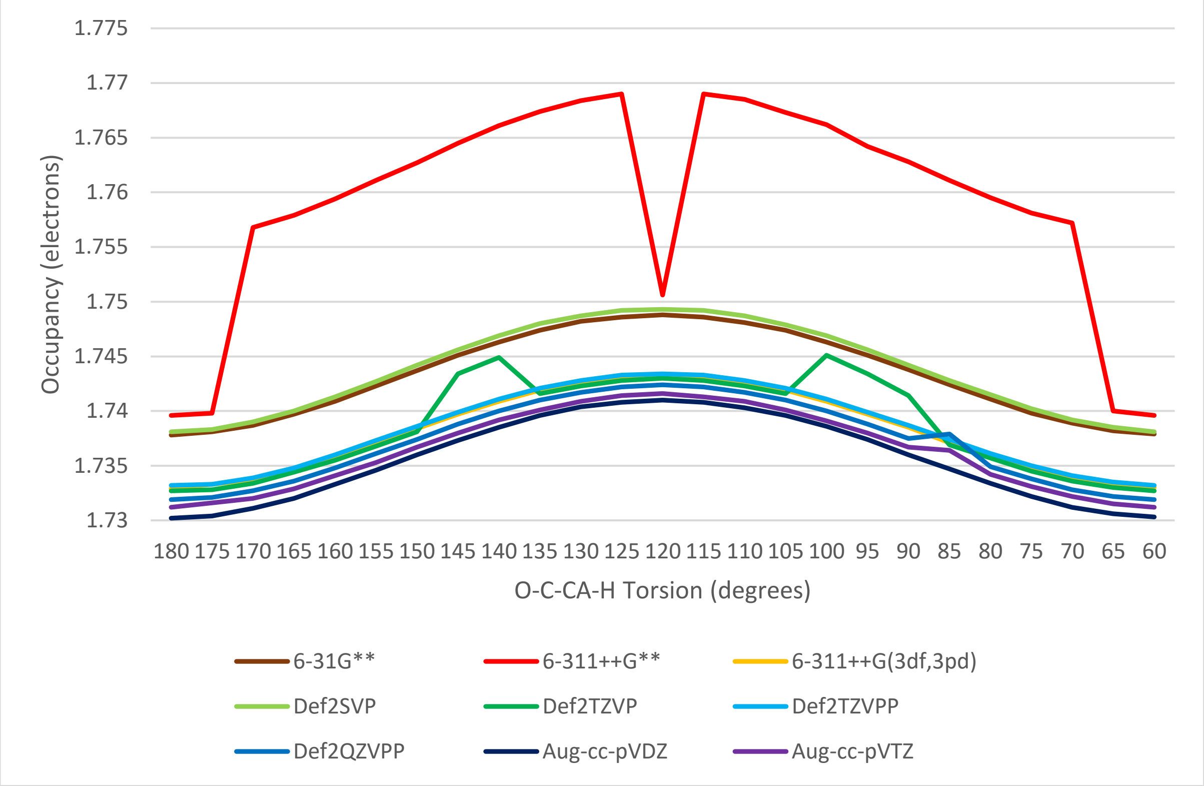


Figure 3.31. Nitrogen Lone Pair NBO Occupancy in Ethanamide at Varying O-C-CA-H Torsion and Basis Set at MP2

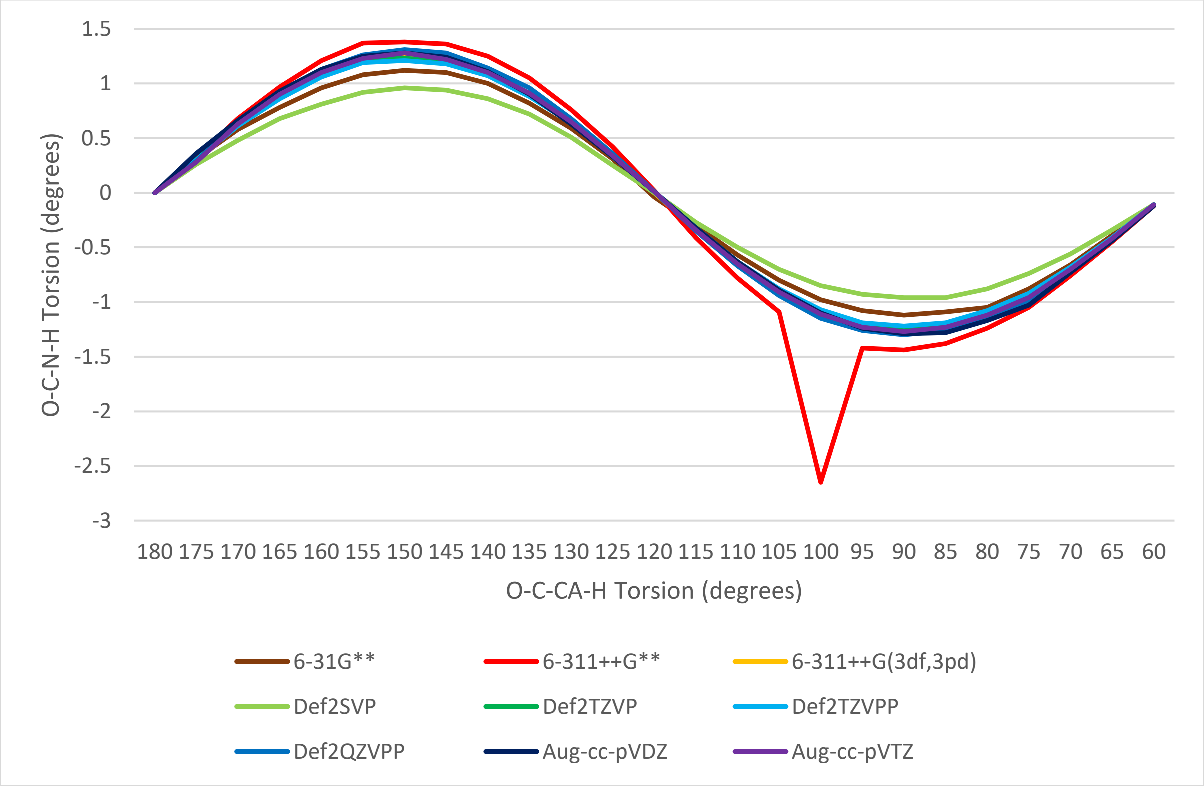


Figure 3.32. O-C-N-H Torsion in Ethanamide with Varying O-C-CA-H Torsion and Basis Set at B3LYP

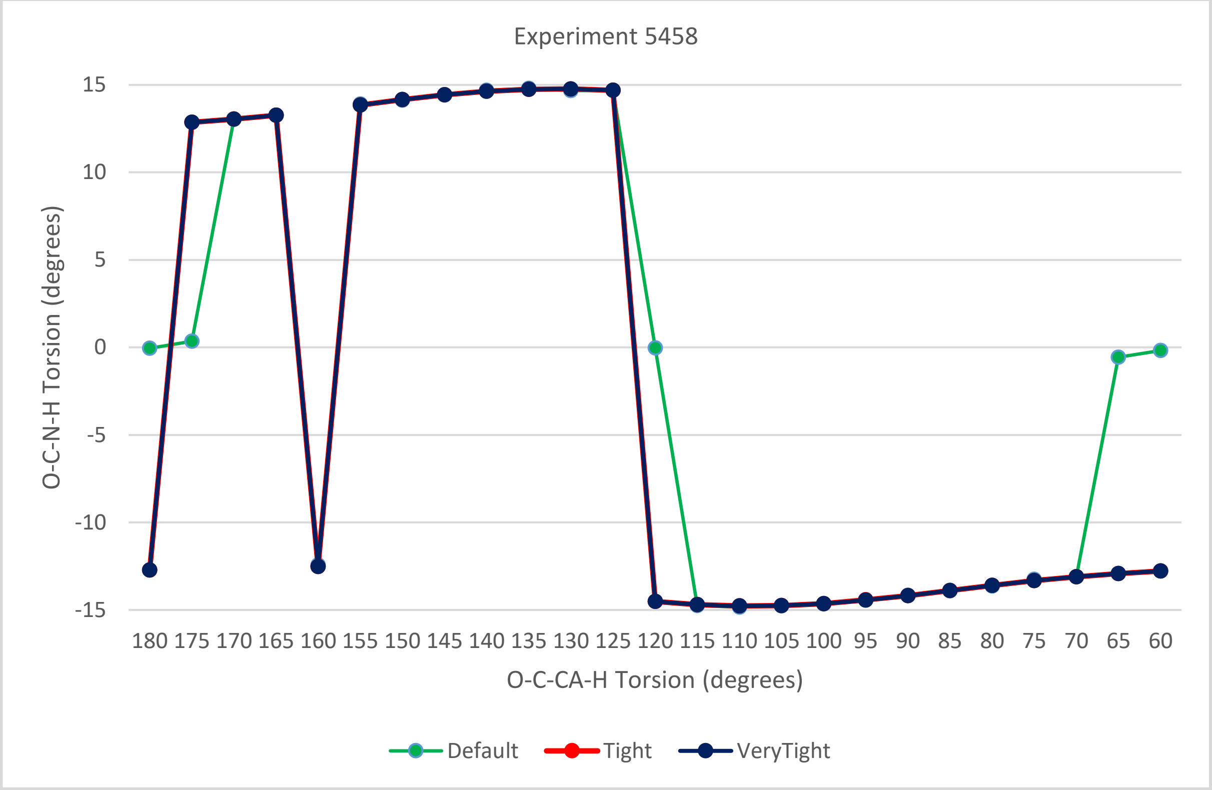


Figure 3.33. Ethanamide O-C-N-H Torsion with MP2/6-311++G** UltraFineGrid at Varying O-C-CA-H Torsion and Gaussian Optimization Convergence Label

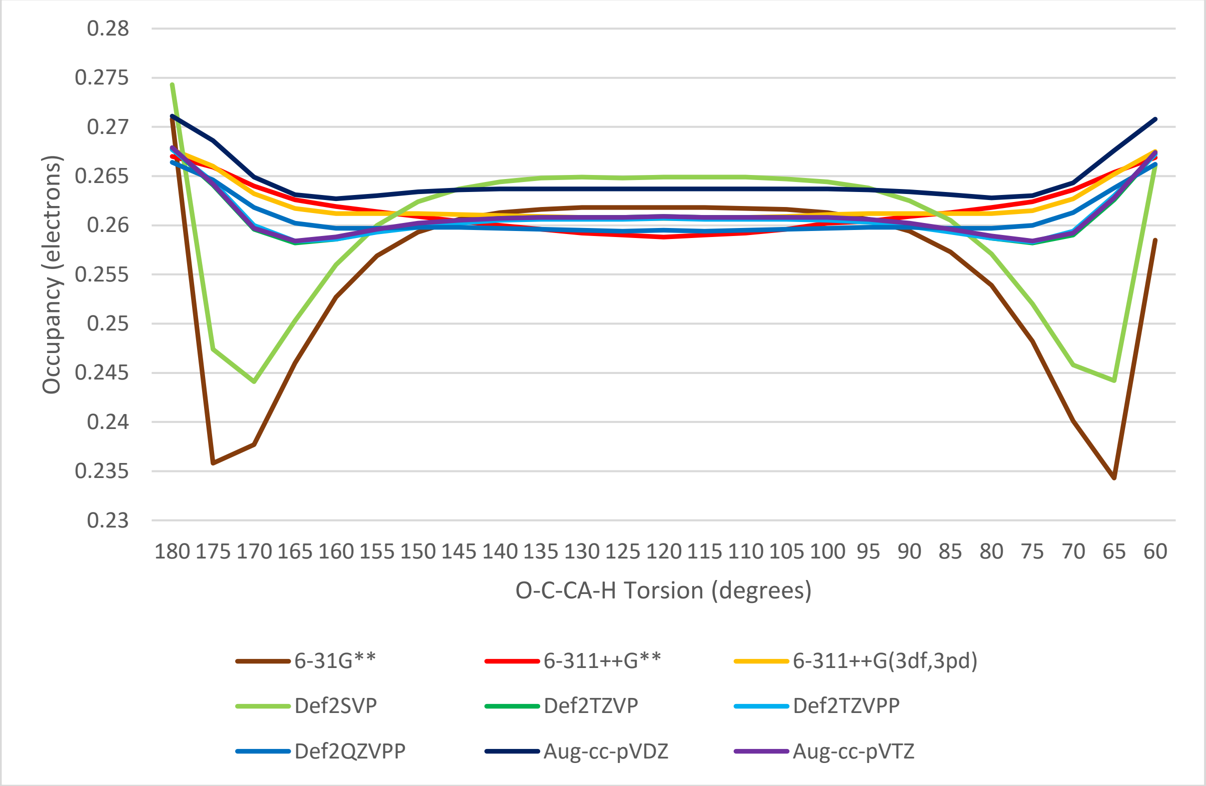


Figure 3.34. C-O(pi)* NBO Occupancy at Varying O-C-CA-H Torsion and Basis Set at B3LYP

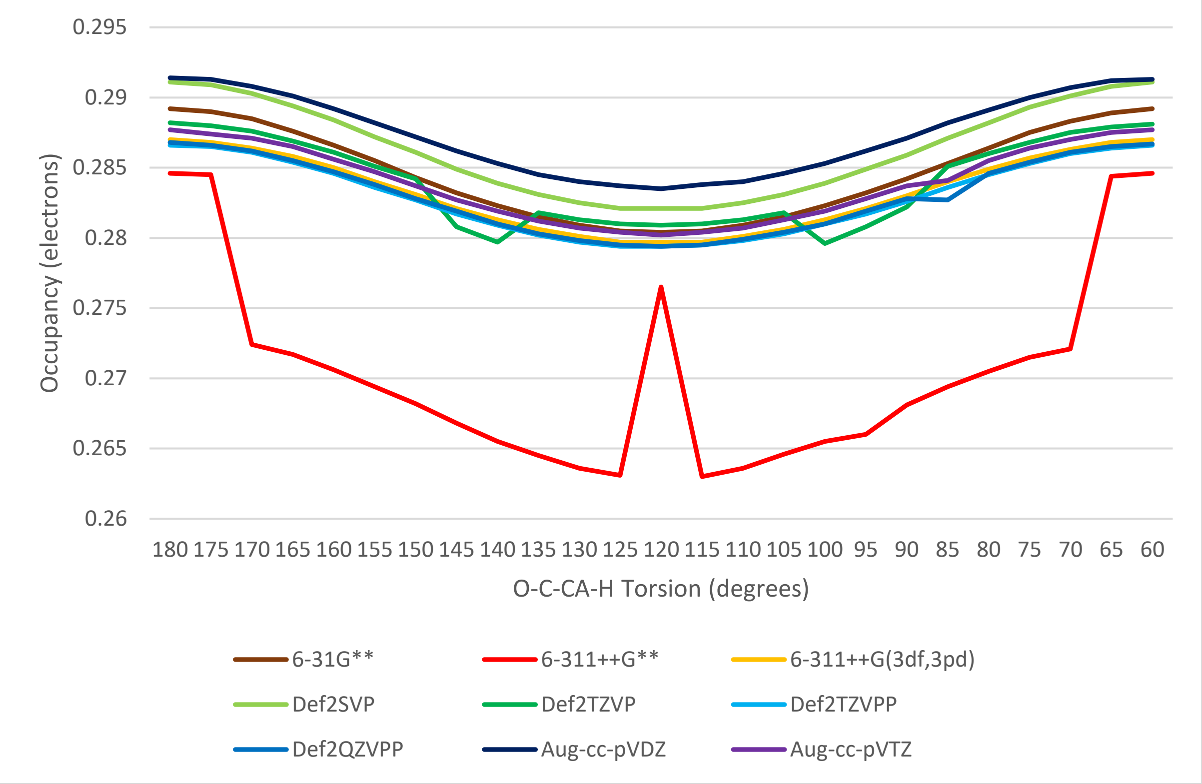


Figure 3.35. C-O(pi)* NBO Occupancy at Varying O-C-CA-H Torsion and Basis Set at MP2

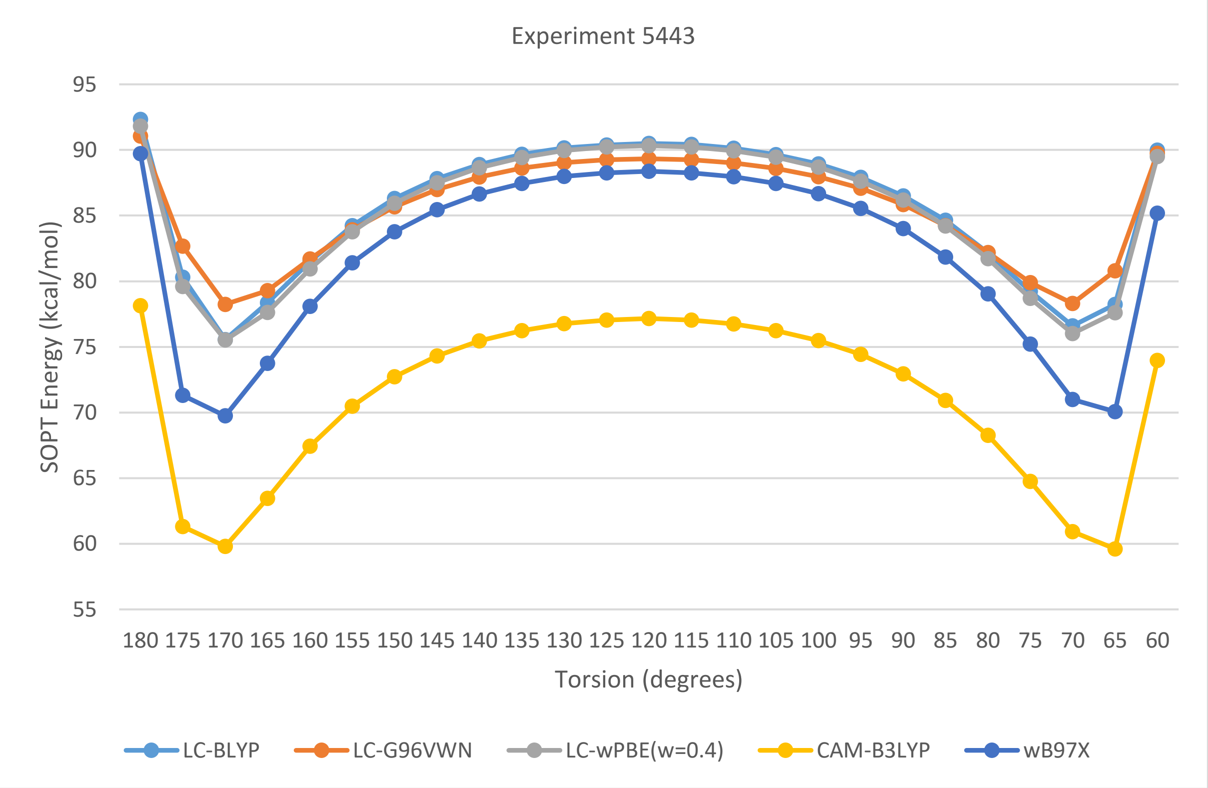


Figure 3.36. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Long Range Corrected Method

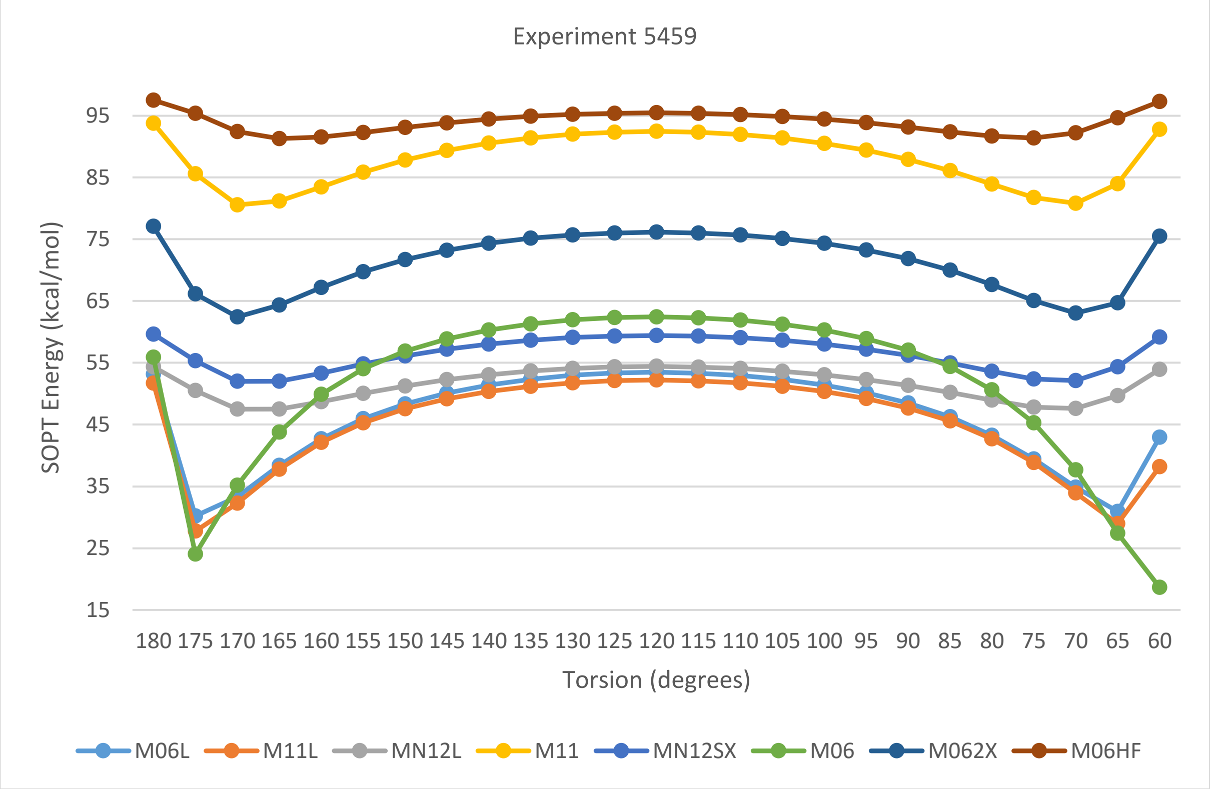


Figure 3.37. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion and Method

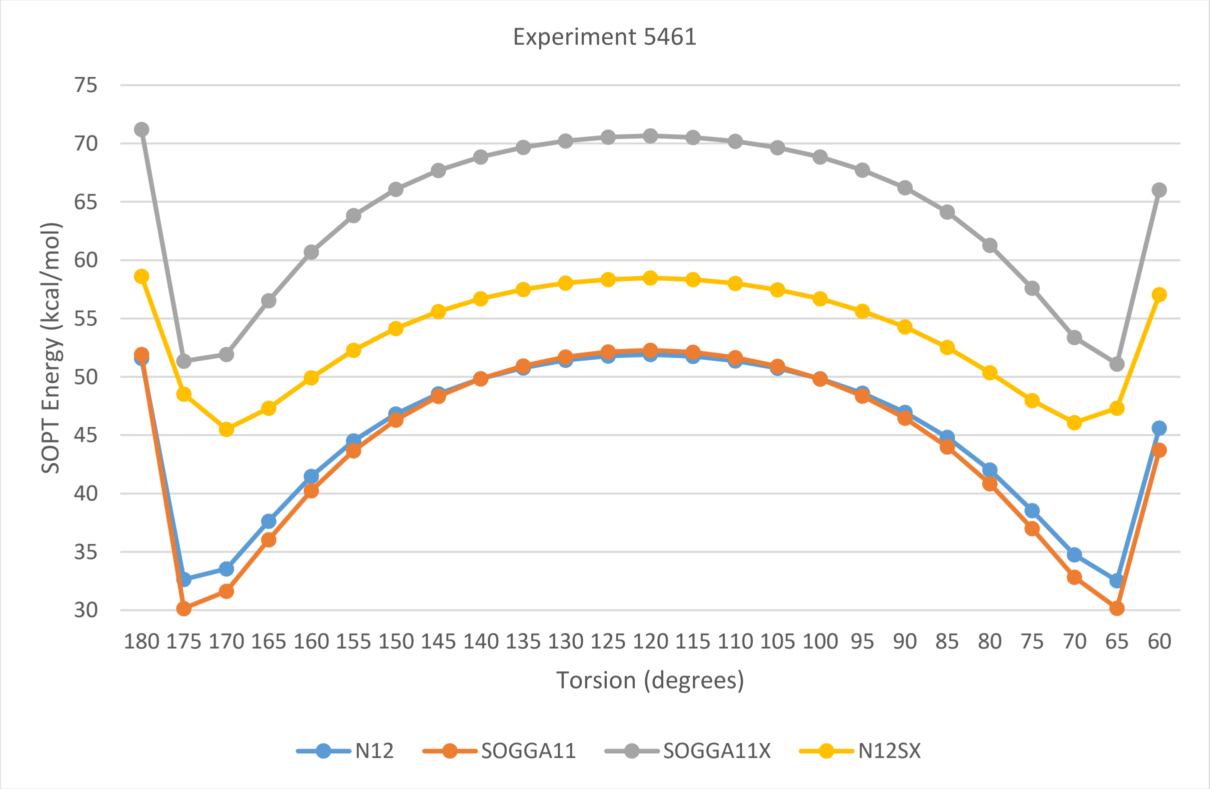


Figure 3.38. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion and Method

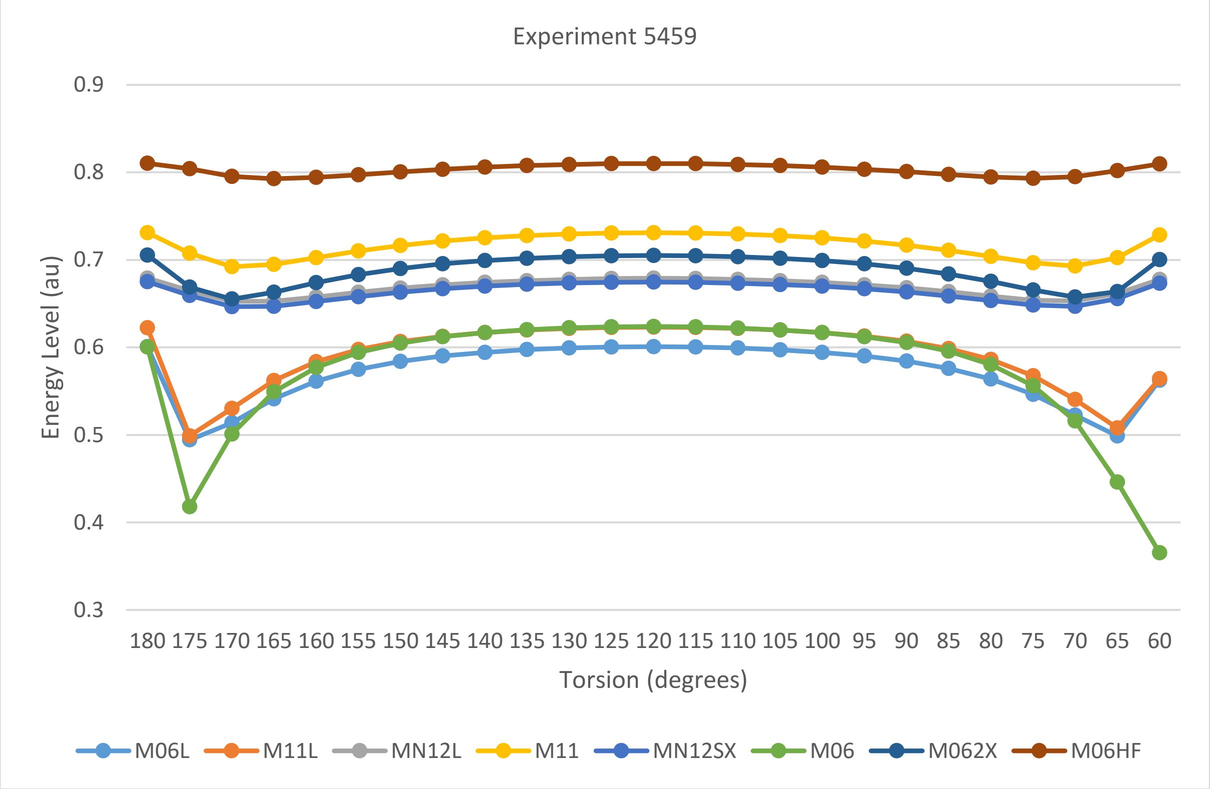


Figure 3.39. Ethanamide C-O(s)* Energy Level with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method

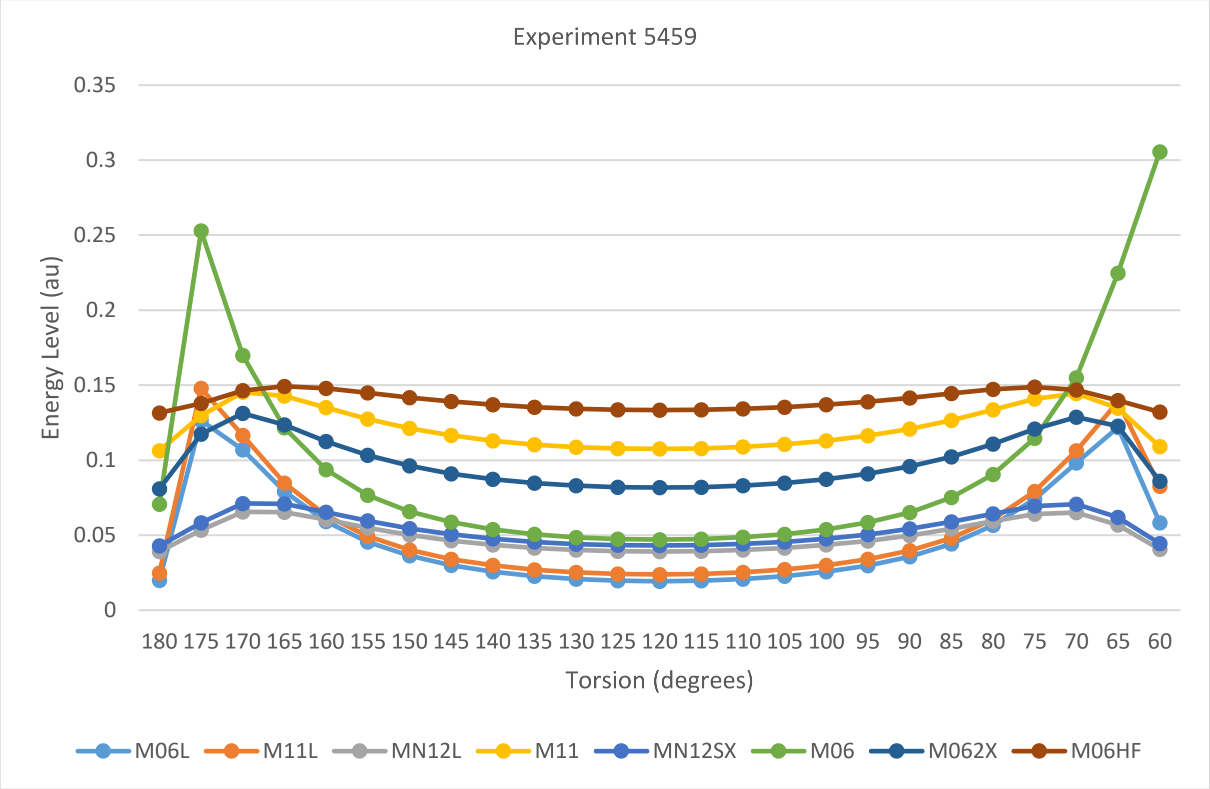


Figure 3.40. Ethanamide C-O(p)* Energy Level with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method

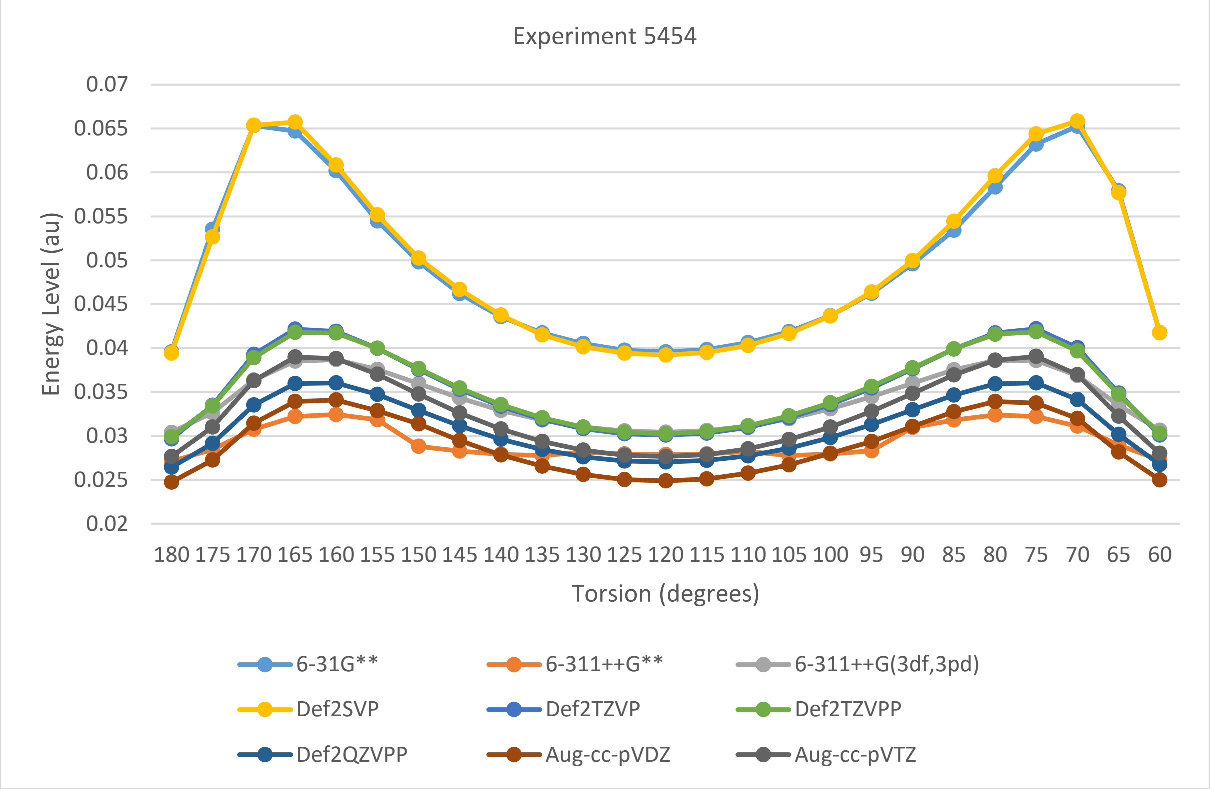


Figure 3.41. Ethanamide C-O(p)* Energy Level with MN12-L UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

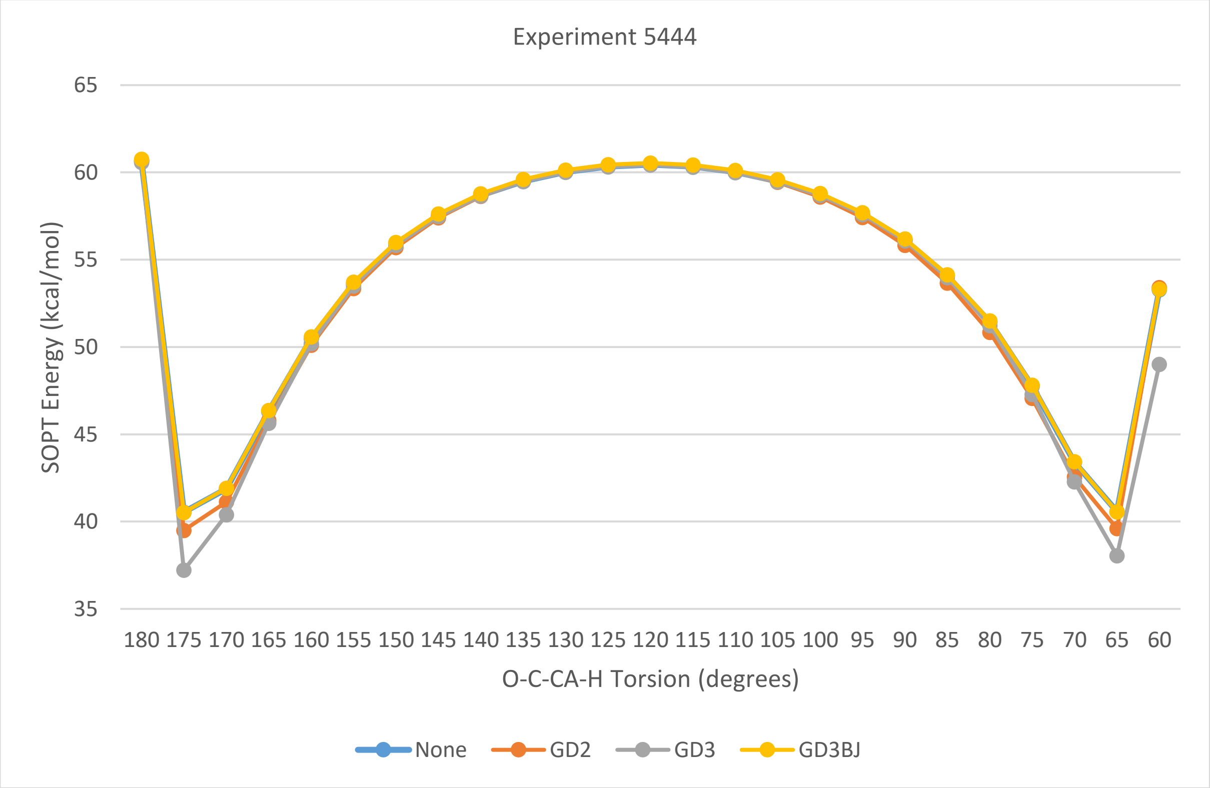


Figure 3.42. Ethanamide N(lp)->C-O(p)* SOPT Energy with B3LYP/6-31G** and UltraFineGrid at Varying O-C-CA-H Torsion and Empirical Dispersion Scheme

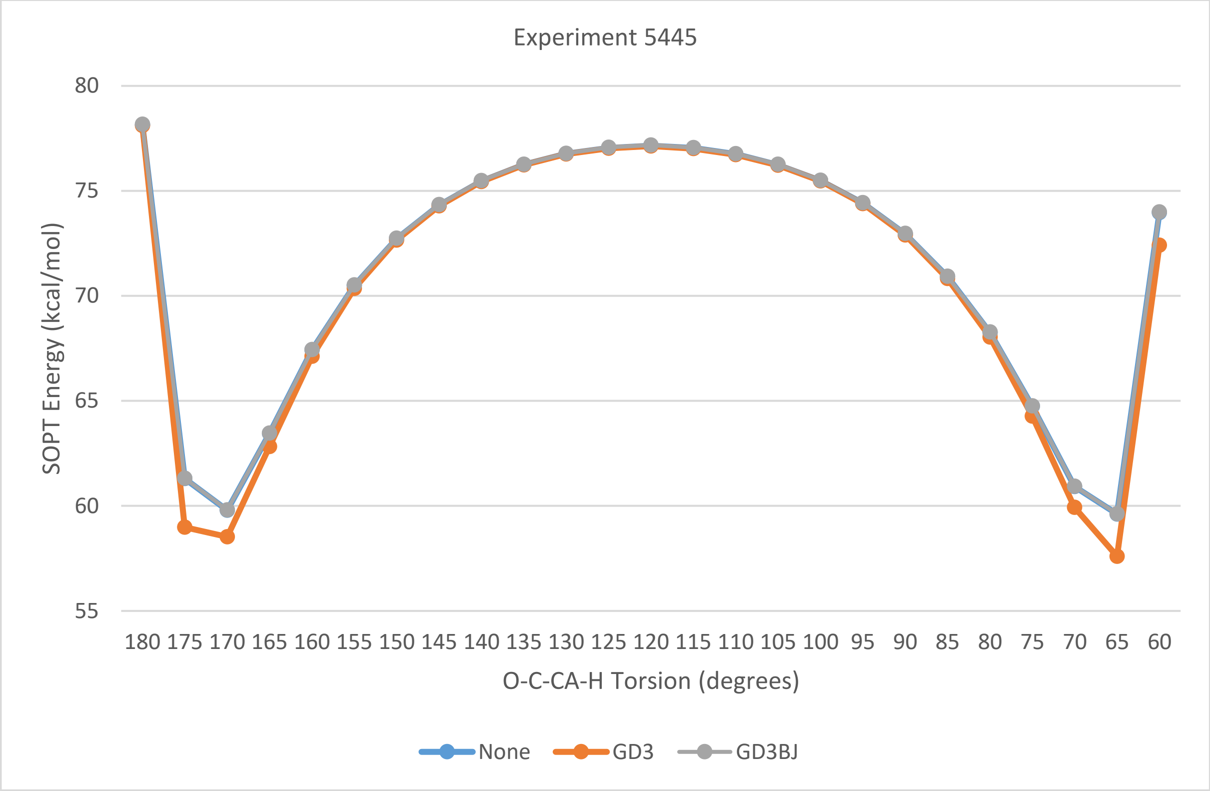


Figure 3.43. Ethanamide N(lp)->C-O(p)* SOPT Energy with CAM-B3LYP/6-31G** and UltraFineGrid at Varying O-C-CA-H Torsion and Empirical Dispersion Scheme

Figure 3.44. Ethanamide N(lp)->C-O(p)* SOPT Energy with LC-wPBE(w=0.4)/6-31G** and UltraFineGrid at Varying O-C-CA-H Torsion and Empirical Dispersion Scheme

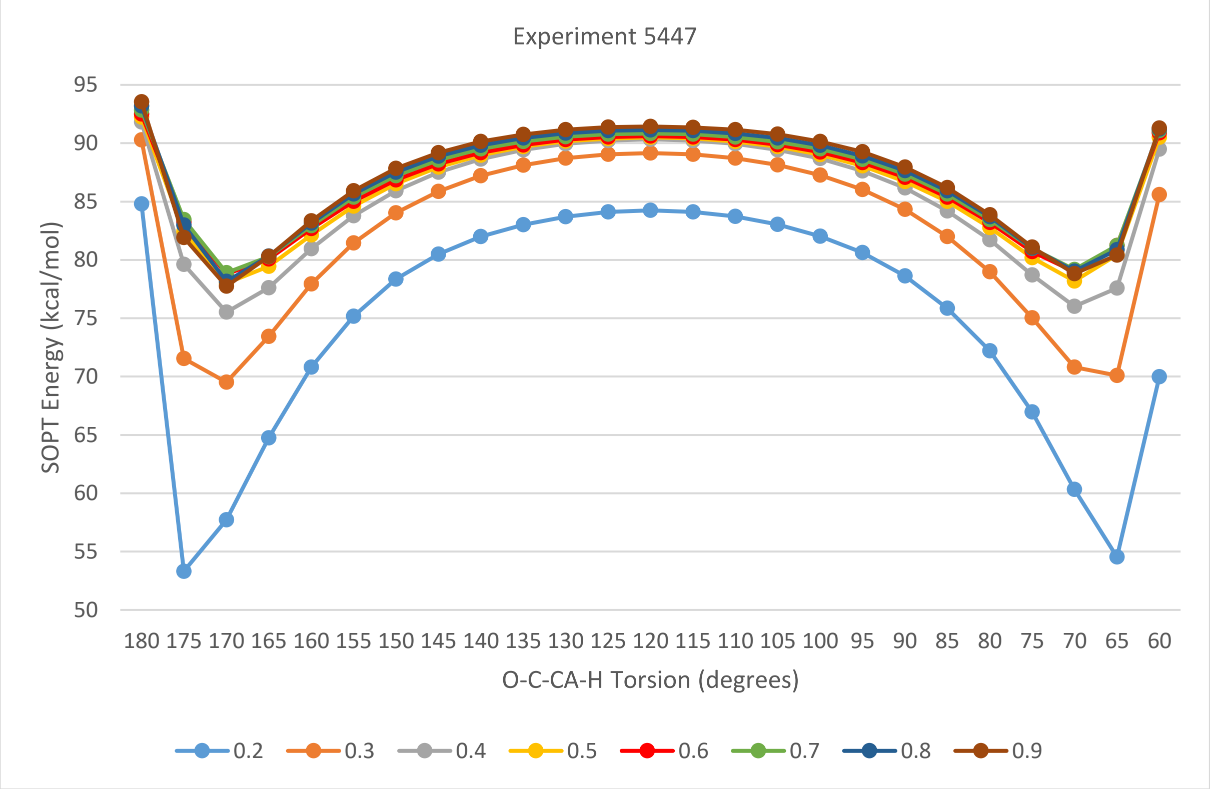


Figure 3.45. Ethanamide N(lp)->C-O(p)* SOPT Energy with LC-wPBE/6-31G** and UltraFineGrid at Varying O-C-CA-H Torsion and Omega

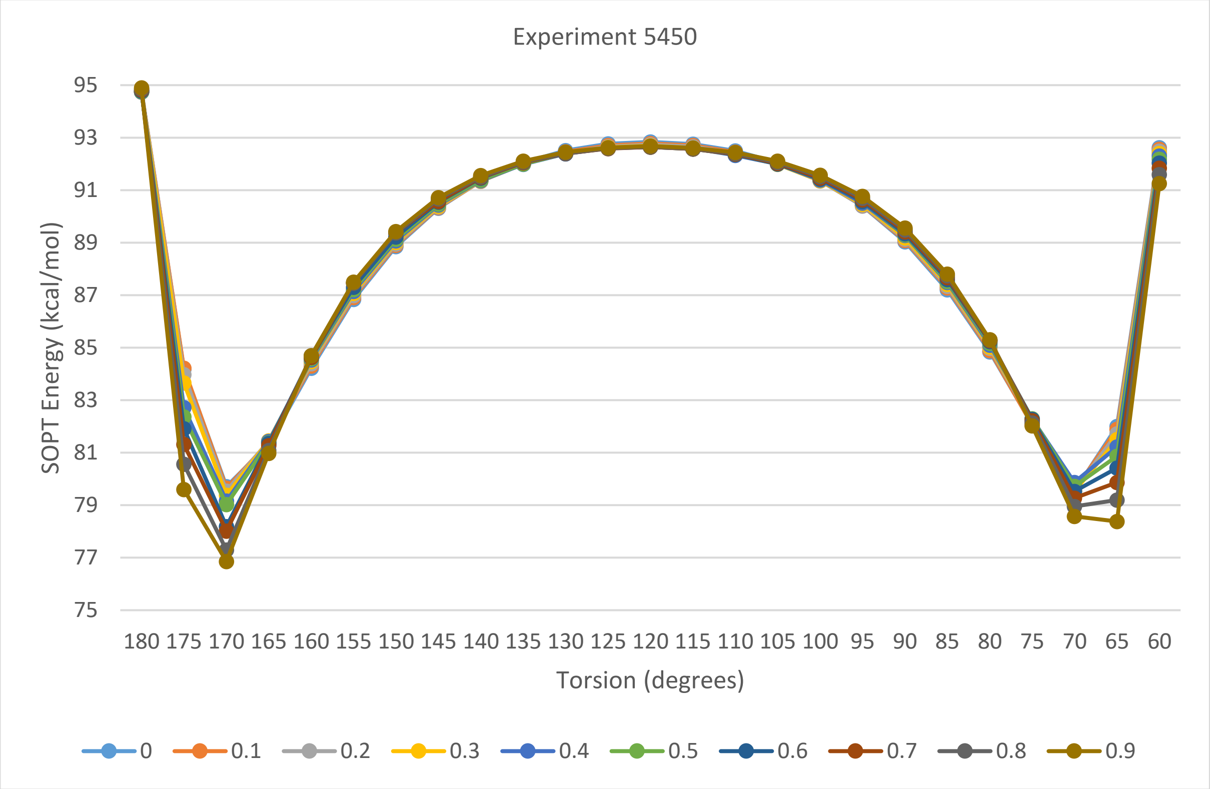


Figure 3.46. Ethanamide N(lp)->C-O(p)* SOPT Energy with LC-PBEhPBE at Omega=0.6 and Varying O-C-CA-H Torsion and HF Exchange Coefficient

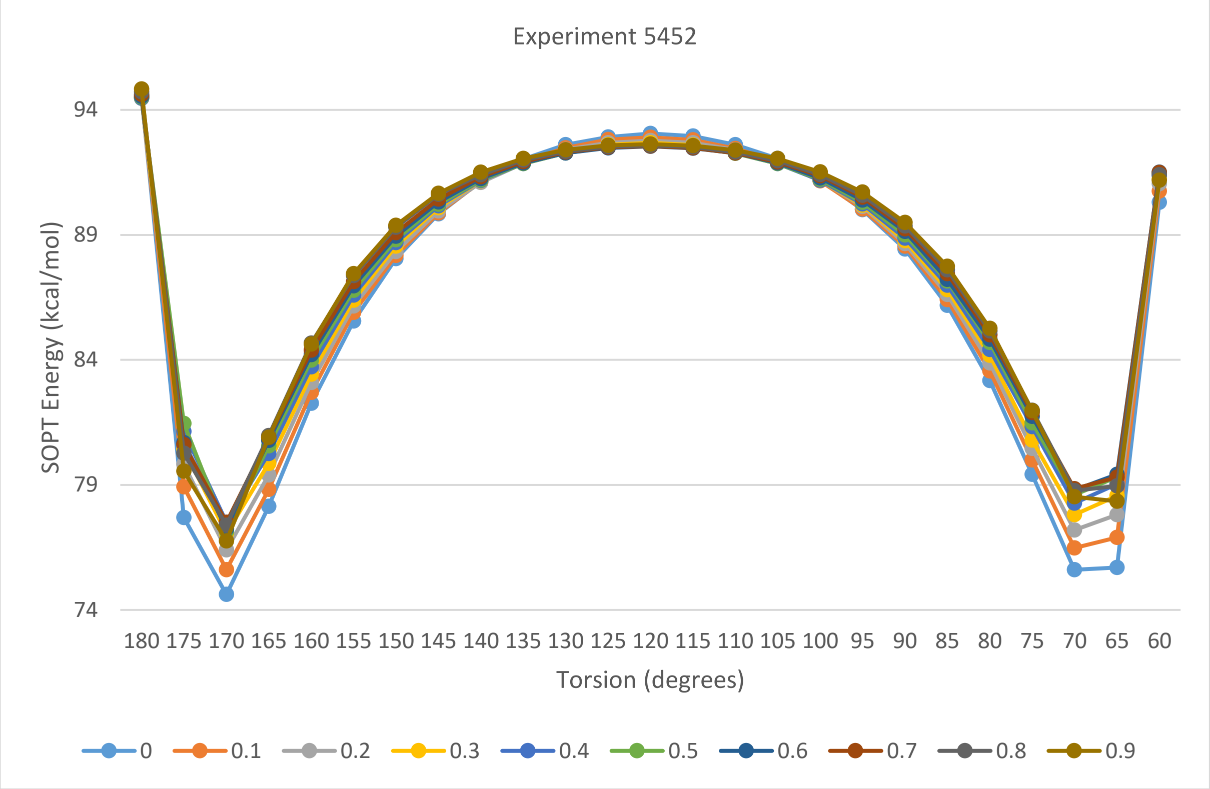


Figure 3.47. Ethanamide N(lp)->C-O(p)* SOPT Energy with LC-PBEhPBE at Omega=0.4 and Varying O-C-CA-H Torsion and HF Exchange Coefficient

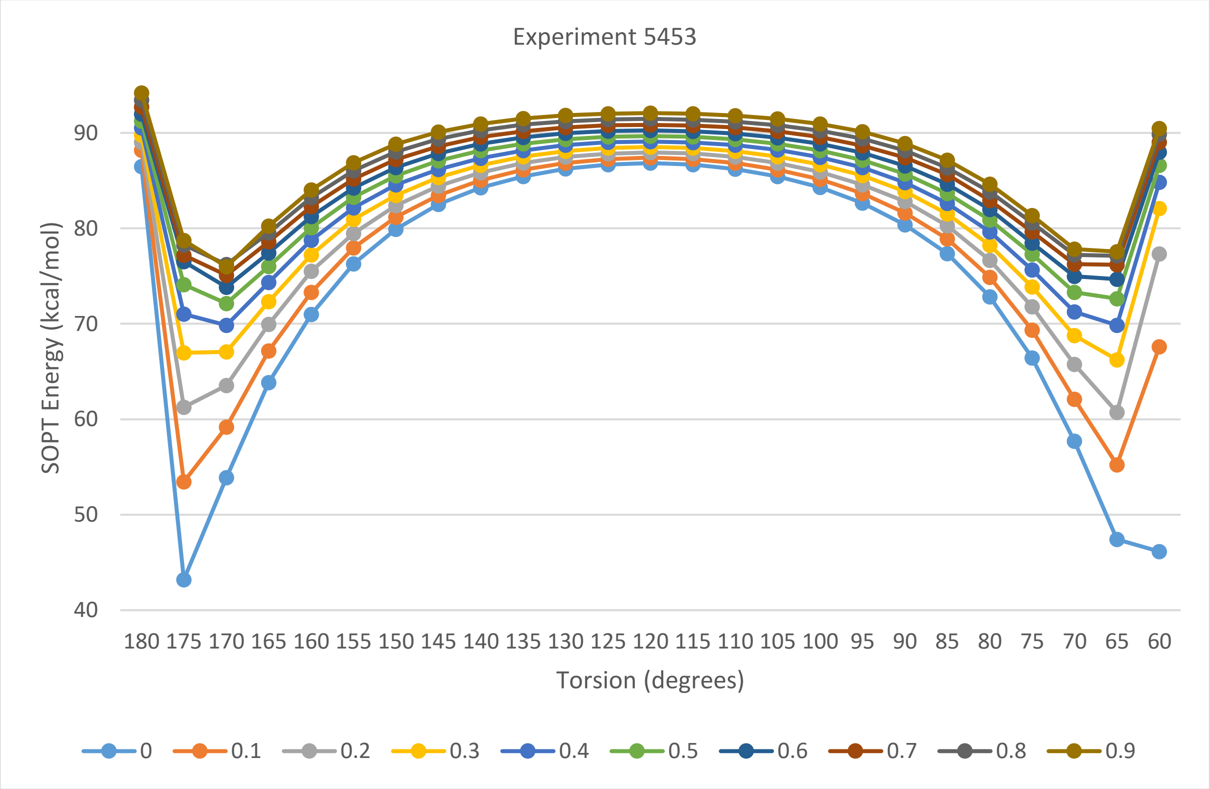


Figure 3.48. Ethanamide N(lp)->C-O(p)* SOPT Energy with LC-PBEhPBE Omega=0.2 and Varying O-C-CA-H Torsion and HF Exchange Coefficient

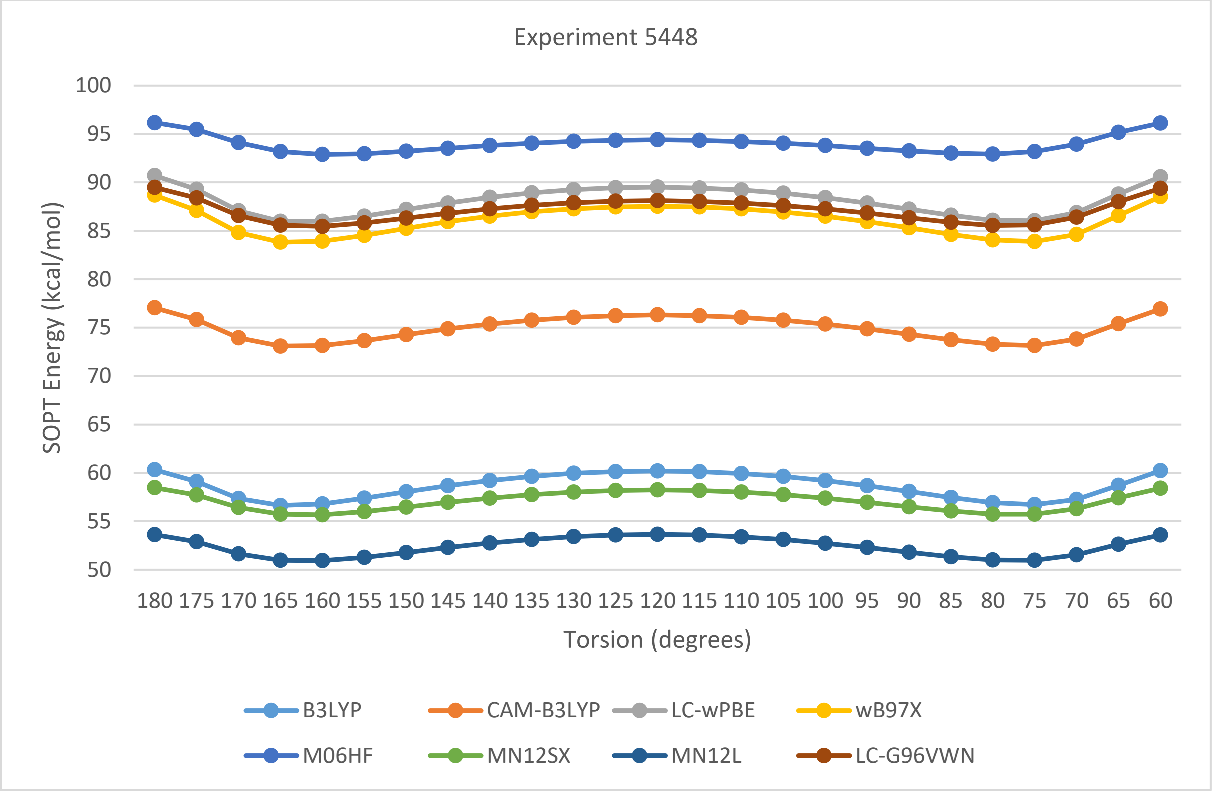


Figure 3.49. Ethanamide N(lp)->C-O(p)* SOPT Energy with def2-QZVPP UltraFineGrid over SCS-MP2/def2-QZVPP Optimized Geometry at Varying O-C-CA-H Torsion and Method

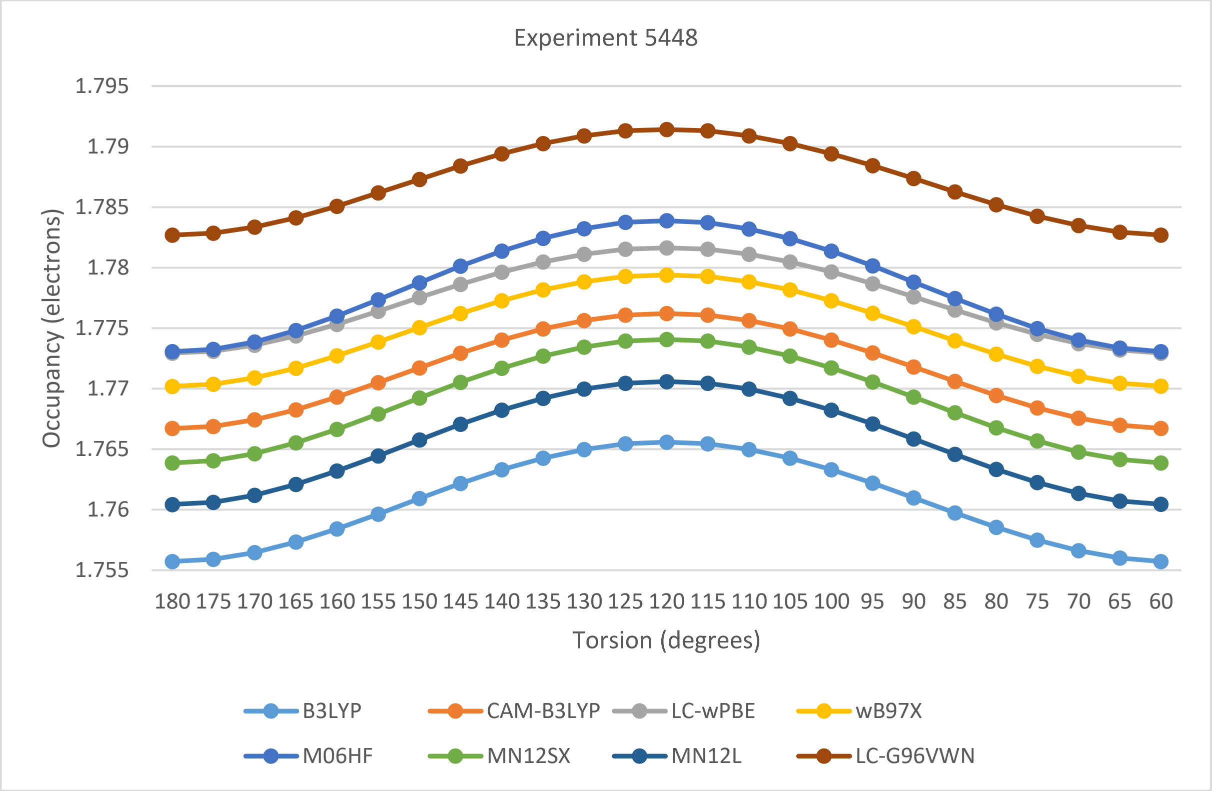


Figure 3.50. Ethanamide N(lp) Occupancy with def2-QZVPP UltraFineGrid and SCS-MP2/def2-QZVPP Optimized Geometry with Varying O-C-CA-H Torsion and Method

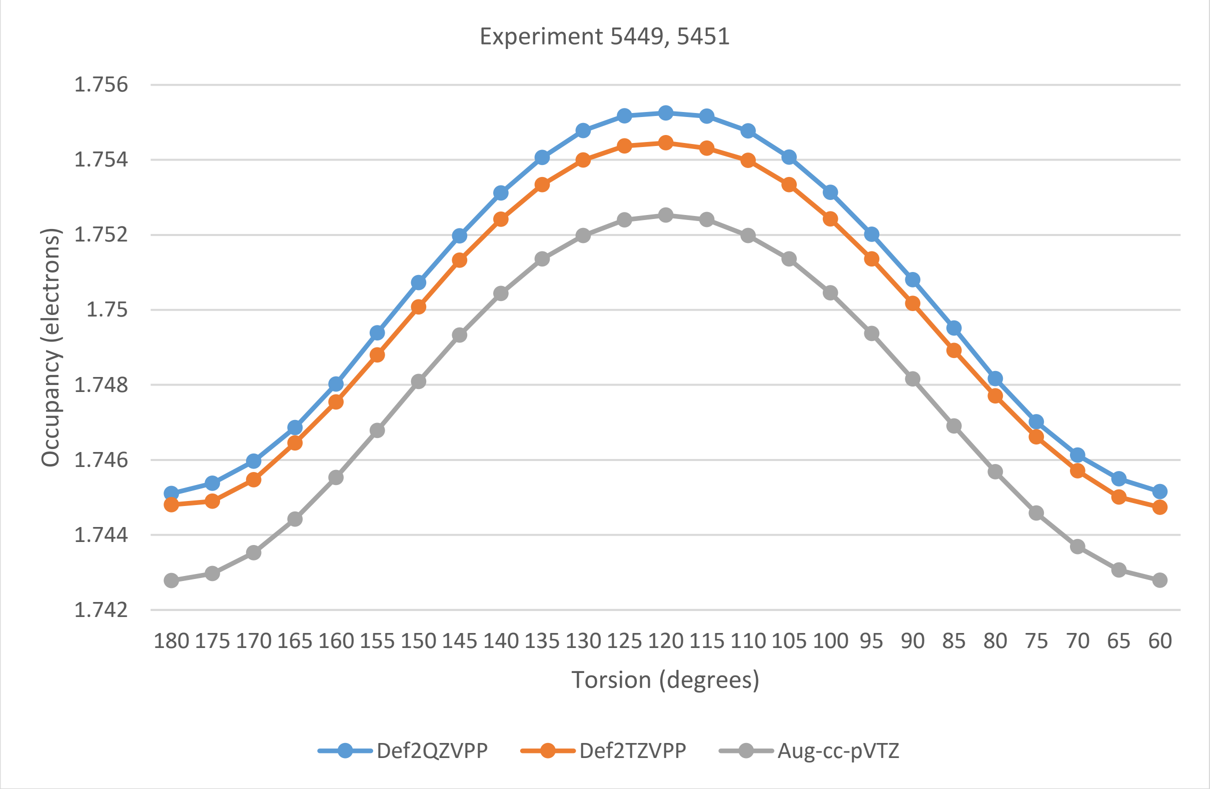


Figure 3.51. Ethanamide N(lp) Occupancy with DLPNO-CCSD(T) over SCS-MP2/def2-QZVPP Optimized Geometry with Varying O-C-CA-H Torsion and Basis Set

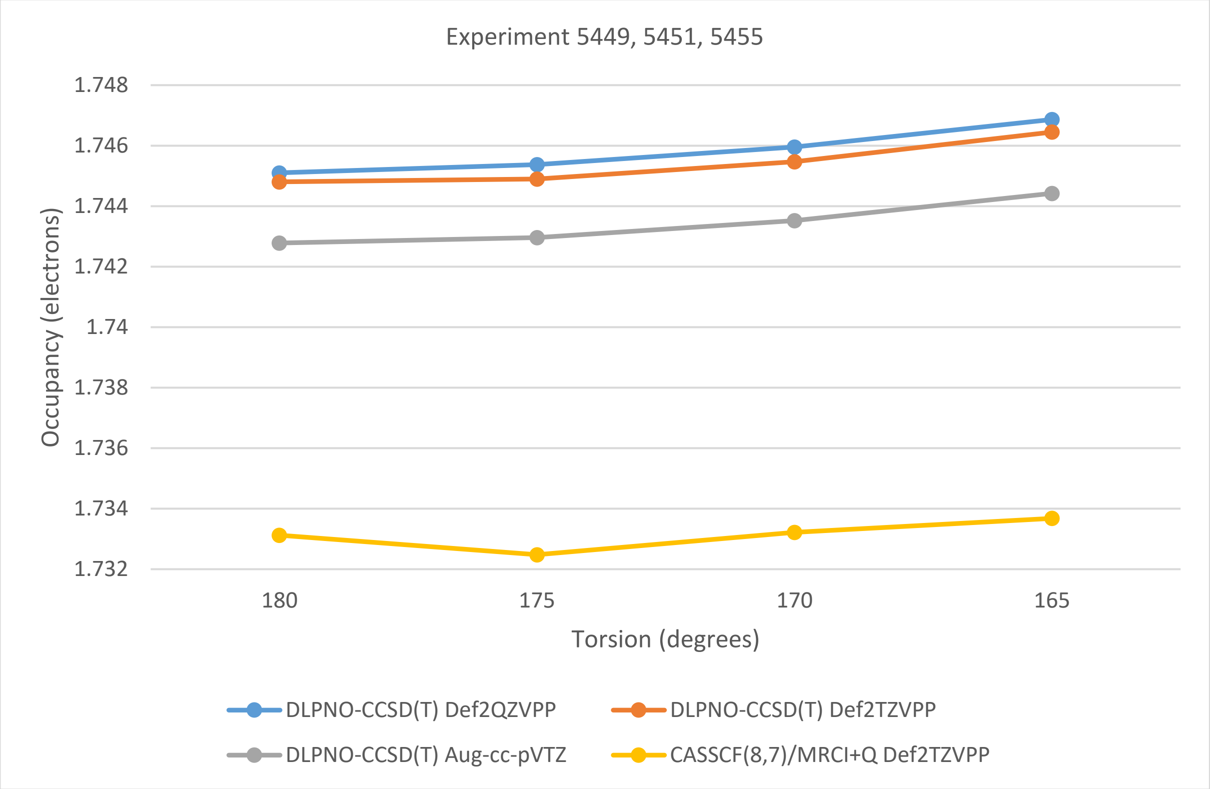


Figure 3.52. Ethanamide N(lp) Occupancy at SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion and Method/Basis Set

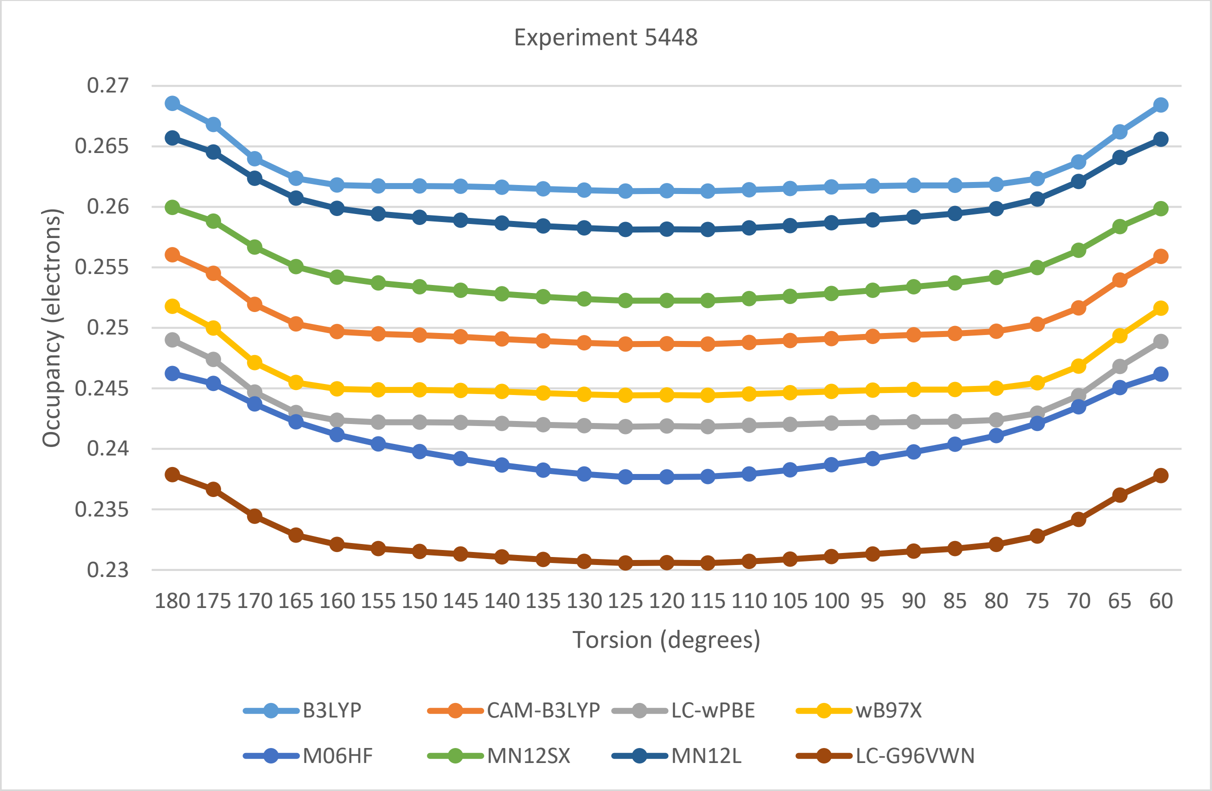


Figure 3.53. Ethanamide C-O(p)* Occupancy with def2-QZVPP UltraFineGrid and SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion and Method

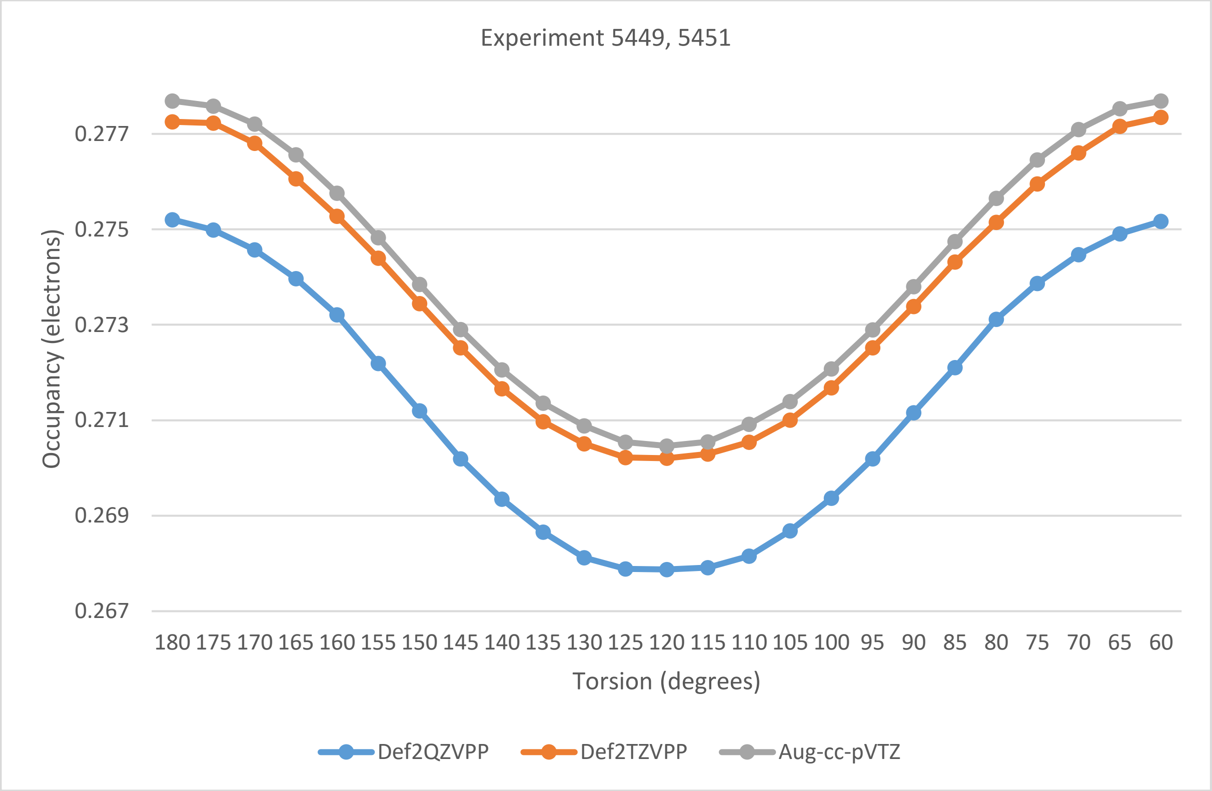


Figure 3.54. Ethanamide C-O(p)* Occupancy with DLPNO-CCSD(T) at SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion Basis Set

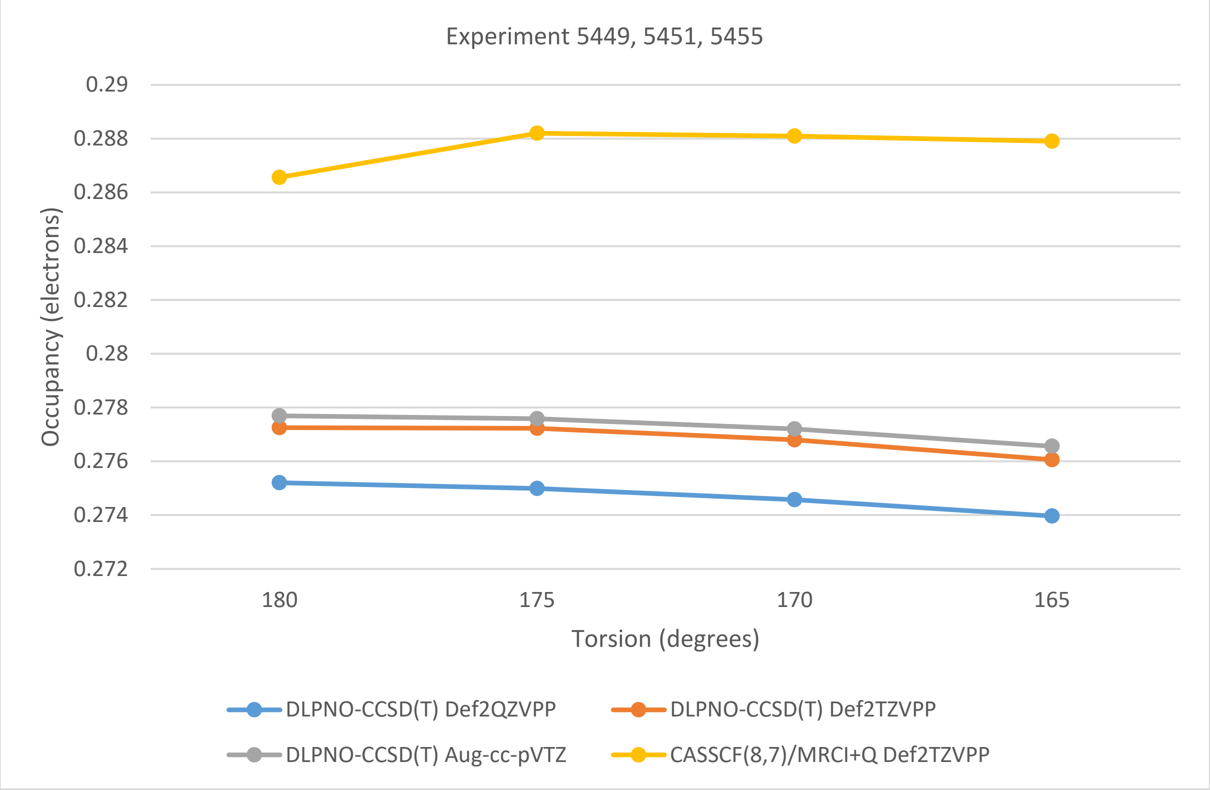


Figure 3.55. Ethanamide C-O(p)* Occupancy with SCS-MP2/def2-QZVPP Optimized Geometry and Varying O-C-CA-H Torsion and Method/Basis

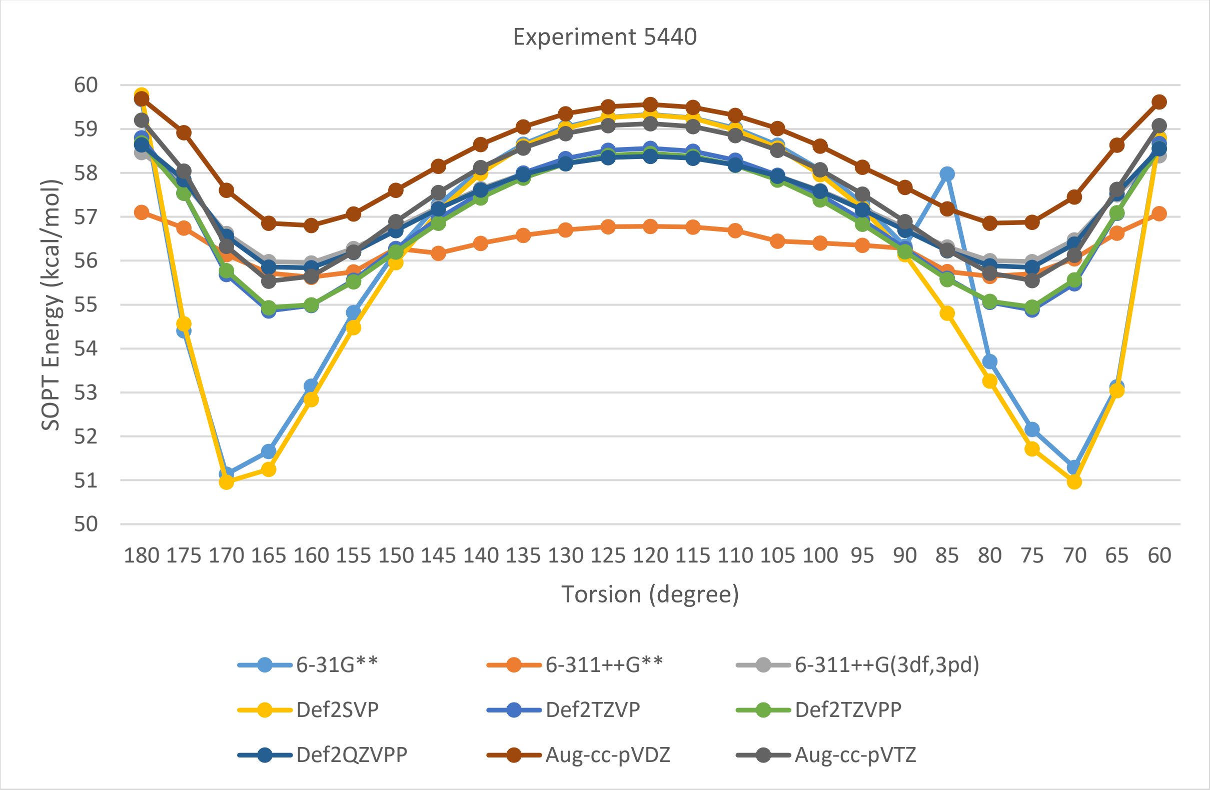


Figure 3.56. Ethanamide N(lp)->C-O(p)* SOPT Energy with MN12-SX UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

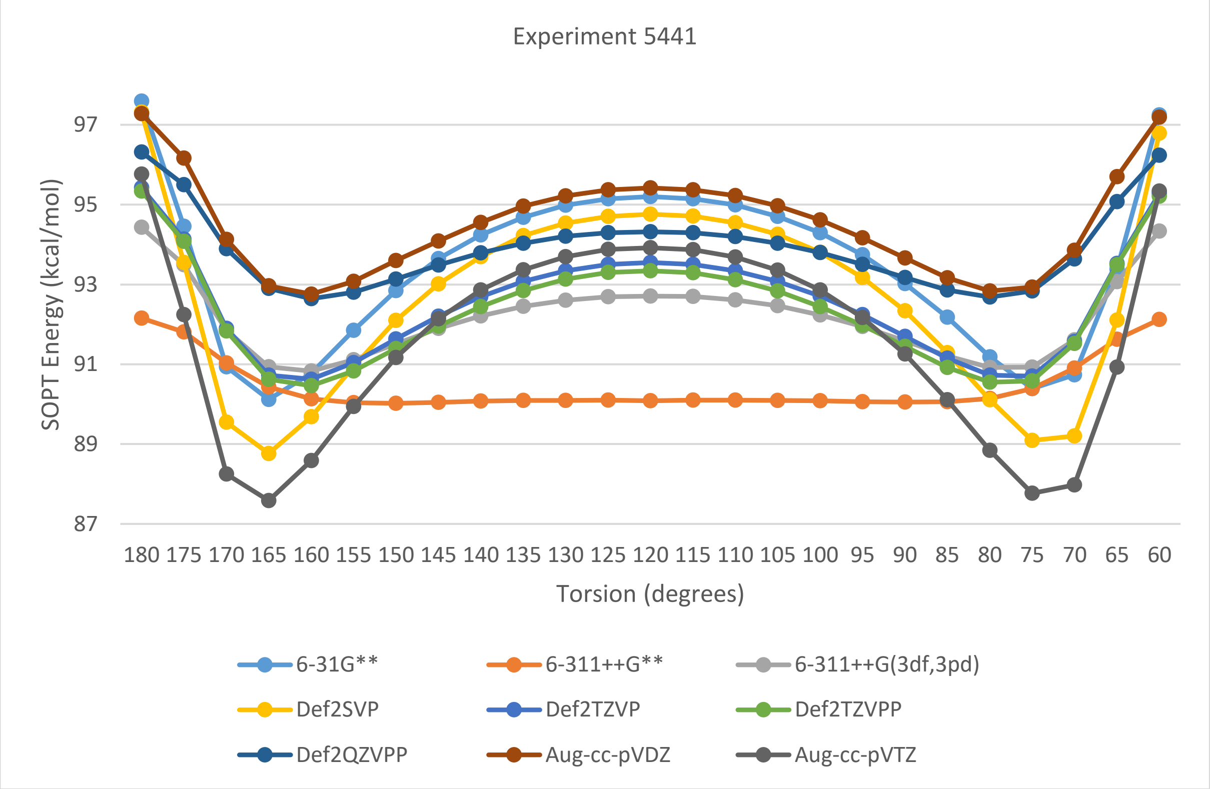


Figure 3.57. Ethanamide N(lp)->C-O(p)* SOPT Energy with M06HF UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

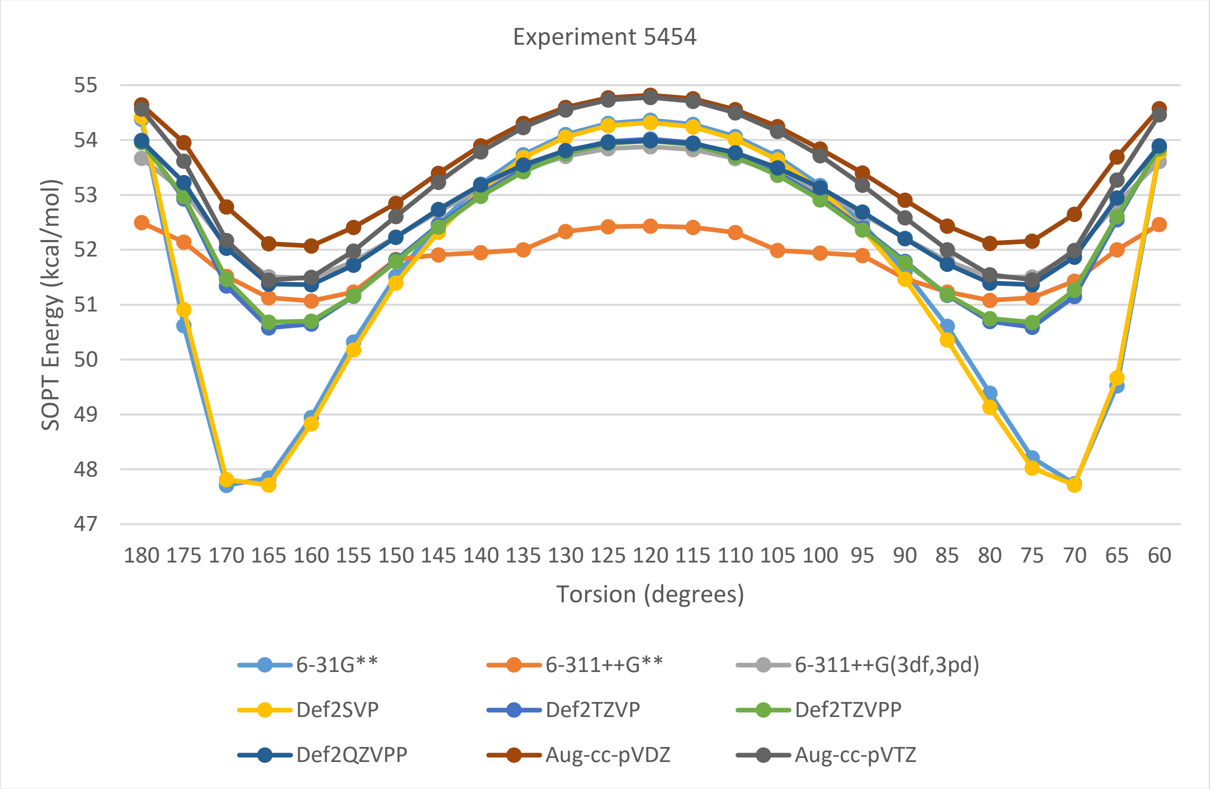


Figure 3.58. Ethanamide N(lp)->C-O(p)* SOPT Energy with MN12-L UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

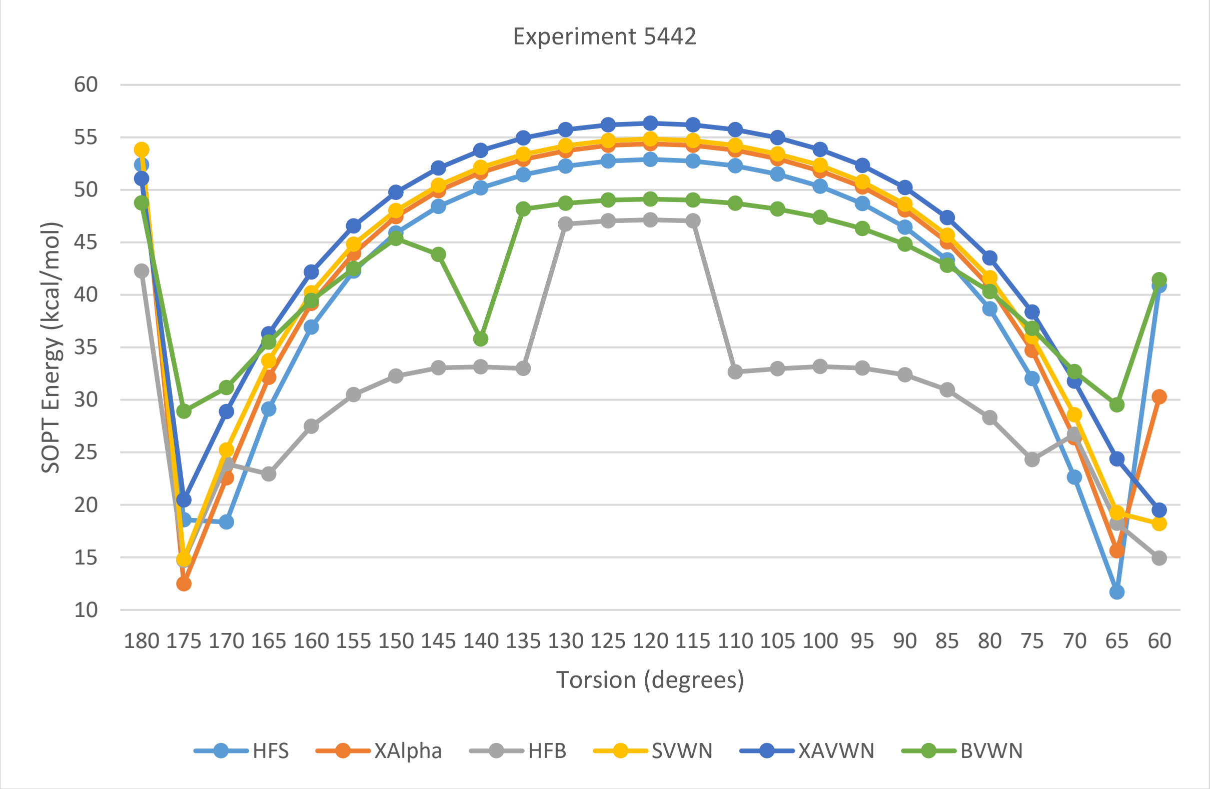


Figure 3.59. Ethanamide N(lp)->C-O(p)* SOPT Energy at 6-31G** UltraFineGrid at Varying O-C-CA-H Torsion and Method

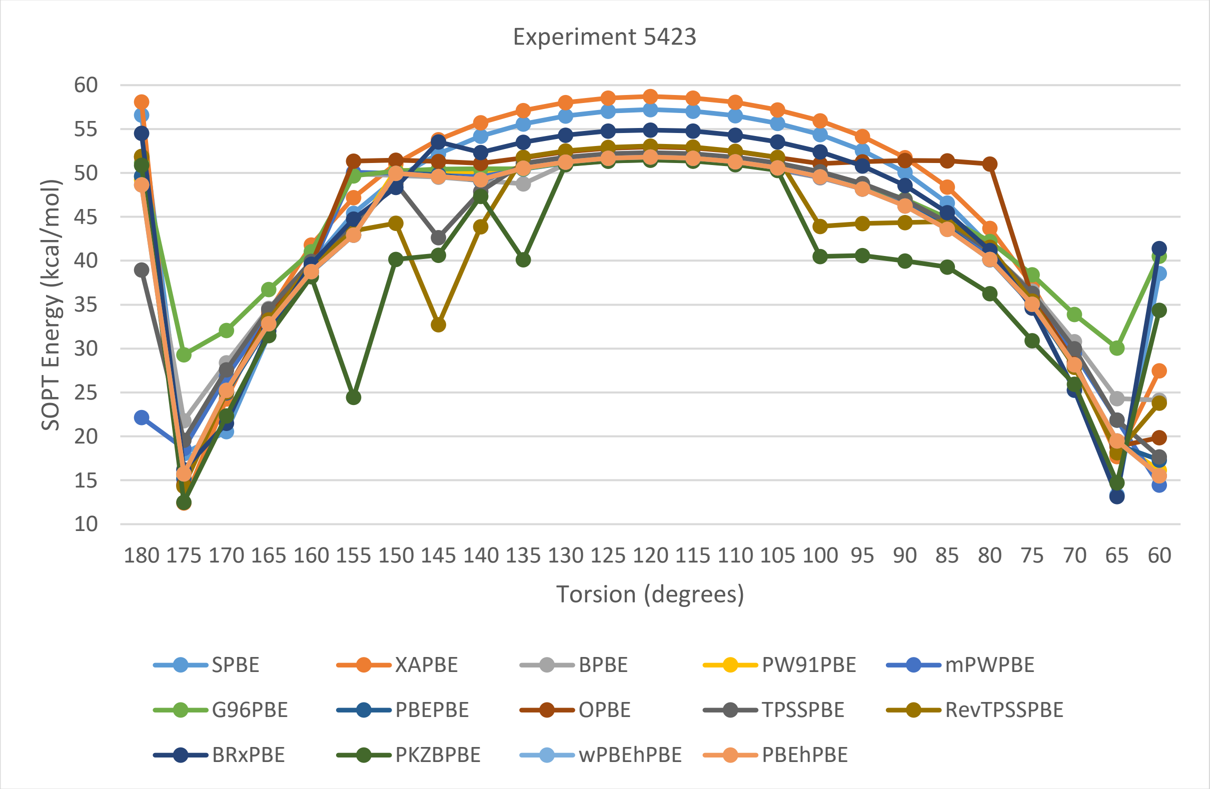


Figure 3.60. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** and UltraFineGrid and Varying O-C-CA-H Torsion and Method having PBE Correlation

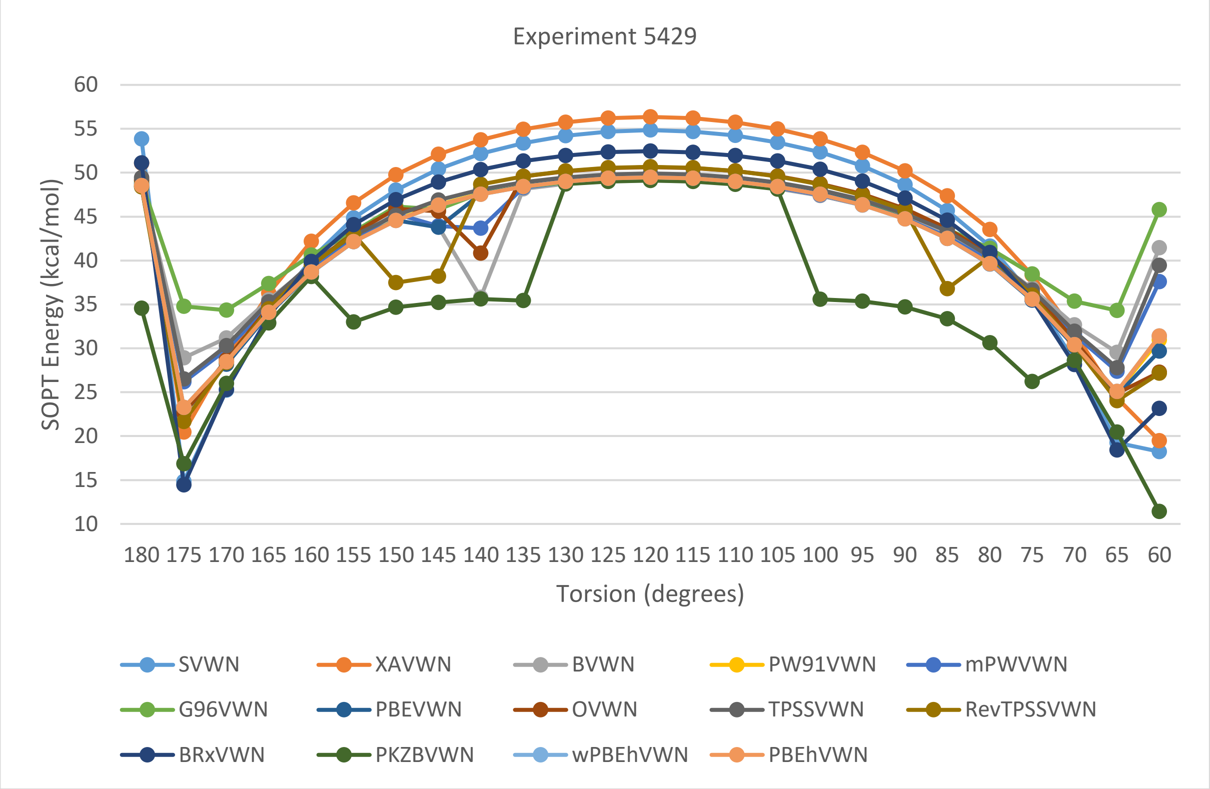


Figure 3.61. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method having VWN Correlation

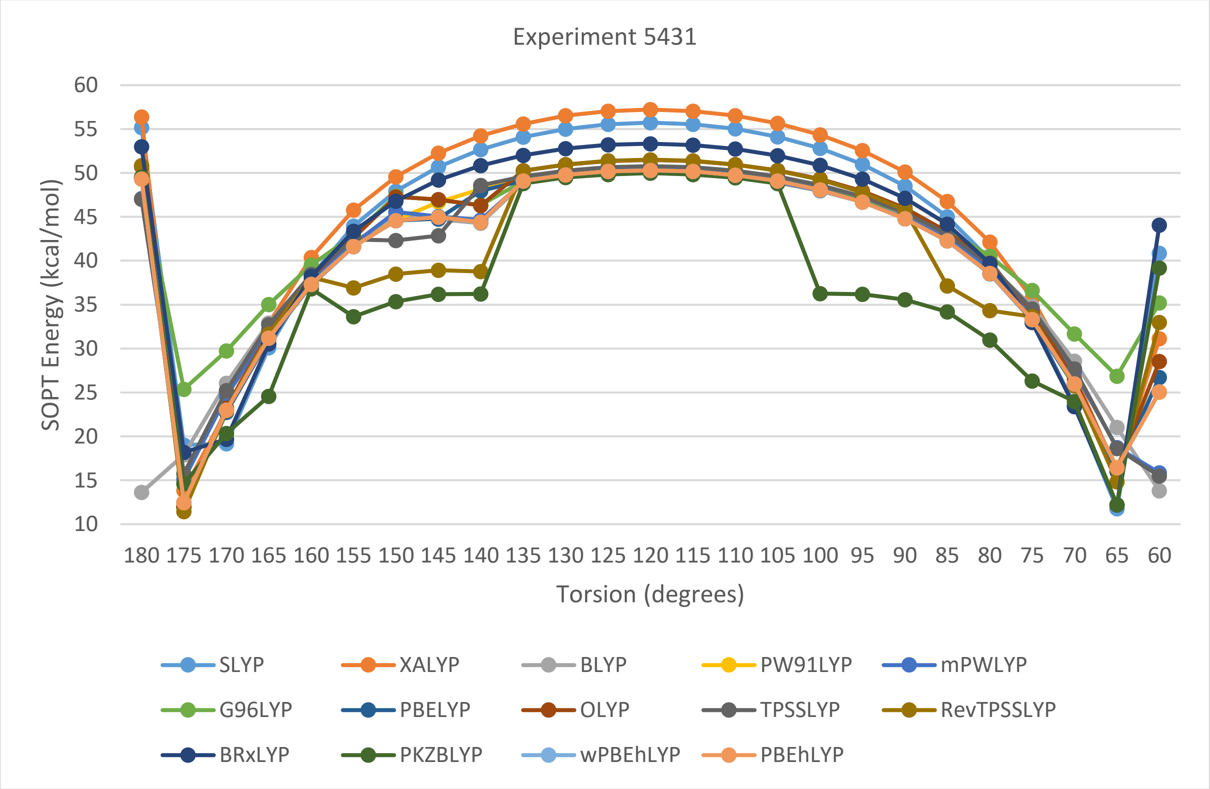


Figure 3.62. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method having LYP Correlation

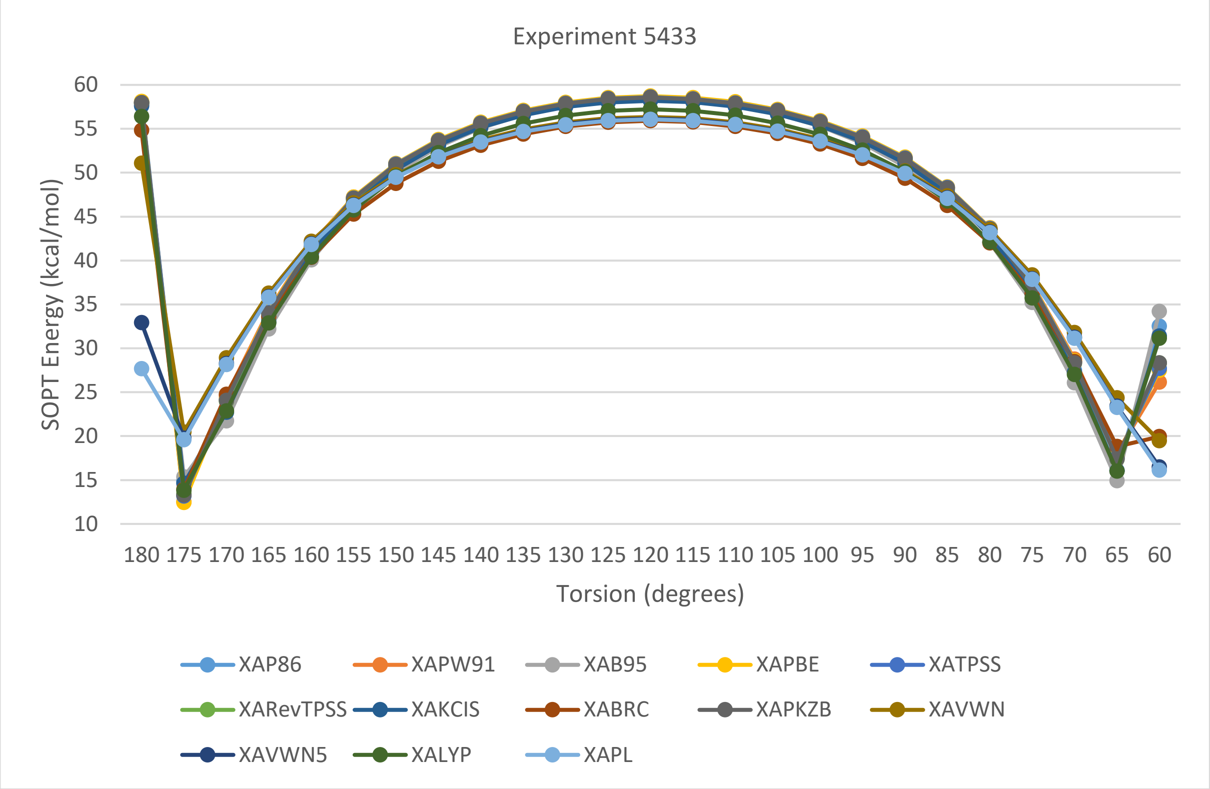


Figure 3.63. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method having XA Exchange

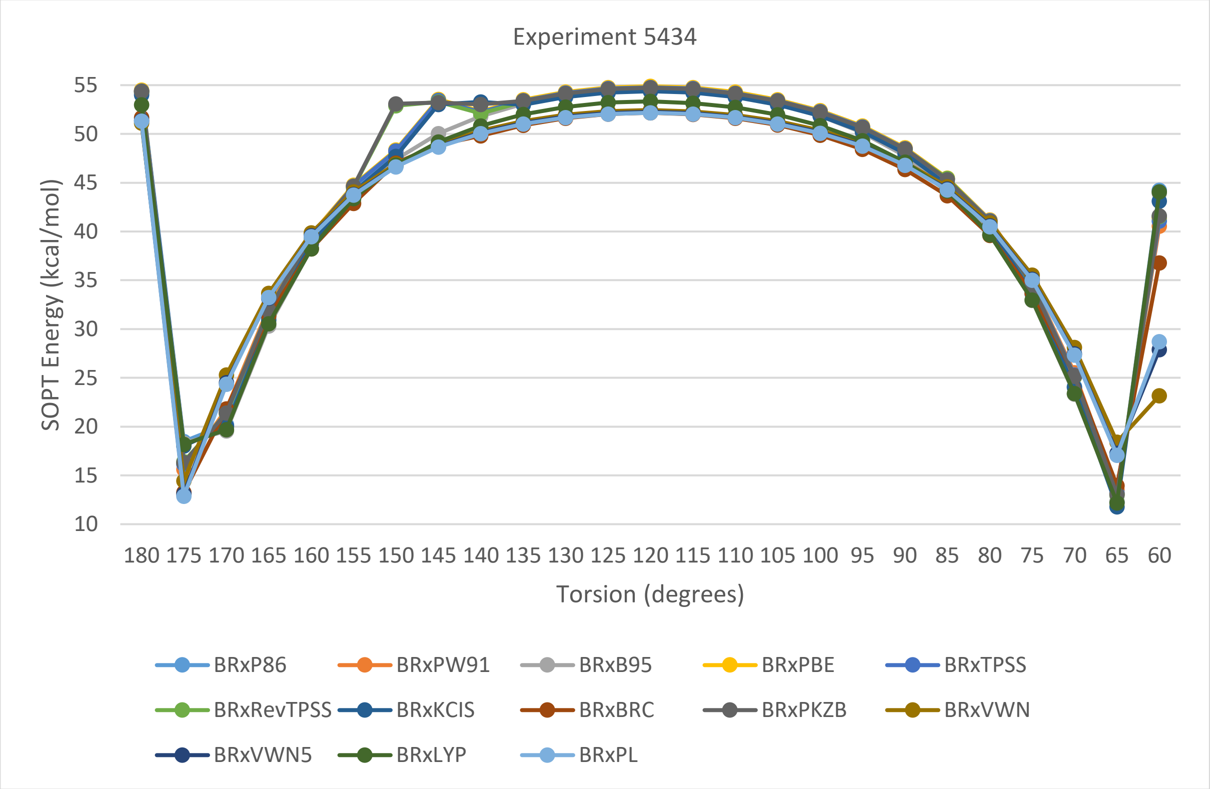


Figure 3.64. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method having BRx Exchange

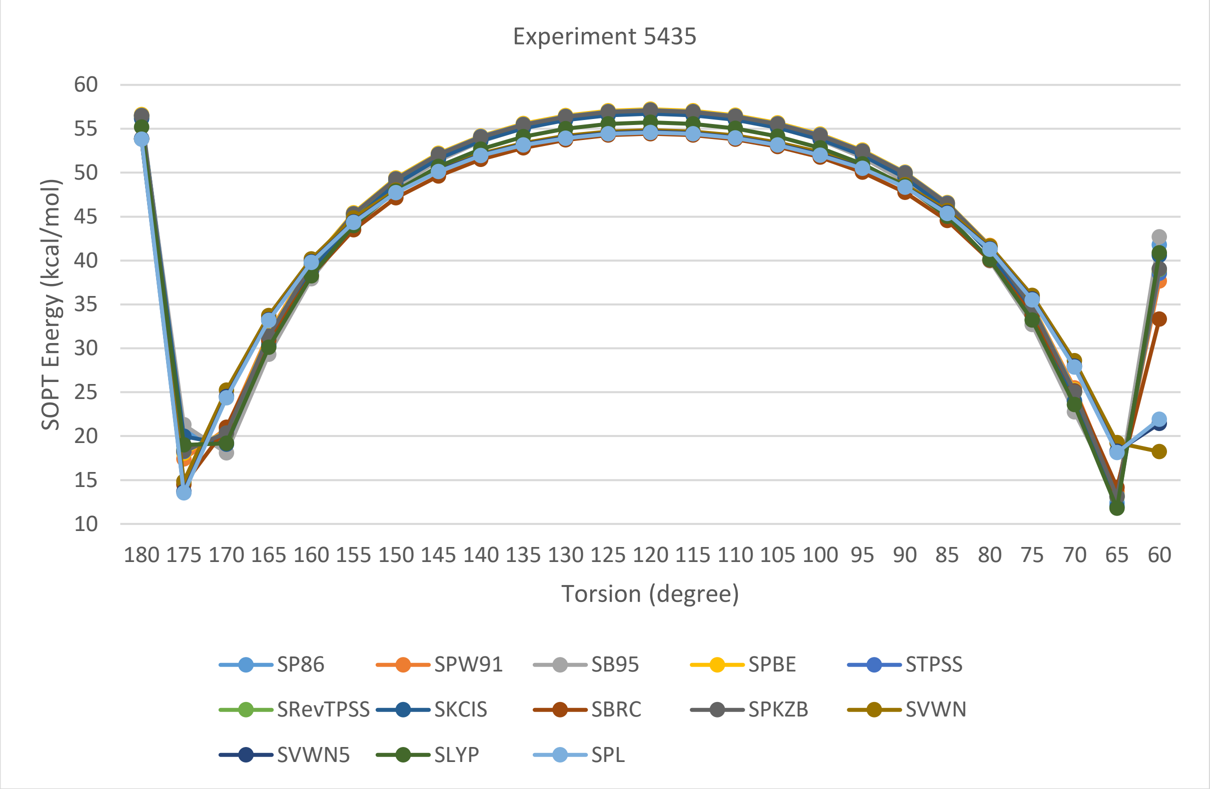


Figure 3.65. Ethanamide N(lp)->C-O(p)* SOPT Energy with 6-31G** UltraFineGrid and Varying O-C-CA-H Torsion and Method having Slater Exchange

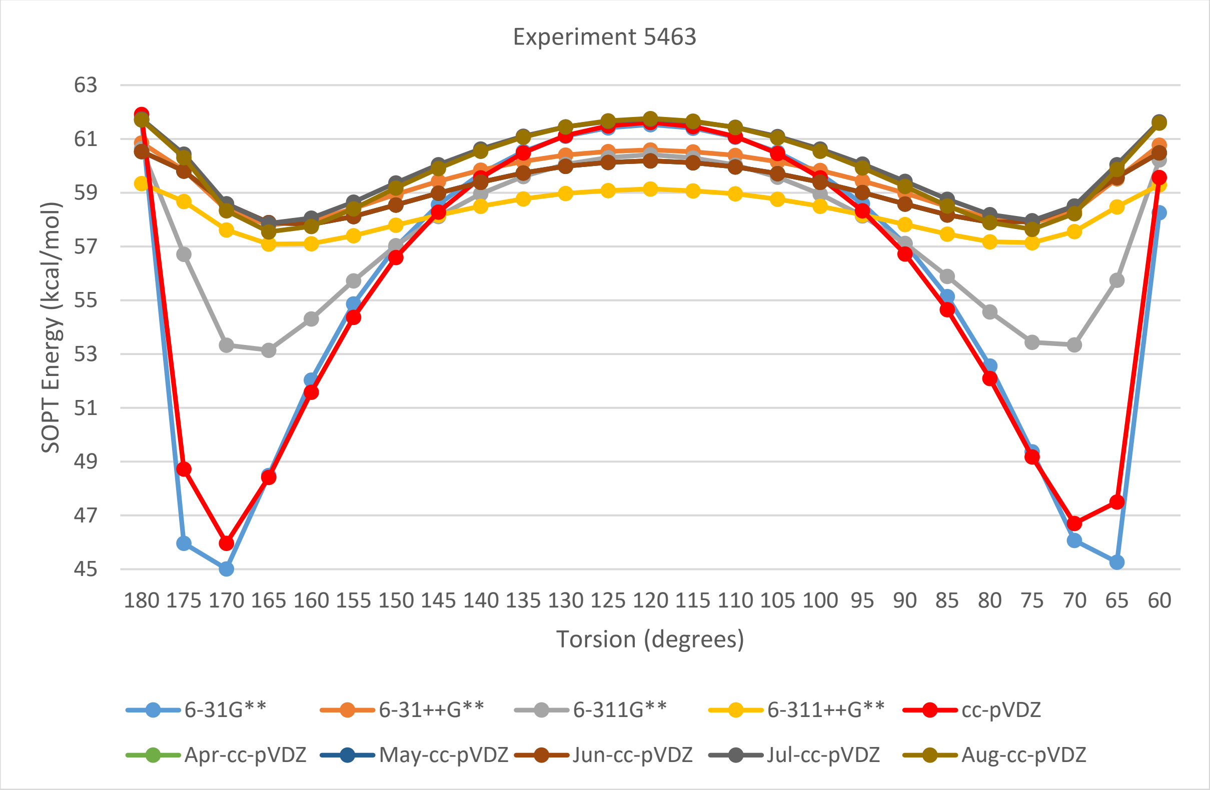


Figure 3.66. Ethanamide N(lp)->C-O(p)* SOPT Energy with B3LYP over SCS-MP2/def2-QZVPP Optimized Geometry, UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

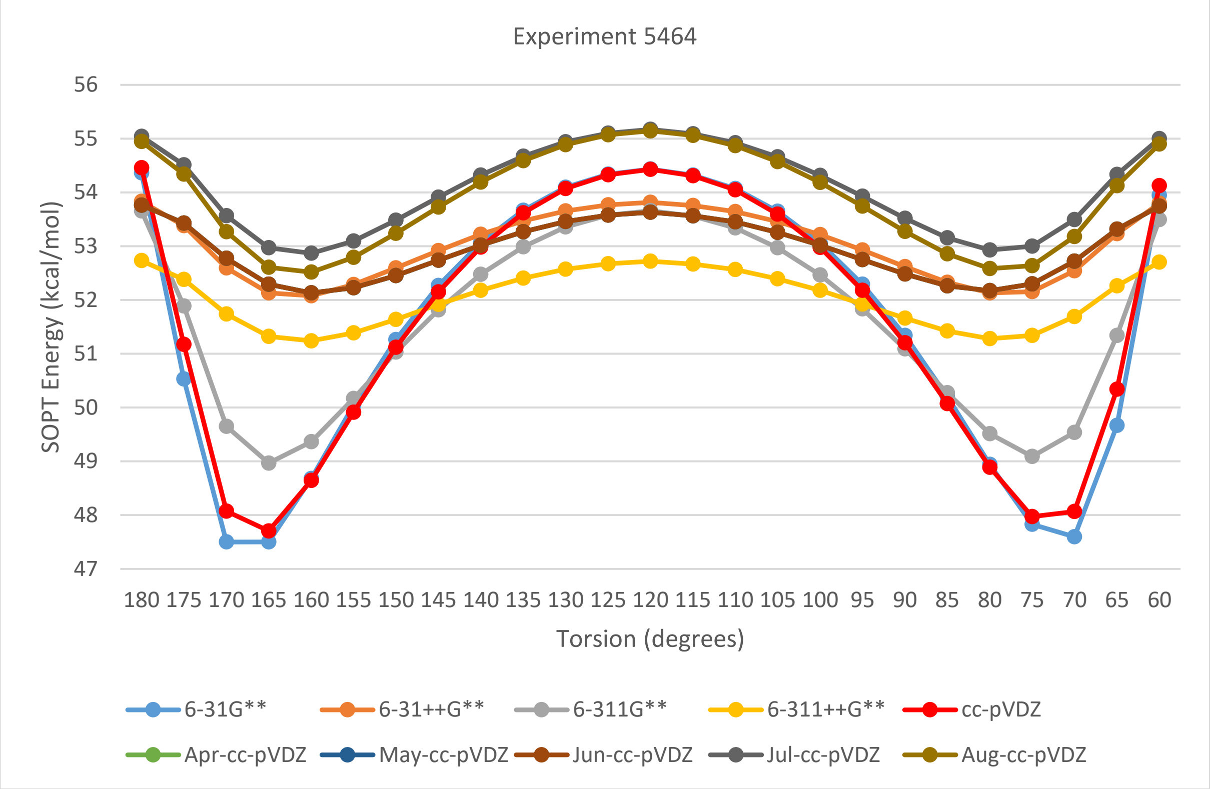


Figure 3.67. Ethanamide N(lp)->C-O(p)* SOPT Energy with MN12-L and SCS-MP2/def2-QZVPP Optimized Geometry, UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

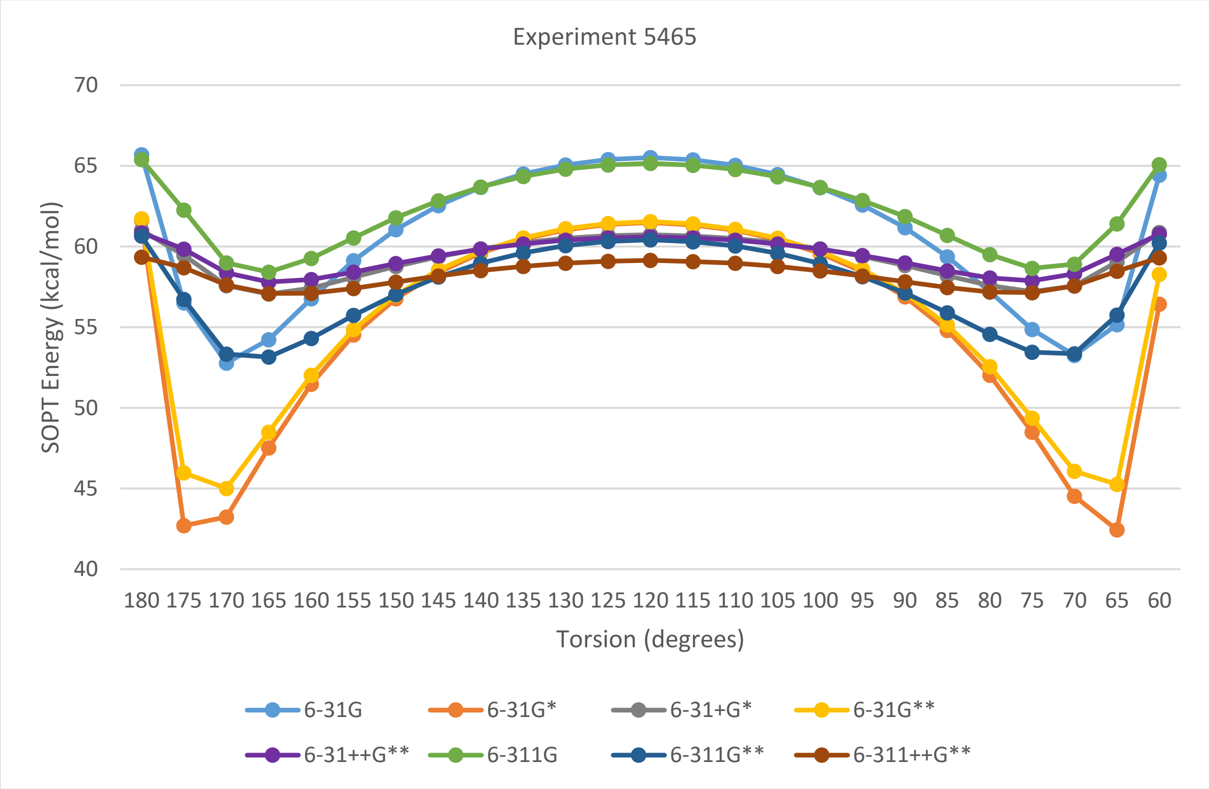


Figure 3.68. Ethanamide N(lp)->C-O(p)* SOPT Energy with B3LYP over SCS-MP2/def2-QZVPP Optimized Geometry, UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

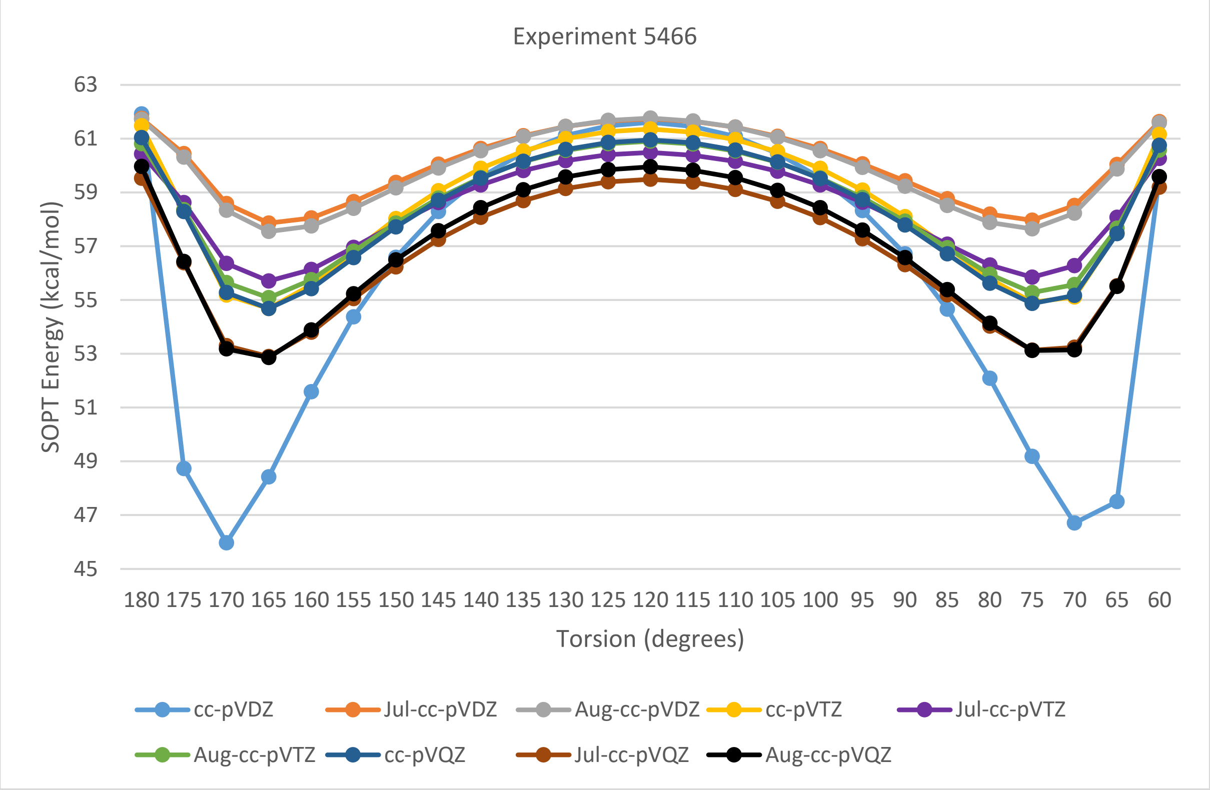


Figure 3.69. Ethanamide N(lp)->C-O(p)* SOPT Energy with B3LYP over SCS-MP2/def2-QZVPP Optimized Geometry, UltraFineGrid and Varying O-C-CA-H Torsion and Basis Set

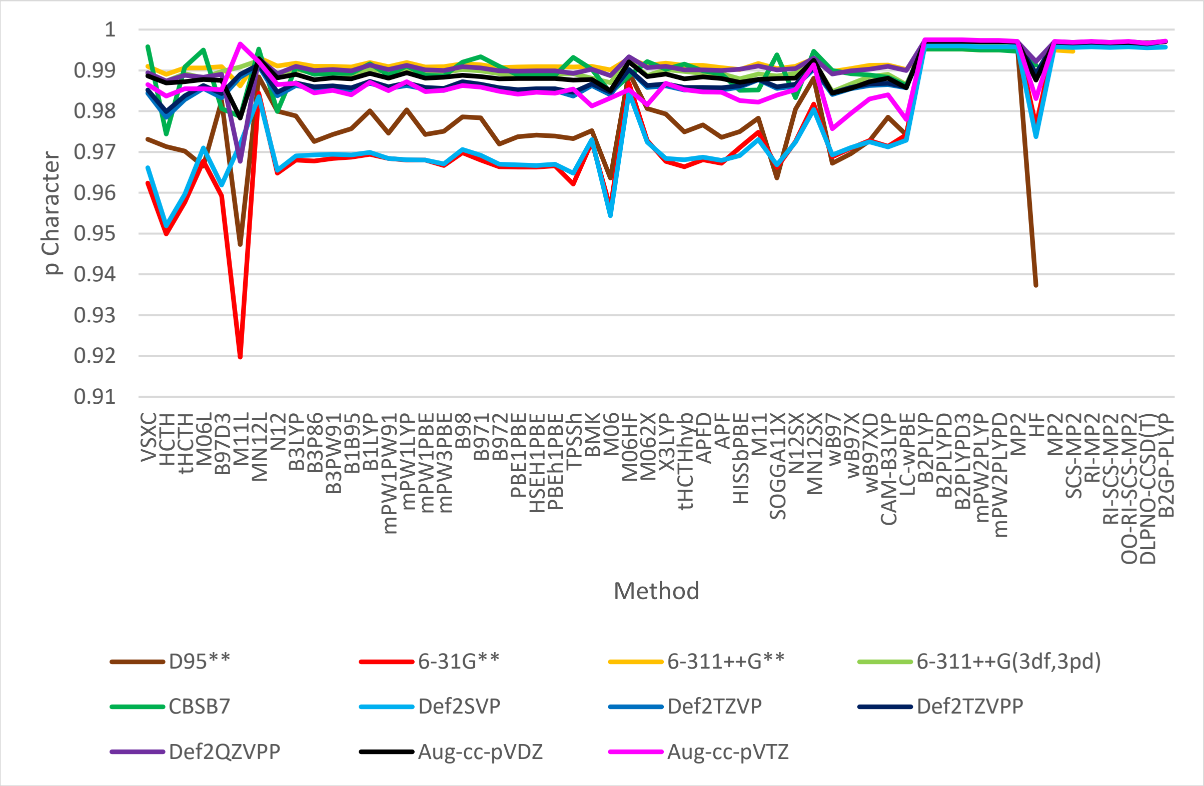


Figure 3.70. C-O(pi) Carbon NHO p Character in Ethanamide at Antiparallel Beta Strand Psi Torsion with Varying Method and Basis Set

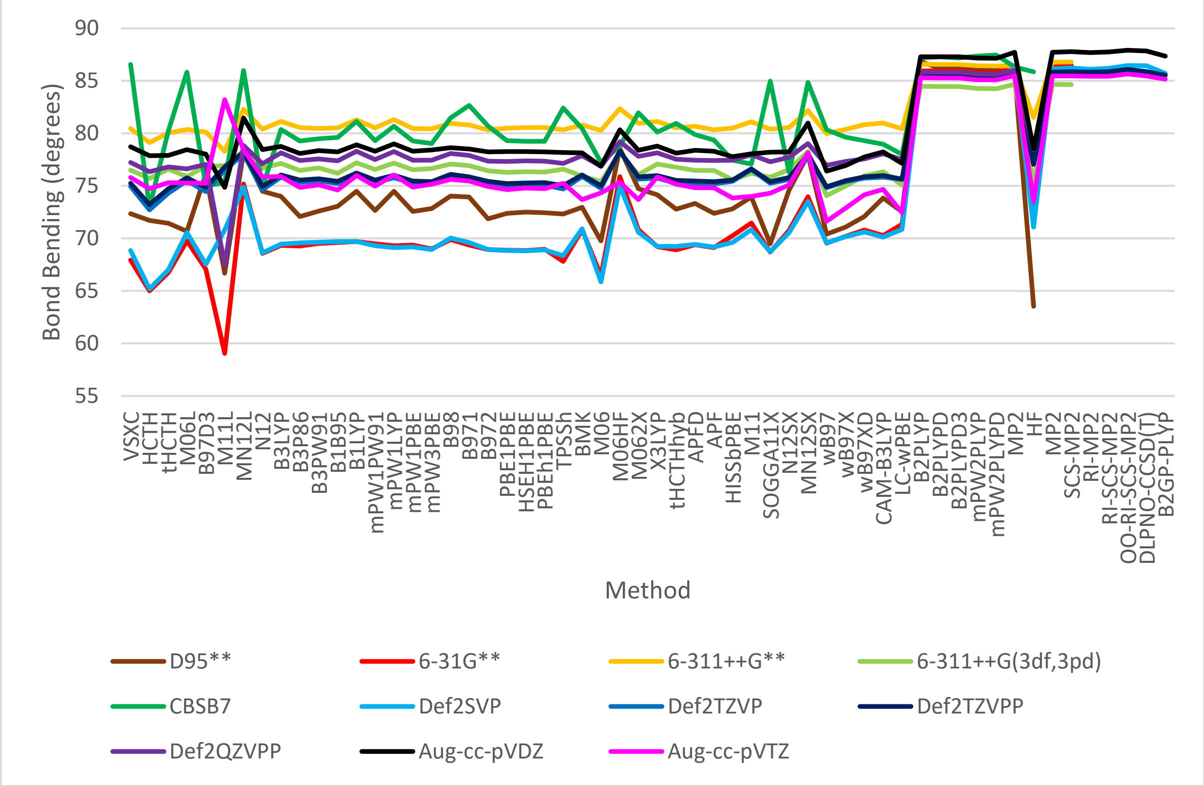


Figure 3.71. C-O(pi) Carbon NHO Bond Bending at Antiparallel Beta Strand Psi Torsion with Varying Method and Basis Set

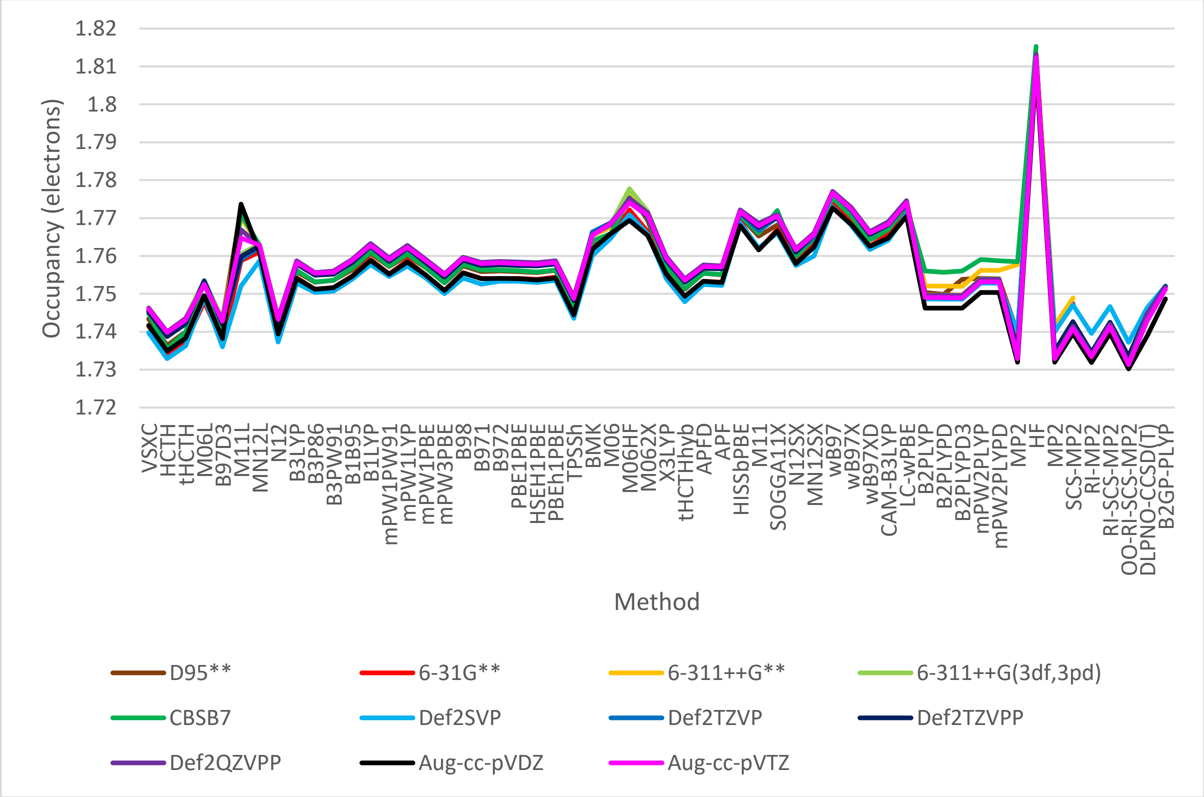


Figure 3.72. Nitrogen Lone Pair NBO Occupancy in Ethanamide at Antiparallel Beta Strand Psi Torsion with Varying Method and Basis Set

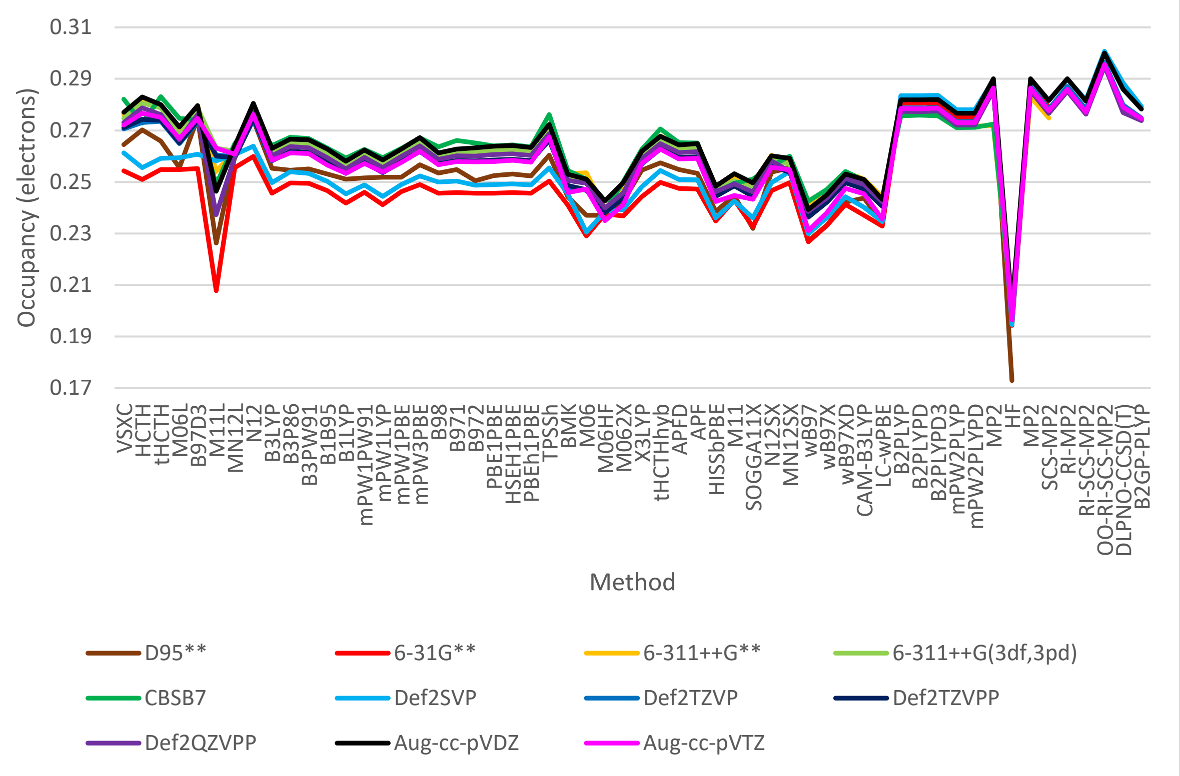


Figure 3.73. C-O(pi)* NBO Occupancy in Ethanamide at Antiparallel Beta Strand Psi Torsion with Varying Method and Basis Set

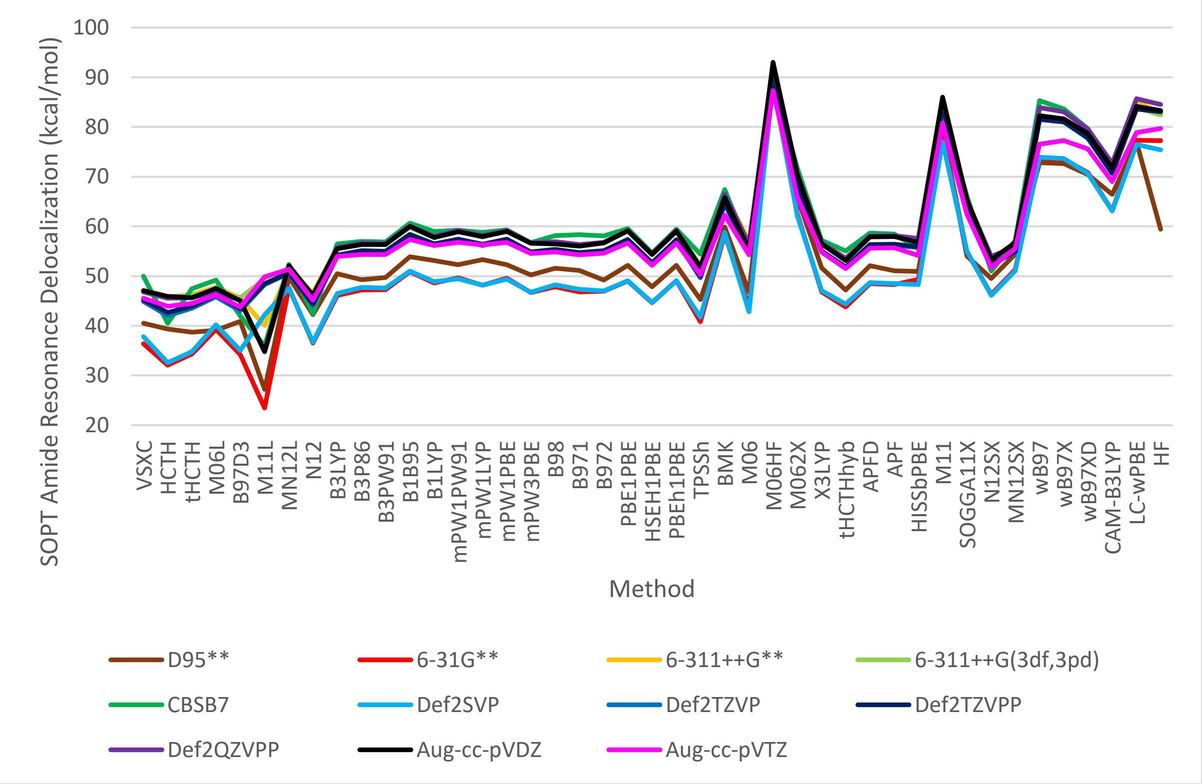


Figure 3.74. Second-Order Perturbation Theory Amide Resonance Delocalization (N(LP)->C-O(pi)*) in Ethanamide at Antiparallel Beta Strand Psi Torsion and Varying Method and Basis Set

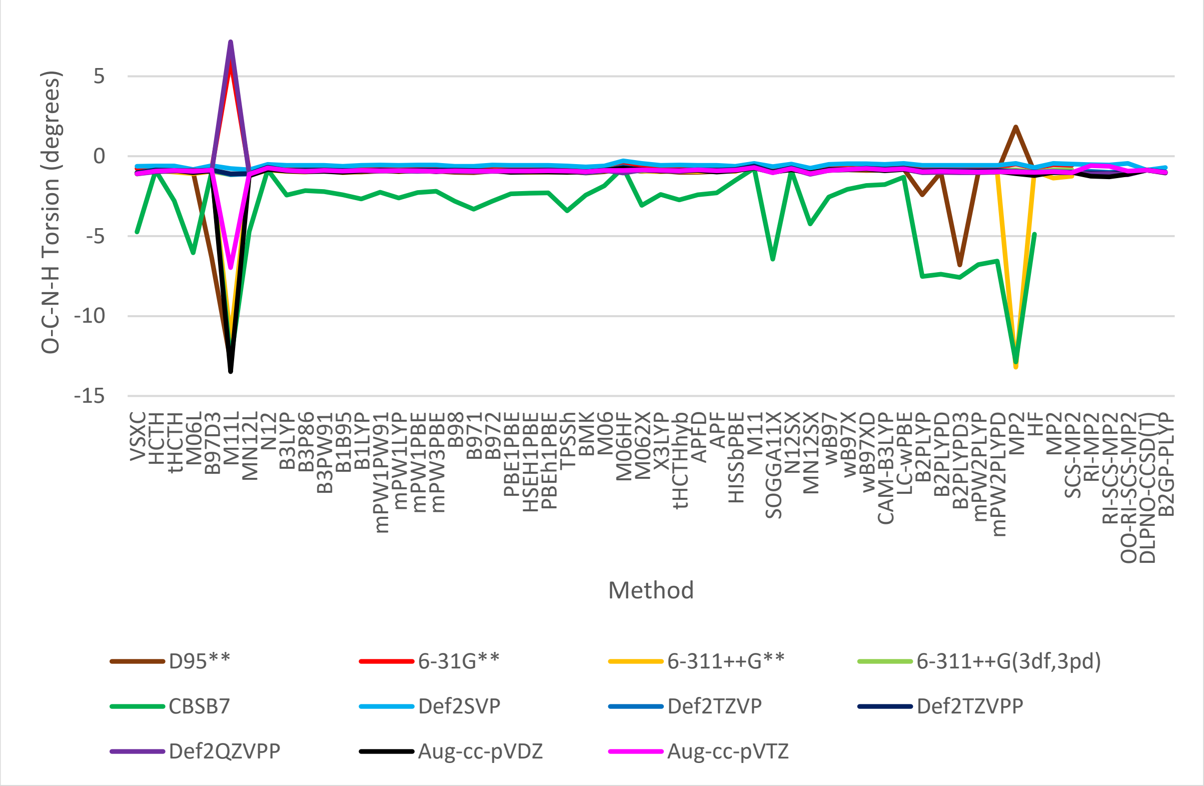


Figure 3.75. O-C-N-H Torsion in Ethanamide at Antiparallel Beta Strand Psi Torsion with Varying Method and Basis Set

3.12 Appendix 2

Table 3.1. CASSCF(8,7)/MRCI+Q TZVPP Ethanamide Symmetry and Bond Bending at Beta Sheet Psi Torsions

Beta Psi Torsion	Energy at Ground	Largest Config Weight	Second Config Weight	p Character of C-O(p) C NHO	Bond Bending of C-O(p) C NHO
Parallel	-208.862366	0.8435	0.013	0.9968	86.2
Antiparallel	-208.871637	0.8387	0.0186	0.9968	86.2

Table 3.2. CASSCF(8,7)/MRCI+Q TZVPP Ethanamide N(lp) Occupancy at Varying O-C-CA-H Torsion

O-C-CA-H deg	Energy at Ground	Largest Config Weight	Second Config Weight	N(lp) Occ
180	-208.868559	0.8372	0.0183	1.73312
175	-208.870524	0.839	0.0181	1.73248
170	-208.871203	0.8397	0.0181	1.73322
165	-208.871525	0.8401	0.018	1.73368

Table 3.3. Dihedral Angles (degrees) in Ethanamide with O-C-CA-H Constrained to -165 at MP2/6-311++G** with CP Fragment Boundaries at Bond Shown

Fragment	O-C-N-H1	O-C-N-H2
none	-163.359	-13.248
C-C	-165.651	-11.899
C-N	178.312	-0.373

Column Names for Table 3.4 and Table 3.5

s: s character of C-O(pi)* Carbon NHO

p: p character of C-O(pi)* Carbon NHO

d: d character of C-O(pi)* Carbon NHO

Dev: bond bending of C-O(pi)* Carbon NHO (degrees)

Dih: O-C-CA-HA dihedral angle

PolP: polarization of C-O(pi) NBO

PolS: polarization of C-O(sigma) NBO

CoapOcc: occupancy of C-O(pi)* NBO

CoasOcc: occupancy of C-O(sigma)* NBO

SoptP: N(LP)->C-O(pi)* SOPT (kcal/mol)

SoptS: N(LP)->C-O(sigma)* SOPT (kcal/mol)

CaHaCos: CA-HA->C-O(sigma)* SOPT (kcal/mol)

CaCbCop: CA-CB->C-O(pi)* SOPT (kcal/mol)

CaNCop: CA-N->C-O(pi)* SOPT (kcal/mol)

CopCaN: C-O(pi)>Ca-N* SOPT (kcal/mol)

Cid: identity of Carbon atom central to backbone amide group

Strand: beta strand in sheet

Chain: cross-strand hydrogen bonding chain

Table 3.4. Polyvaline Antiparallel Beta Sheet Backbone Amide Resonance Delocalization, Bond Bending and Loss of sigma/pi Symmetry at LC-wPBE/6-31G**

s	p	d	Dev	Dih	PolP	PolS	CoasOcc	CoapOcc	SoptP	SoptS	CaHaCos	CaCbCop	CaNCop	CopCaN	Cid	Strand	Chain
0.0452	0.9514	0.0034	65.33	179.3	0.2698	0.3348	0.058	0.2841	88.32	7.33	5.16	2.98			80	2	1
0.0436	0.953	0.0034	65.66	-180	0.2693	0.335	0.0566	0.2854	89.32	7.05	5.23	3.12			158	3	4
0.0226	0.9742	0.0032	70.86	158.1	0.2895	0.3465	0.0347	0.2583	85.52	4.28	3.71				2	1	4
0.0214	0.9754	0.0032	71.29	158.7	0.2891	0.3469	0.0333	0.2603	87.24	3.85	3.8	1.03			236	4	1
0.0105	0.9859	0.0036	76.24	-165.6	0.2651	0.3424	0.0241	0.3	110.91	1.36	5.38	4.42	2.01		252	4	1
0.0096	0.9869	0.0036	76.78	-166	0.2644	0.3427	0.0229	0.3013	111.94	1.18	5.38	4.42	2.01		18	1	3
0.0034	0.993	0.0036	82.1	-171.8	0.2699	0.3453	0.0177	0.2961	112.24		4.83	4.1	2.22	1.05	50	1	1
0.0023	0.9942	0.0036	83.4	-170.6	0.2695	0.3454	0.0166	0.2968	113.23		4.83	4.27	2.16	1.01	284	4	4
0.0009	0.9957	0.0034	83.84	-167.5	0.2821	0.3504	0.0149	0.2901	111.74		4.53	4.41	1.85		268	4	3
0.0006	0.996	0.0034	84.45	-168.4	0.2821	0.3505	0.0145	0.2907	112.24		4.57	4.34	1.96		34	1	2
0.0005	0.9958	0.0037	85.07	-164.9	0.2586	0.3443	0.0166	0.3258	128.3		4.76	5.03	1.62		96	2	2
0.0002	0.9961	0.0037	86.16	-164.1	0.2582	0.3445	0.016	0.3265	128.93		4.81	5.06	1.6		174	3	3
0.0002	0.9961	0.0037	86.35	-165.3	0.2608	0.3447	0.0159	0.3235	127.29		4.8	5.17	1.59		190	3	2
0.0001	0.9962	0.0037	86.93	-165.7	0.2607	0.3447	0.0156	0.3235	127.48		4.84	5.21	1.58		112	2	3
0.0003	0.9961	0.0036	87.46	-173	0.2639	0.3455	0.0154	0.3234	128.47		5.07	4.2	2.1	1.12	128	2	4
0.0002	0.9962	0.0036	87.84	-173.6	0.2639	0.3456	0.0153	0.3233	128.45		5.05	4.13	2.17	1.14	206	3	1

Table 3.5. Polyalanine Antiparallel Beta Sheet Backbone Amide Resonance Delocalization, Bond Bending and Loss of sigma/pi Symmetry at LC-wPBE/6-31G**

s	p	d	Dev	Dih	PolP	PolS	CoasOcc	CoapOcc	SoptP	SoptS	CaHaCos	CaCbCop	CopCaCb	Cid	Strand	Chain
0.0215	0.9753	0.0032	71.5	-169.16	0.2922	0.3476	0.0344	0.2602	86.01	4.07	4.71			64	1	6
0.0158	0.981	0.0032	73.75	-171.07	0.2905	0.3488	0.0288	0.2662	91.11	2.91	4.94			168	4	1
0.0139	0.9826	0.0035	73.78	-151.02	0.2784	0.3433	0.0292	0.2744	96.89	3.13	3.42	4.72	1.13	133	4	6
0.0124	0.9841	0.0035	74.47	-149.44	0.2785	0.3437	0.0275	0.2753	98.12	2.79	3.36	4.74	1.16	26	1	1
0.0131	0.9833	0.0036	74.73	-175.2	0.2615	0.3428	0.0275	0.3139	113.01	1.71	5.75			48	2	1
0.0099	0.9865	0.0036	76.42	-173.57	0.2607	0.3434	0.0242	0.3163	115.78	1.19	5.75			184	3	6
0.0076	0.9888	0.0036	78.23	-134.89	0.2674	0.3453	0.0204	0.302	112.74	1.36	4.24	6.38	1.66	123	4	4
0.0057	0.9906	0.0036	79.42	-134.25	0.2665	0.3458	0.0187	0.3047	115.6		4.08	6.2	1.64	32	1	3
0.0039	0.9926	0.0035	79.8	-135.81	0.2817	0.3503	0.018	0.2859	106.46	1.54	2.67	4.34	1.44	128	4	5
0.0032	0.9933	0.0035	80.37	-137.2	0.2813	0.3506	0.0172	0.2872	108.25	1.24	2.8	4.59	1.45	21	1	2
0.0025	0.994	0.0035	81.11	-138.34	0.2815	0.3511	0.0164	0.2887	110.49		2.98	4.71	1.49	55	1	4
0.0025	0.9941	0.0035	81.26	-137.36	0.2815	0.351	0.0162	0.2887	110.48		2.91	4.63	1.49	170	4	3
0.0035	0.9928	0.0037	81.37	-141.14	0.2623	0.3454	0.0169	0.308	118.71		4.55	6.59	1.57	163	4	2
0.0034	0.993	0.0037	81.59	-143.92	0.262	0.3453	0.0169	0.3083	119.06		4.68	6.63	1.54	59	1	5
0.0017	0.9946	0.0037	83.76	-145.73	0.261	0.3458	0.0153	0.3218	127.37		4.37	6.14	1.5	50	2	3
0.0013	0.995	0.0037	84.27	-144.83	0.2611	0.3459	0.015	0.3217	127.45		4.3	6.07	1.51	175	3	4
0.0012	0.9951	0.0037	84.47	-143.88	0.262	0.3466	0.0149	0.3209	126.95		4.25	6.21	1.53	152	3	3
0.0008	0.9955	0.0037	85.21	-144.36	0.262	0.3465	0.0147	0.3213	127.31		4.21	6.13	1.51	4	2	4
0.0007	0.9955	0.0037	85.4	-145.35	0.257	0.3453	0.0151	0.3258	129.51		4.37	6.17	1.46	179	3	5
0.0006	0.9956	0.0037	85.55	-143.09	0.2576	0.3456	0.015	0.3248	128.93		4.26	6.08	1.49	43	2	2
0.0004	0.9958	0.0038	86.19	-147.33	0.2578	0.3455	0.0147	0.3286	130.93		4.27	6.05	1.42	141	3	2
0	0.9964	0.0036	86.58	-150.1	0.2696	0.347	0.0145	0.3103	122.22		4.34	5.89	1.43	14	2	6
0.0003	0.9961	0.0036	86.67	-150.96	0.2696	0.3472	0.0143	0.3108	122.79		4.47	5.98	1.44	146	3	1
0.0001	0.9974	0.0025	88.02	-146.83	0.7422	0.6546	0.0147	0.3285	130.83		4.26	6.05	1.43	9	2	5

4 Variation of protein backbone amide resonance by electrostatic field

4.1 Statement of authorship

Title of Paper	Variation of protein backbone amide resonance by electrostatic field	
Publication Status	☐ Published ☐ Submitted for Publication	☐ Accepted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	arXiv:1512.05488 First version appeared on arXiv (arxiv.org) on 2015-12-17.	

Name of Principal Author (Candidate)	John Neville Sharley		
Contribution to the Paper	all		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	2016-09-26

4.2 Author contact

john.sharley@pobox.com

4.3 Abstract

Amide resonance is found to be sensitive to electrostatic field with component parallel or antiparallel to the amide C-N bond, an effect we refer to here as EVPR-CN. EVPR-CN is linear and without threshold in the biologically plausible electrostatic field range -0.005 to 0.005 au. Variation of amide resonance varies Resonance-Assisted Hydrogen Bonding [4], RAHB, such as occurs in the hydrogen bonded chains of backbone amides of protein secondary structures such as beta sheets [7] and non-polyproline helices [11] such as alpha helices [8], varying the stability of the secondary structure. The electrostatic properties including permittivity of amino acid residue sidegroups influence the electrostatic field component parallel or antiparallel to the C-N bond of each amide, giving a novel relationship between residue sequence and protein structure. Additionally, a backbone-based theory of protein folding [3] which includes this effect is presented in Section 4.10.4.

The significance of EVPR-CN relative to other factors in protein folding depends on field C-N component at each backbone amide at a given time. Calculation indicates that backbone amides do not occupy an intrinsically electrostatically-protected niche. We propose that EVPR-CN warrants investigation in any study of stable protein structure or protein folding pathway [30]. EVPR-CN is somewhat associated with hydrophobia [141], since hydrophobia creates low permittivity environments. EVPR-CN is more directionally and hence structurally specific than hydrophobia.

Hypotheses concerning the stability of beta sheets and amyloid fibrils [10] and of protein complexation and molecular chaperone function [13] are offered. An analogous effect in nitrogenous base pairing [12] is proposed.

EVPR-CN is energetically significant in biologically plausible electrostatic fields even without considering a hydrogen bonding context, and a hypothesis concerning the stability of polyproline helices types I and II is offered.

4.4 Key phrases

- protein folding
- conformational change
- molecular chaperone
- secondary structure
- beta sheet
- helix
- polyproline
- amide resonance
- electrostatic

- resonance assisted hydrogen bond
- backbone torsion
- base pairing
- stark effect

4.5 Notation

“->” denotes Natural Bond Orbital [6], NBO, resonance-type charge transfer and “|” denotes NBO steric exchange repulsion. “(“ and “)” enclose specification of an orbital type and follow an atom name for single-center NBOs and a pair of atom names separated by “-“ for two-center NBOs.

Examples: N(lp) for the amide nitrogen lone pair NBO, O(lp-p) for the oxygen p-type lone pair NBO, O(lp-s) for the s-rich lone pair NBO, C-O(p)* for the pi carbonyl antibonding orbital NBO and N(lp)->C-O(p)* for the primary amide resonance type charge transfer.

4.6 Overview

The cooperativity of hydrogen-bonded chains of protein backbone amides depends on the resonance at each amide in the chain. This is an example of Resonance-Assisted Hydrogen Bonding, RAHB. We investigate variation of amide resonance and RAHB due to electrostatic field.

Resonance-type charge transfer from the nitrogen lone pair NBO to the carbonyl pi antibonding NBO is primary to amide resonance. An NBO charge transfer variation of 0.001 electrons is nominated as the threshold of chemical significance, having associated energy of ~0.6 kcal/mol (p104-105 of [80]).

A striking change in amide resonance occurs when an electrostatic field is applied parallel or antiparallel to the amide C-N vector, an effect referred to here as EVPR-CN. Using two monovalent ions to create an electrostatic field, we find variation in the primary amide charge transfer of 0.2 electrons, some 200 times the nominated level of chemical significance, and note that this is not the full extent of what is physically possible. Using a uniform electrostatic field, we find that the variation in charge transfer is linear with the field magnitude in the biologically plausible range of -0.005 to 0.005 au. In this range the primary amide charge transfer in N-methylformamide varies by 0.042 electrons. An applied field magnitude of 0.000238 au varies the primary amide charge transfer by 0.001 electrons. We demonstrate changes to RAHB in a chain of N-methylformamides in the presence of an electrostatic field.

EVPR-CN is a previously unrecognized factor in protein folding. The significance of EVPR-CN depends on the C-N vector component of electrostatic field environmental to backbone amides.

The directional sensitivity of the resonance of the amide group to electrostatic field allows amino acid residue specification of protein structure through residue electrostatic properties including permittivity. We propose that EVPR-CN is central to the structure and function of proteins, and that evolution selected a polymer having an amide group in its hydrogen-bonded chains for this effect. The amide group is a remarkably simple molecular mechanism for introducing directional sensitivity to electrostatic field resulting in variation in hydrogen bonding cooperativity, and it may be that no simpler organic molecular mechanism exists for this.

Rose *et al*. [3] propose a backbone-based theory of protein folding in which “the energetics of backbone hydrogen bonding dominate protein folding” and backbone hydrogen bonding is a “universal folding mechanism”, but do not propose that this mechanism is directly varied by residue sidechain properties. Electrostatic variation of backbone amide resonance and hence RAHB depends on the electrostatic properties including permittivity of the sidechains in the residue sequence. In addition to EVPR-CN, we propose a backbone-based theory of protein folding in which the universal folding mechanism becomes peptide group (PG) resonance (PR) rather than backbone hydrogen bonding (Section 4.10.4).

We offer hypotheses concerning the stability of beta sheets, amyloid fibrils, polyproline helices of types I and II and of protein complexation and molecular chaperone action, and of an analogous effect in nitrogenous base pairing.

4.7 Introduction

In some protein secondary structures, such as beta sheets and non-polyproline helices, there exist hydrogen bonded chains of backbone amide groups. These distinctive protein hydrogen bonding patterns were discovered by Pauling *et al.* in 1951 [7, 8]. The hydrogen bonds in these chains are cooperative, with cooperativity mediated by the resonance of the backbone amides. The phenomenon of hydrogen bonding cooperatively depending on resonance is referred to as RAHB.

Any variation of backbone amide resonance is of relevance to hydrogen bonding in backbone amide RAHB chains and hence the stability of RAHB protein secondary structures, the appearance of secondary structures in proteins and the stability of the folded protein structure.

Variation in the primary amide resonance-type charge transfer, N(lp)->C-O(pi)*, is associated with a variation in energy which can be considerable and must be taken into account when considering protein structure stability, regardless of whether this variation is due to EVPR-CN, RAHB or any other source.

Variation of resonance by electrostatic field in general is known, but we are not aware of a report of quantification of this effect in amides or the peptide groups of polypeptide backbones, which is surprising since the resonance of amide or peptide groups is large. The general effect has been described by Shaik and co-workers and Coote and co-workers and investigated by these groups as a mean of catalysis [142-148]. Though electrostatic catalysis is used in enzymes [149-153], it is the least developed form of synthetic chemistry [146]. Karafiloglou observed that an electrostatic field has an important directionally-modulated on delocalization energy and hence the weights of alternative resonance structures even in a non-polar molecule, noting that the same principles must apply to other pi-systems [154]. The general effect appears to be in want of a name and we propose “Field-Dependent Resonance Mixing”. The underlying physics is the very familiar influence of electrostatic field on charge transfer [155]. Electrostatic field variation of amide resonance in particular is mentioned in lecture notes offering an account of the means by which hydrogen bonding to backbone amides changes amide resonance (i.e. the mechanism of RAHB), remarking that “The resonant state could be influenced by an external field in different ways, but the primary cause is hydrogen bonding” [156], though no quantification of the influence of external field on amide resonance is cited. As far as we are aware, no quantification of the variation in backbone amide resonance due to electrostatic field has been reported.

4.8 Software

Methods used in experiments are as implemented by Gaussian 09 D.01 [75], Orca 3.0.3 [72-74] and TeraChem 1.5K [91-94]. Unless otherwise stated, default grids and optimization and SCF convergence limits were used, except that the Orca option VeryTightSCF was used throughout as were cartesian coordinates for geometry optimization with TeraChem.

A development version of NBO [95] was used for its XML [59] output option. The XML was queried with XQuery 3.0 [65] or XSLT 3.0 [63] as implemented by Saxon-PE 9.6.0.4 [69], and the results imported into Excel 2013 [96].

Molecular coordinates are depicted by UCSF Chimera 1.10.2 [111].

4.9 Results and discussion

4.9.1 Constrained to O-C-N plane

We used N-methylethanamide to study amide resonance in the presence of hyperconjugative interactions with the amide carbonyl orbitals, and N-methylformamide otherwise. These molecules have greater amide resonance than formamide and are more representative of protein backbone amides.

With the atoms of N-methylformamide that are in the O-C-N plane in the absence of an electrostatic field constrained to that plane, and one of the methyl hydrogens constrained to this plane in cis configuration with the N-H bond, but all bond lengths and bond angles free to re-optimize, variation in amide resonance due to variation of electrostatic field direction and magnitude is calculated at SCS-MP2 [105] and MP2 [84] with the aug-cc-pVTZ [85] basis set. The term “X-Y vector” refers here to the normalized vector from the X nuclear center to the Y nuclear center. Figure 4.2 shows that in 360 degree rotation in 10 degree intervals in the amide O-C-N plane, the maximum increase in amide resonance as indicated by C-O(pi)* NBO occupancy occurs when the electrostatic field vector has the same direction as the C-N vector and maximum decrease in the opposite direction. Refining this rotation to 1 degree intervals in a range of -10 to 10 degrees from the C-N vector in the O-C-N plane, the peak is found at 0 or 1 degree (Ap1:Figure 4.17). Conducting a similar experiment with N-methylethanamide with a hydrogen of the extra methyl group constrained to O-C-N planarity in trans configuration to C-O, the peak is 5 degrees away from the C-N vector, passing through the C-O bond (Ap1:Figure 4.18). The substituent at the amide carbon makes a difference, but so little elevated is the peak at 5 degrees over the value at 0 degrees, being less than 0.0001 electron which is an order of magnitude lower than the nominated level of chemical significance, that we will take the C-N vector as being a good approximation to the field direction giving maximum effect.

Figure 4.3 shows that the vectors for maximum resonance increase and decrease in 360 degree rotation in the plane though the O-C-N plane normal and C-N are the same as occur in rotation in the O-C-N plane. Figure 4.4 shows the results of 360 degree rotation of the electrostatic field normal to the C-N vector about that vector which are that these field directions make little difference to amide resonance.

Available physically determined data for electrostatic variation in proteins [155, 157-161] give that the range of magnitudes without regard to direction is at least -0.005 to 0.005 au.

The variation of C-O(pi)* NBO occupancy with fields of C-N vector and range of magnitudes between minus 0.005 au and 0.005 au is shown in Figure 4.5, demonstrating a linear variation to C-O(pi)* NBO occupancy with field. DLPNO-CCSD(T) [74] with aug-cc-pVTZ and def2-TZVPP [137] basis sets was used over coordinates optimized at SCS-MP2/aug-cc-pVTZ, again showing linearity. DFT methods B3LYP [101], B97-D3 [94], BLYP [122, 162] and PBE0 [163] in combination with basis sets aug-cc-pVTZ, def2-TZVPP and 6-311++G** [129] all show this same linearity (Ap1:Figure 4.19), showing that DFT captures this effect well. Ap1:Figure 4.20 shows the linearity of the variation in C-O(pi)* NBO occupancy with variation of magnitude of N-C vector field in N-methylethanamide at SCS-MP2 and MP2/aug-cc-pVTZ.

The SCS-MP2 and MP2 data shown Figure 4.5 was obtained with N-C-O, N-C-H, C-N-H and C-N-CA angles free to re-optimize, and there is variation of these angles in the field. At 0.005 au, the N-C-O angle is 124.81 degrees, for the zero field case is 124.25 degrees and for -0.005 au is 123.68. SCS-MP2 and MP2 data was obtained again with these angles fixed to the zero field case, but the data was not discernibly different from that with angles free and is not shown.

We must state to forestall non-confirmatory report that we found an absence of concurrence between the uniform electrostatic field results calculated by Orca [72] and Gaussian [75]. In Gaussian, we used Field=Read and 6 decimal places. Uniform electrostatic field implementation was found by the Orca developers to be in error at Orca 3.0.2, and this was revised for 3.0.3. Because of the discrepancy between Gaussian and Orca, we also conduct computational experiments using ions, and then find that Orca and Gaussian results are in agreement and in accord with Orca's uniform electrostatic field results. We use Orca 3.0.3 and Gaussian 09 D.01 for these experiments, since these have the most robust NBO interfaces of the quantum chemistry packages available to us.

Figure 4.6, Ap1:Figure 4.23, Ap1:Figure 4.24 and Ap1:Figure 4.25 show the C-O(pi)* NBO occupancy resulting from constraining the ions Li+ or F- to the C-N line, constraining an ion to be either at distance from C further away from N or an ion at distance from N further away from C or both. As can be expected, the largest variation in primary amide charge transfer occurs between the cases of Li- and F+ bracketing C-N versus that with the positions of the ions swapped giving a C-O(pi)* NBO occupancy difference of 0.2 electron (Figure 4.6), some 200 times the nominated level of chemical significance being 0.001 electron. We have not explored fields of larger and increasingly biologically implausible magnitude and do not give an upper bound on the magnitude of the variation that is physically possible.

Ap1:Figure 4.21 and Ap1:Figure 4.22 show the variation in Natural Localized Molecular Orbital [6], NLMO, dipole moment magnitude with field. NLMOs are unitarily equivalent to canonical molecular orbitals and the vector sum of their dipole moments for the molecular system is that also given by canonical molecular orbitals. Significant contributions of the dipoles of lone pairs to the overall dipole of formamide are noted on p150-151 of [80]. The dipole of the oxygen lone pairs of formamide are aligned with the C-O axis, and importantly, the dipole of N(lp) is aligned with the C-N axis. This is surely associated with the N(lp)->C-O(p)* NBO charge transfer. Since this charge transfer is contained within the N(lp) NLMO, the associated dipole is best viewed in terms of NLMOs. The change in this dipole is directly physically related to electrostatic field, whereas the change to N(lp)->C-O(p)* does not by itself reveal the underlying physics.

Ap1:Figure 4.26 shows the Second-Order Perturbation Theory Energy [6], SOPT energy, associated with the primary amide charge transfer for Li+ without F-, giving variation of up to 16 kcal/mol in the range of distances considered. The equation given in Figure 4.1 for second-order correction has the square of the Fock NBO off-diagonal matrix element in the numerator and the gap in the energy levels between acceptor and donor orbitals in the denominator. Ap1:Figure 4.27 to Ap1:Figure 4.32 show variation in energy level of N(lp), C-O(pi)*, the gap between them, and the square of Fock NBO matrix element with ions placed in the C-N line. Electrostatic variation of resonance-type charge transfer might be dissected as being primarily due to variation in the overlap of the NBOs which then causes variation in the energy levels of the NBOs, but change in the dipole of the resonance-type charge transfer is perhaps more compelling.

$$E_{i\to j}{}^{(2)} = -n_i{}^{(0)} \frac{\langle \phi_i{}^{(0)} | \hat{F} | \phi_{j*}{}^{(0)} \rangle^2}{\epsilon_{j*}{}^{(0)} - \epsilon_i{}^{(0)}}$$

Figure 4.1. NBO Donor-Acceptor Second-Order Correction (Eq 1.24 p19 [6]) for stabilizing interaction between filled donor orbital $\boldsymbol{\phi}_i^{(0)}$ and an unfilled acceptor orbital $\boldsymbol{\phi}_j^{*(0)}$ with energy levels $\boldsymbol{\epsilon}_i^{(0)}$ and $\boldsymbol{\epsilon}_j^{*(0)}$ and orbital occupancy $\boldsymbol{n}_i$ and Fock operator $\hat{F}$

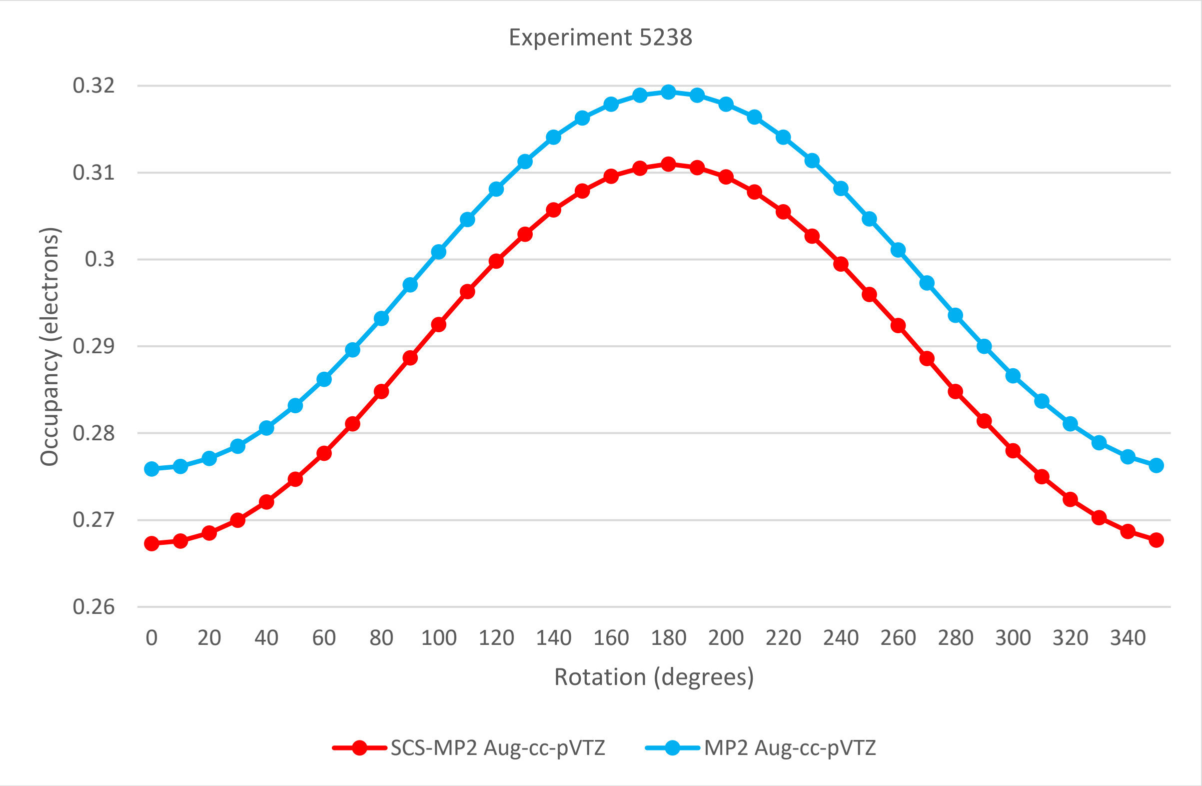


Figure 4.2. C-O(pi)* NBO Occupancy in N-methylformamide in 0.005 au Uniform Electrostatic Field Rotated in O-C-N Plane from N-C Vector

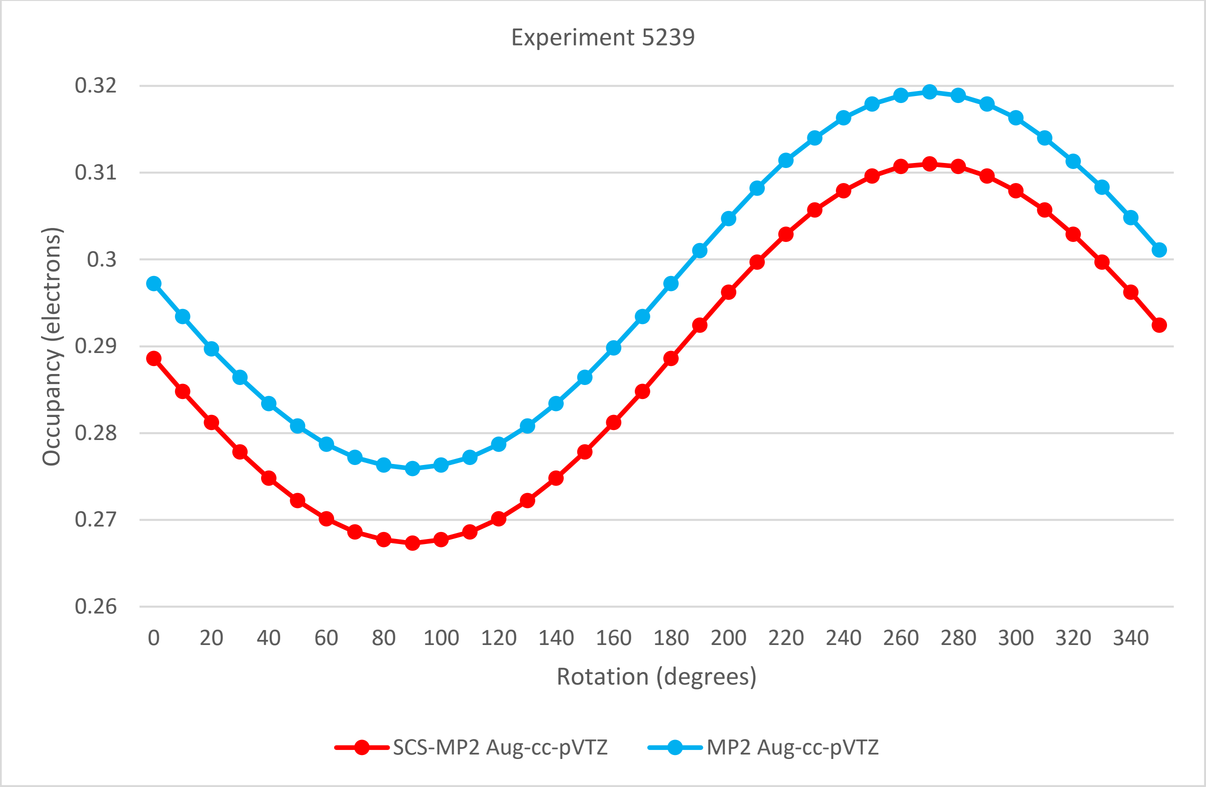


Figure 4.3. C-O(pi)* NBO Occupancy in N-methylformamide in 0.005 au Uniform Electrostatic Field Rotated in Plane Containing O-C-N Normal and N-C Vector Starting from O-C-N Normal

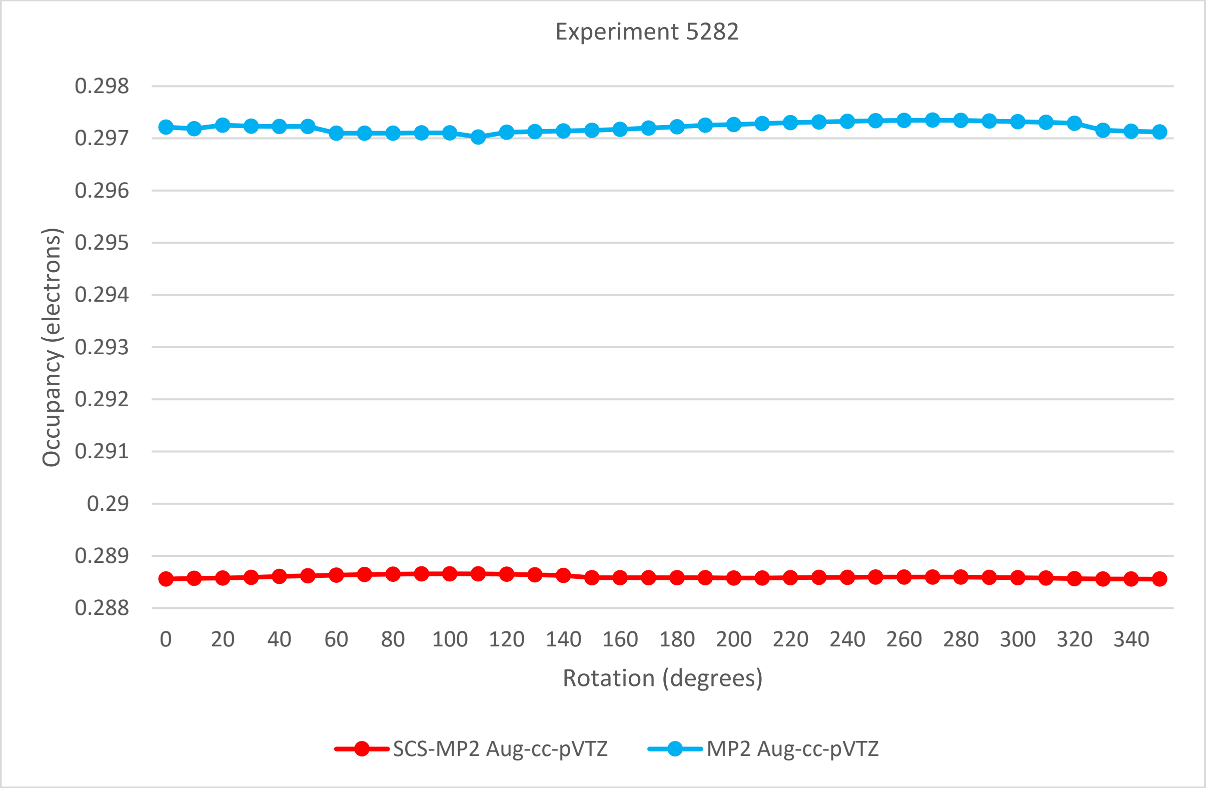


Figure 4.4. C-O(pi)* NBO Occupancy in N-methylformamide in 0.005 au Uniform Electrostatic Field Orthogonal to C-N Vector and Rotated from O-C-N Plane

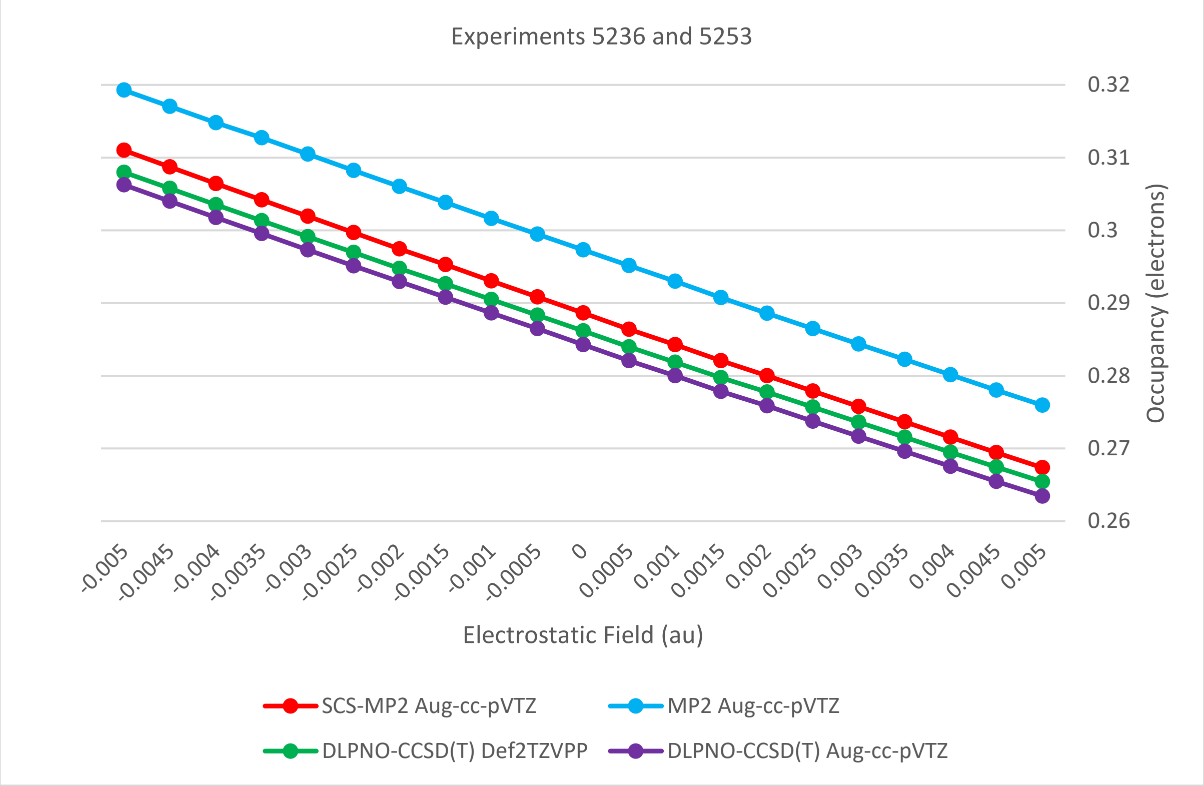


Figure 4.5. C-O(pi)* NBO Occupancy in N-methylformamide in Uniform Electrostatic Field with N-C Vector

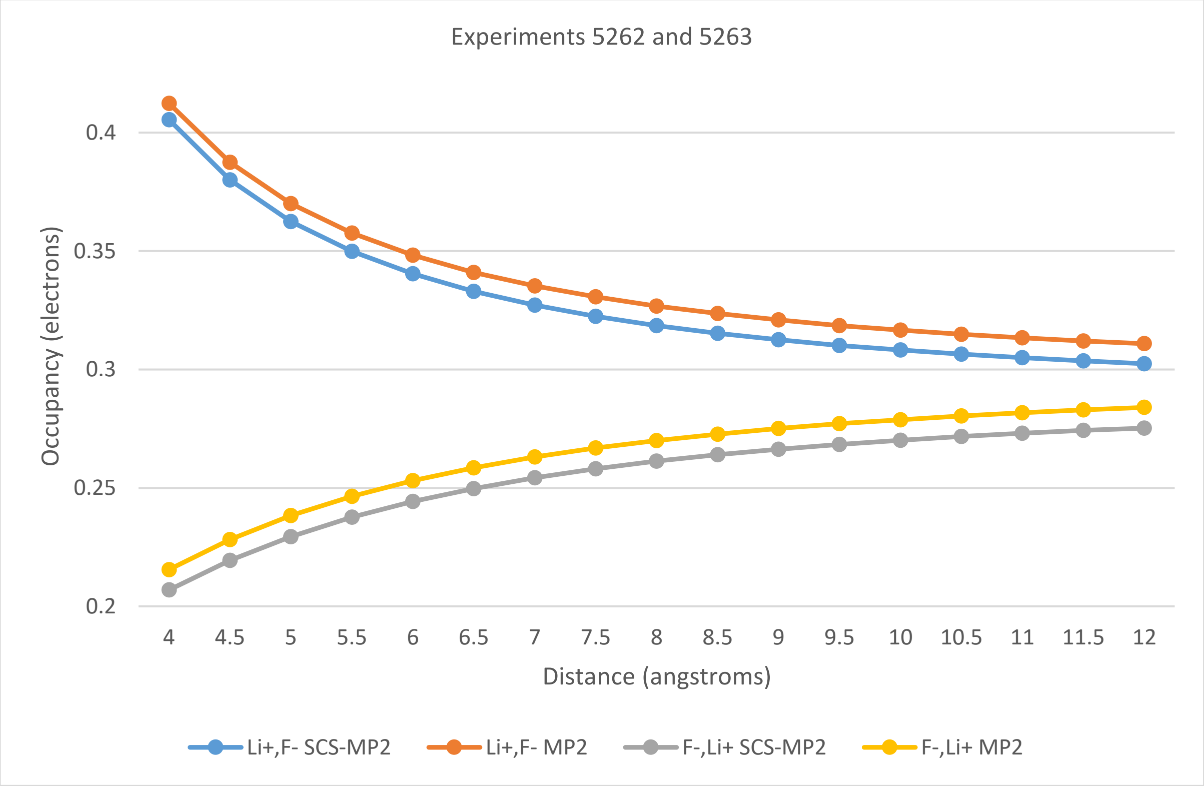


Figure 4.6. C-O(pi)* NBO Occupancy in N-methylformamide with Ions on C-N Line, right-listed Ion at Distance from N, left-listed Ion at Distance in Other Direction from C with aug-cc-pVTZ

4.9.2 Partial sp3 hybridization at N

Amide O-C-N normal electrostatic field can cause partial sp3 hybridization at N depending on amide resonance, but may also arise from mechanical torsion arising in the molecular context. However the sp3 hybridization arises, amide resonance is reduced since the bond order of C-N is reduced. Another view of this is that overlap of the N(lp) with C-O(pi)* becomes less favourable, since N(lp) takes on some s character in keeping with partial sp3 hybridization at N. Reduction in amide resonance enables sp3 hybridization at N, and increasing amide resonance inhibits sp3 hybridization at N [6]. In the absence of amide resonance, N would completely sp3 hybridize.

With sp3 hybridization at N introduced by rotating O-C-N-H and O-C-N-CA so that H and CA move away from the O-C-N plane, Ap1:Figure 4.33 shows that there is no change in sp3 hybridization when O-C-N normal electrostatic field of 0.005 au is applied, and Ap1:Figure 4.34 shows there is no change in primary amide charge transfer. Ap1:Figure 4.35 shows that with O-C-N-H and O-C-N-CA dihedral angles 10 degrees from planar an O-C-N normal field varying from -0.005 to 0.005 au results in no change in the primary amide charge transfer. It can be concluded that the change in amide resonance in O-C-N normal field is entirely due to the partial sp3 hybridization at N.

Ap1:Figure 4.36 shows the change O-C-N-H dihedral angle from planarity with unconstrained H, CA and HA as a 0.005 uniform field is rotated in the plane containing the O-C-N normal and C-N vector, and Ap1:Figure 4.37 shows the primary amide charge transfer. SCS-MP2 is somewhat smoother in calculating the O-C-N-H dihedral, though the primary amide transfer varies smoothly for MP2 as well. In can be concluded that the maximum and minimum primary amide charge transfers in these experiments occurs when the field passes through the O-C-N plane even when the O-C-N-H and O-C-N-CA dihedrals are unconstrained.

Figure 4.7 shows the primary amide charge transfer variation as a Li+ or F- ion moves in a line parallel to N-C but shifted 5 angstroms above the O-C-N plane. The ion moves from 0.5 angstroms before the intersection with the O-C-N normal from N to 2.0 angstroms after. The intersections of similar methods for different ions gives a position neutral to charge transfer variation. For both methods used, the neutral point is close to 0.7 angstroms from above the point above N, and at neutral point the occupancy values are very close to those for no field (Figure 4.5). At this neutral point, the C-N bond length is 1.350 angstroms, so the neutral point is a slightly closer C than N.

Ap1:Figure 4.38 shows that there is no difference in the primary amide charge transfer resulting from having Li+ and F- on the O-C-N normal from N equidistant from N but on different sides of O-C-N plane compared to constraining the O-C-N-H and O-C-N-CA dihedrals to the result of the presence of the ions but removing the ions. Ap1:Figure 4.39 shows that this is not the case when only one ion is used. It can be concluded that the two ions cause a field that passes through the O-C-N plane normal to that plane but the single ion does not and provides a component parallel to the N-C vector, thus decreasing the amide resonance further than that due to sp3 hybridization at N only. Ap1:Figure 4.40 shows the similar result with the ammonium ion used in place of Li+, where the distance shown is to the ammonium nitrogen. Charge transfer from amide N(lp) NBO to any ammonium N-H can be taken to be slight at 5.0 angstroms. Ap1:Figure 4.41 shows the difference in primary amide charge transfer between the ion present and ion removed but dihedrals retained cases for 4 methods and 2 basis sets.

Ap1:Figure 4.42 shows the variation of N(lp) NBO s character with variation in O-C-N normal uniform field rather than ions with unconstrained O-C-N-H and O-C-N-CA dihedrals, and figure Ap1:Figure 4.43 shows the difference in primary amide charge transfer.

As given by [6] and our data which is not shown, increase in amide resonance such as by RAHB inhibits sp3 hybridization at N. Once an RAHB chain has formed and resonance increased, sp3 hybridization at N will be prevented. However, if in the presence of an electrostatic field with component normal to the O-C-N plane the RAHB chain is weakened for example by thermal jolting and partial sp3 hybridization becomes possible at any point in the chain, significant loss of RAHB may result. Also, H-N, C-O or the whole amide group may rotate in the electrostatic field, weakening hydrogen bonds. A field normal to the amide plane can be expected to introduce an RAHB chain nucleation barrier.

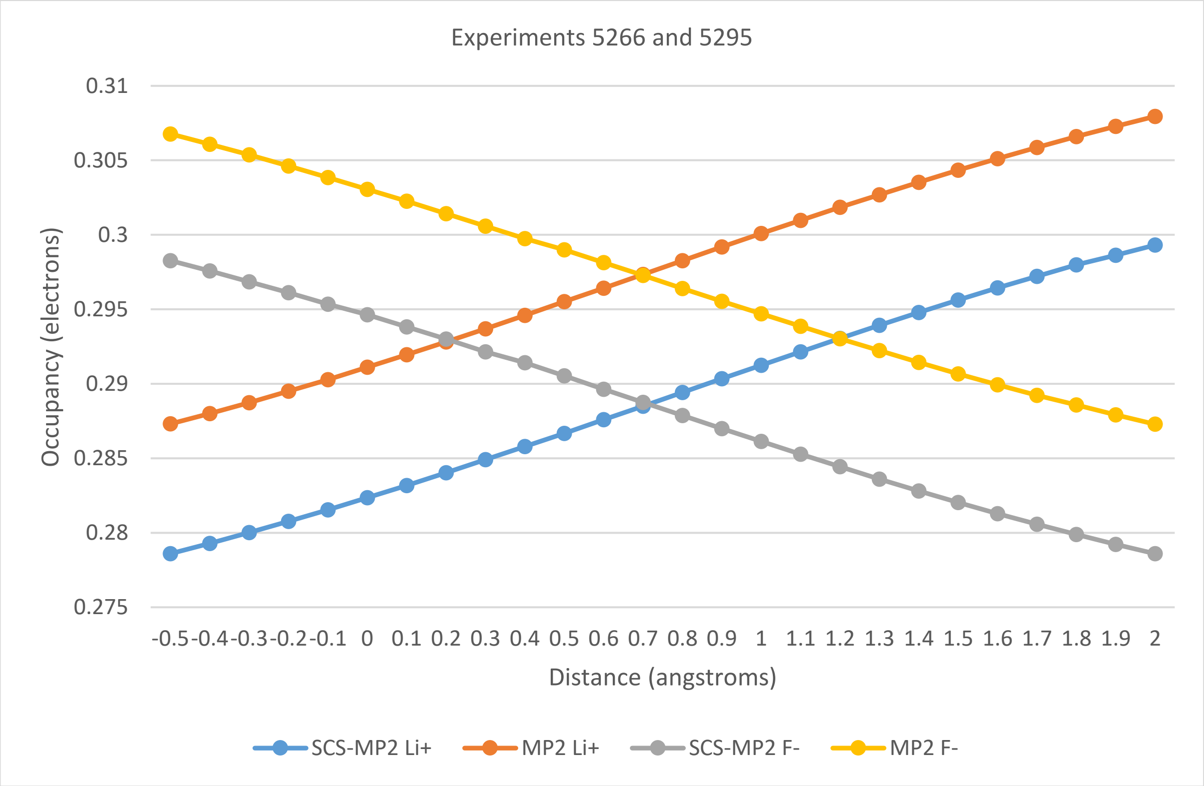


Figure 4.7. C-O(pi)* NBO Occupancy in N-methylformamide Constrained to Planarity with Li+ on Line Parallel to N-C Vector 5 Angstroms Above O-C-N Plane at Distance from Point Above N

4.9.3 Other geometry variations

Ap1:Figure 4.44 shows the primary amide charge transfer with rotation of the methyl group in N-methylformamide in orthogonal electrostatic fields. There is very little difference in the primary amide charge transfer between the cases of the field being 0, the field being along the y axis and the field being along the z axis. The case with the field along the x axis shows the expected variation for field aligned with the N-C vector. The N(lp) NBO's busy donation to the methyl group is not varied by the presence of electrostatic field of 0.005 au.

Ap1:Figure 4.45 shows the primary amide charge transfer with twisting about the C-N bond in N-methylformamide in orthogonal electrostatic fields. There is a slight separation of the curve for z-axis field from that of 0 and y-axis fields, but this is beneath the level of chemical significance.

Ap1:Figure 4.46 shows the primary amide charge transfer with rotation of the carbonyl substituent methyl in N-methylethanamide in orthogonal electrostatic fields. The separation of the curves for y-axis field from the 0 and z-axis fields is likely due to the optimal field vector being offset by 5 degrees from C-N in N-methylethanamide, with variation in primary amide charge transfer at twice the level of chemical significance. The slight separation of the 0 and z-axis field cases is likely due to variation in hyperconjugation between the methyl group and the carbonyl group.

Ap1:Figure 4.47 shows the primary amide charge transfer in N-methylethanamide with pyramidalization at the carbonyl carbon formed by rotating the CA-N-C-CA dihedral while maintaining the CA-N-C-O dihedral. The explanation offered is as for the case in the previous paragraph.

4.9.4 Torsional steering

A means by which change in resonance might change the torsional barrier across a single bond connecting to the atom which is central to resonance, which we refer to here as torsional steering, is discussed by [6] p693-702. An example of this is a methyl group bonded to the amide carbon, forming a bond similar to CA-C in proteins. A torsional barrier might be introduced through hyperconjugation between a CA-H bond or antibond and C-O antibonds or bonds. This is entirely plausible, and appreciable charge transfer CA-H->C-O(sigma)* is seen in Ap1:Figure 4.46, though the torsional steering proposal suggests that greatest stability is found with H-CA in cis with C-O rather than our finding that greatest stability occurs in trans. This difference in findings in undoubtedly due to our use of SCS-MP2/aug-cc-pVTZ rather than B3LYP/6-311++G**. Figure 4.8 shows the energy associated with N-methylethanamide with variation in O-C-CA-H dihedral from the trans to cis conformations. The trans conformation is ~0.3 kcal/mol lower in energy than cis. This is about an order of magnitude less than the torsional barrier of the torsional steering proposal, and slope has opposite sign. The main point of our experiment is to find what change to torsional barrier results from change to amide resonance. The different electrostatic fields cause markedly different amide resonance, and it can be seen that the curves for the fields do not differ other than being offset on the y-axis, that is, the torsional barrier is the same at each field and hence resonance. At the method and basis set used, these findings are not in accord with the torsional steering proposal.

The difference in molecular energy levels between the field strengths at a given dihedral angle may be taken as the upper bound of the variation in energy associated with the primary amide charge transfer, since purely electrostatic interactions may also be present. The change in energy associated with amide resonance between N-C vector field of -0.005 au and 0.005 au is ~7 kcal/mol per backbone amide. In terms of protein stability, this is considerable and occurs even without any hydrogen bonding to the amide O or H.

In reproducing these data, be aware that while the N-methylformamide experiments used dummy atoms to introduce constraint to planarity constraints and constrain the N-C bond to the x-axis, the N-methylethanamide experiments use two helium atoms at 8 angstroms from N, and the two helium atoms are on the y and z axes respectively.

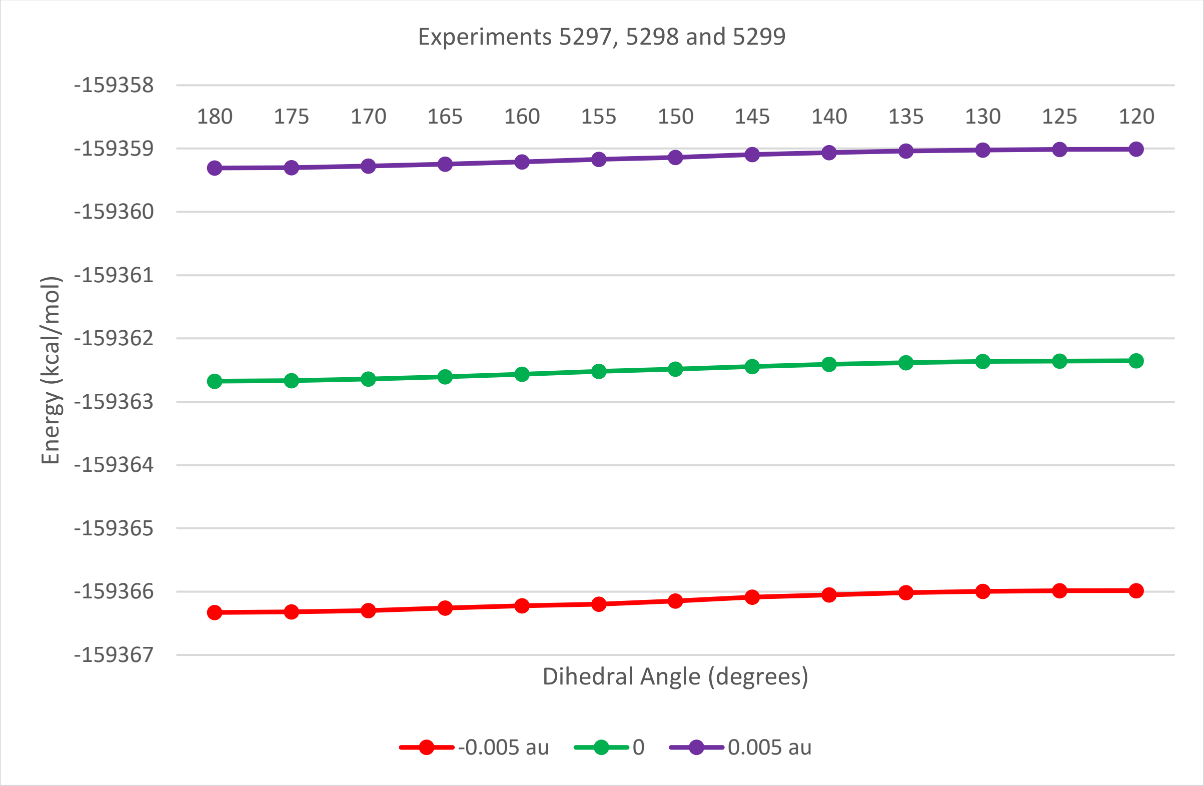


Figure 4.8. Energy Associated with N-methylethanamide at O-C-CA-H Dihedral Angle in Uniform Electrostatic Field with N-C Vector at SCS-MP2/aug-cc-pVTZ

4.9.5 Backbone amide nitrogen as hydrogen bond acceptor

If the backbone amide nitrogen participated directly in hydrogen bonds, amide resonance would be reduced, since charge transferred from N(lp) for the hydrogen bond would diminish the disposition of N(lp) to participate in the primary amide charge transfer, an instance of the busy donor effect [6]. This manner of variation of backbone amide resonance is proposed by Weinhold and Landis [6] as a means by which their resonance-based torsional effect comes into play during protein folding. In this section, we investigate the existence of this hydrogen bonding in the folded state on the possibility that phenomena occurring during folding might be seen in some aspect of folded state such as random coil.

In a collection of experiments, the central atom of a molecule of water, hydrogen sulfide, methane, ammonia and the ammonium ion was constrained to the O-C-N normal from N, but its distance from N was not constrained. The hydrogen bond length and C-O(pi)* occupancy following geometry optimization is shown in Ap2:Table 4.2. The hydrogen bond in the case of methane which represents a residue sidechain methyl group is of a length not conducive to charge transfer from N(lp) to methane, and the C-O(pi)* occupancy is largest. Modestly less C-O(pi)* occupancy is seen with ammonia which represents H-N bonds in uncharged groups. Less again C-O(pi)*occupancy is seen with hydrogen sulfide which represents a non-ionized cysteine head group. The C-O(pi)* occupancy with water is less again, and at 0.025 electrons less than the C-O(pi)* occupancy with methane this leads to the expectation that backbone amide N(lp) in RAHB secondary structures will be protected from access by water. Greatest reduction in C-O(pi)* occupancy occurs with the ammonium ion. As demonstrated above, a positively charged ion on the O-C-N normal from N creates an electrostatic field with C-N vector component at the midpoint of the C-N, reducing amide resonance directly and by pyramidalization at N. Also, positive charge in the hydrogen bond donor assists hydrogen bonding generally. We investigate hydrogen bonding with backbone amide N by positively charged residue sidegroups.

Spatial queries over an extract of the Protein Data Bank [164], PDB, to discover the incidence of charged arginine or lysine hydrogen bonding to backbone amide nitrogen were performed. X-ray crystallography [165] structures were not considered since coordinates for the amide proton were required. The extract was for solution Nuclear Magnetic Resonance [166], NMR, PDB files containing protein with no modified residues and no nucleic acid and at the date it was executed (2014-09-08) it returned 8,378 files, and only the first model of each was considered. These models were automatically checked for errors, such as non-contiguous residue sequence number, missing atoms and helices with C-terminal residue sequence number not larger than the N-terminal residue sequence number, eliminating 1,690 models and passing 6,688 models. The following constraints were imposed on identifying hydrogen bonds: N-H..amide N to less than 45 degrees, angle between amide normal and amide N to H vector not in the range 45 to 135 degrees and amide N..H distance to less than or equal 3.0 angstroms. In alternative search, no angular restrictions were imposed and the distance was restricted to 5 angstroms. The 3 polar hydrogens of charged lysine and the 5 of charged arginine were evaluated.

There is no instance of ammonium group of lysine or guanidinium group of arginine hydrogen bonded to backbone amide N according to these criteria in any of the 6,688 PDB NMR models queried. It is likely that more standard amide hydrogen bonding and increased resonance is always more favourable than N lone pair participation in hydrogen bonding which reduces resonance. The proximity of these groups to the backbone amide O-C-N normal from N might occur during protein folding but does not persist even in random coil.

4.9.6 RAHB chains in electrostatic field

We demonstrate the significant consequences of electrostatic field for RAHB in hydrogen bonded chains of amides. With co-planar N-methylformamide units hydrogen bonded C-O...H-N in the manner of RAHB protein secondary structures with the hydrogen bond angle set to 5.209 degrees so that all amide C-N vectors are parallel, the effect of electrostatic field on RAHB in multi-amide systems of molecules was investigated. In this multi-amide experiment, all atoms except 2 methyl hydrogens in each molecule constrained to the common planarity of all the molecules, and the amide N-C-O and C-N-H angles were constrained to that optimal for the molecules in isolation and the hydrogen bond angles constrained to the initial setting. No bond length was constrained, and the only non-dummy atom constrained to coordinates was that of the nitrogen of the first unit. This multiple N-methylformamide chain is shown in Figure 4.9.

Ap1:Figure 4.48 supports the use of RI-SCS-MP2/aug-cc-pVDZ with Coulomb and correlation auxiliary basis sets for experiments in which a uniform electrostatic field varies amide resonance. This pairing of method and basis set is used for its computational efficiency at larger atom counts.

Variation in C-O(pi)* NBO occupancy at each amide in this N-methylformamide chain geometry optimized for three different field directions, with constraints described above so that the chain remains linear are shown in Figure 4.10 to Ap1:Figure 4.55. The first field (Figure 4.10) has N-C vector, and the second (Ap1:Figure 4.49) is orthogonal to this. The second field has no N-C component, but both have component along the hydrogen bonds, so a third field (Ap1:Figure 4.55) is used which is orthogonal to the H-N vectors so that change is due to variation of amide resonance rather than hydrogen bond resonance. The data shown in Figure 4.10 and Ap1:Figure 4.55 is in accord with expectation, being that amide resonance peaks in the middle of the hydrogen bonded chain, and is varied as for the monomer case by electrostatic field. Also, RAHB in the chain is varied with hydrogen bond resonance alone (Ap1:Figure 4.49).

Hydrogen bond lengths for the three experiments described in the previous paragraph are shown in Figure 4.11, Ap1:Figure 4.50 and Ap1:Figure 4.56. It is apparent that in this unidirectional RAHB chain that the hydrogen bond length becomes less related to the C-O(pi)* occupancy along the chain and more related to O(lp) occupancy buildup which occurs because the final unit has no further acceptor to which to donate charge. It is clear that electrostatic field can eliminate the energetic penalty associated with this charge buildup, and this is in accord with a positive charge cap at the end of the chain being stabilizing [167]. A negative cap at the beginning of the chain would also assist in so much that it contributed to field throughout the chain. Of course, an extra charge donating unit, rather than only electrostatics, at the beginning of the chain would increase the resonance of the chain. Note that hydrogen bond length variation is somewhat muted because the inter-amide charge transfer and steric interactions give a net energetic result close to zero and the binding energy is primarily associated with the N(lp)->C-O(pi)* of the amides on each side of the hydrogen bond (Chapter 5).

The remainder of the figures in this section relate to the third experiment in which the field was orthogonal to the H-N bonds to eliminate variation to hydrogen bond resonance. Ap1:Figure 4.51 shows the oxygen Natural Atomic charge along the chain, and Ap1:Figure 4.52 shows the amide proton NAO charge along the chain. The Natural Atomic charge of first amide proton in the chain is not varied by this electrostatic field, hence it is not varied by change in amide resonance.

Ap1:Figure 4.53 and Ap1:Figure 4.54 show the change in occupancy in O(sigma-lp) and O(pi-lp) NBOs in the H-N vector orthogonal field. The O(sigma-lp) NBO of the terminal units are not varied by this field, and in the non-terminal units there is slight variation. In the final unit, the occupancy is higher than in other units, in keeping with the unavailability of an acceptor. In contrast, the terminal O(pi-lp) NBO occupancy is markedly less than that of the other units which is in keeping with decreased C-O(pi)* NBO occupancy, but also declines at the terminal unit with increasing C-O(pi)* NBO occupancy due to field. This divergence of the occupancy of the 2 oxygen lone pairs could be explored further, but we do not do so here. Confirmation that electrostatic field orthogonal to the H-N vector does not differentially vary the O(sigma-lp)->H-N* and O(pi-lp)->H-N* resonance-type charge transfers might also be sought.

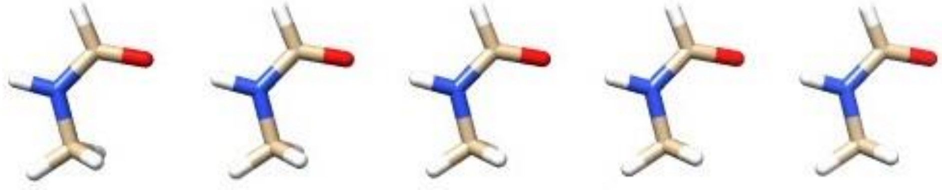


Figure 4.9. N-methylformamide Hydrogen Bonded Chain

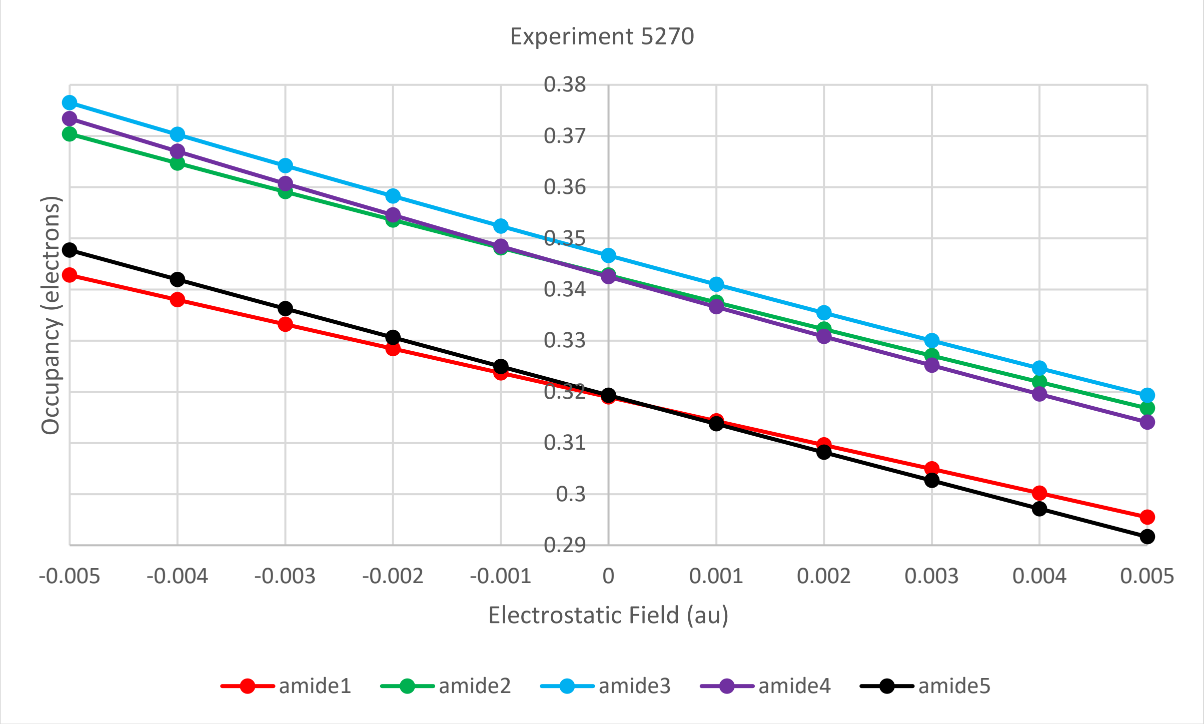


Figure 4.10. C-O(pi)* NBO Occupancy in Hydrogen Bonded N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field with N-C Vector at RI-SCS-MP2/aug-cc-pVDZ with Coulomb and Correlation Auxiliary Basis Sets

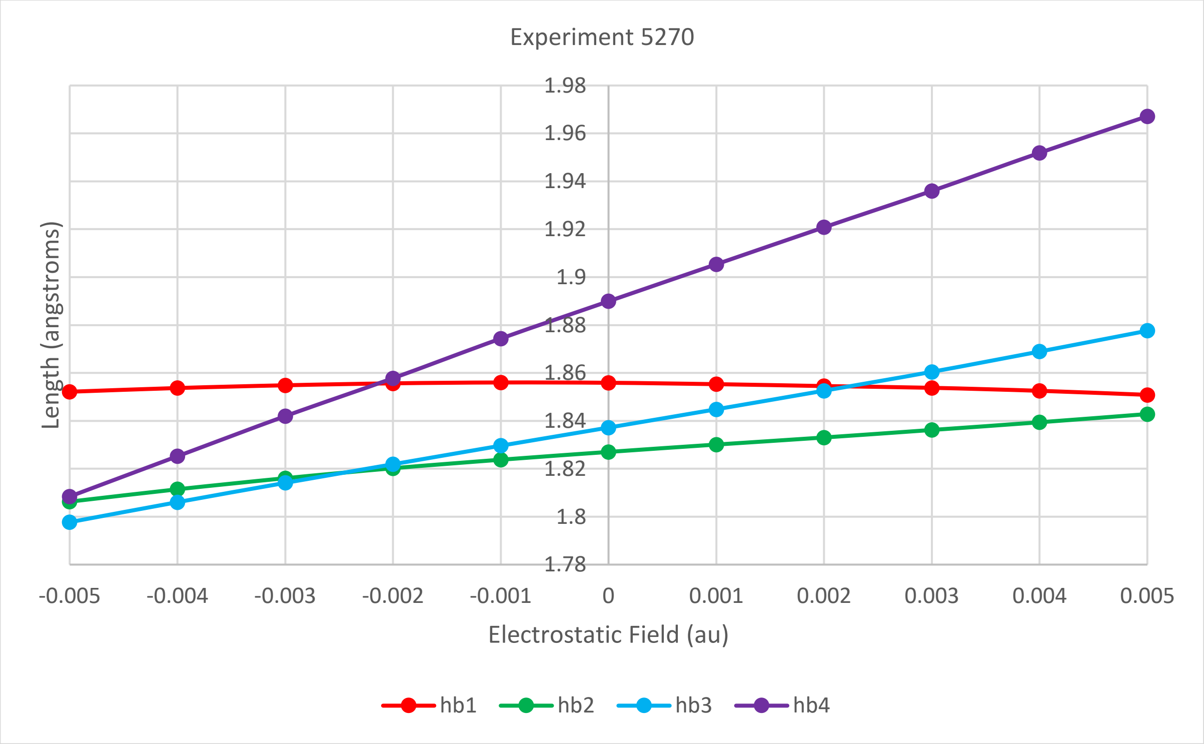


Figure 4.11. Hydrogen Bond length in Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic with N-C Vector at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

4.9.7 Electrostatic field vectors at backbone amides

The foregoing demonstrates or otherwise implies the relevance of electrostatic field to RAHB and hence to RAHB protein secondary structures. This leads to an interest in observation of the electrostatic field vector at backbone amides in proteins, and that the backbone amide is not protected from variation in electrostatic field with C-N vector component.

Emphasizing that Newtonian or classical simulators have an uncertain account of resonance, but wishing to obtain some estimate of electrostatic field vectors at backbone amides for proteins, we used the Tinker 7.1.2's [168] AMOEBA Protein 2013 [169] force field and Generalized Kirkwood continuous solvent [170] to estimate the electrostatic field vectors at backbone amide nitrogen (Figure 4.12) and carbon (Figure 4.57) in accordance with the calculation described in [171] for 10 minimized protein structures having initial coordinates derived from the PDB, entries 1D27, 1H1J, 1IMP, 1UUA, 2JOF, 2LHD, 2LJI, 2LT8, 2LX9 and 2LXR.

The variation in C-N component of electrostatic field is larger than the general intra-protein fields observed by [155, 157-161]. These calculations offer no support for a notion that the C-N component of electrostatic field is constant or protected at protein backbone amides. An appreciably smaller field than calculated by AMOEBA with Generalized Kirkwood continuous solvent can be expected to result in significant influence on protein structure.

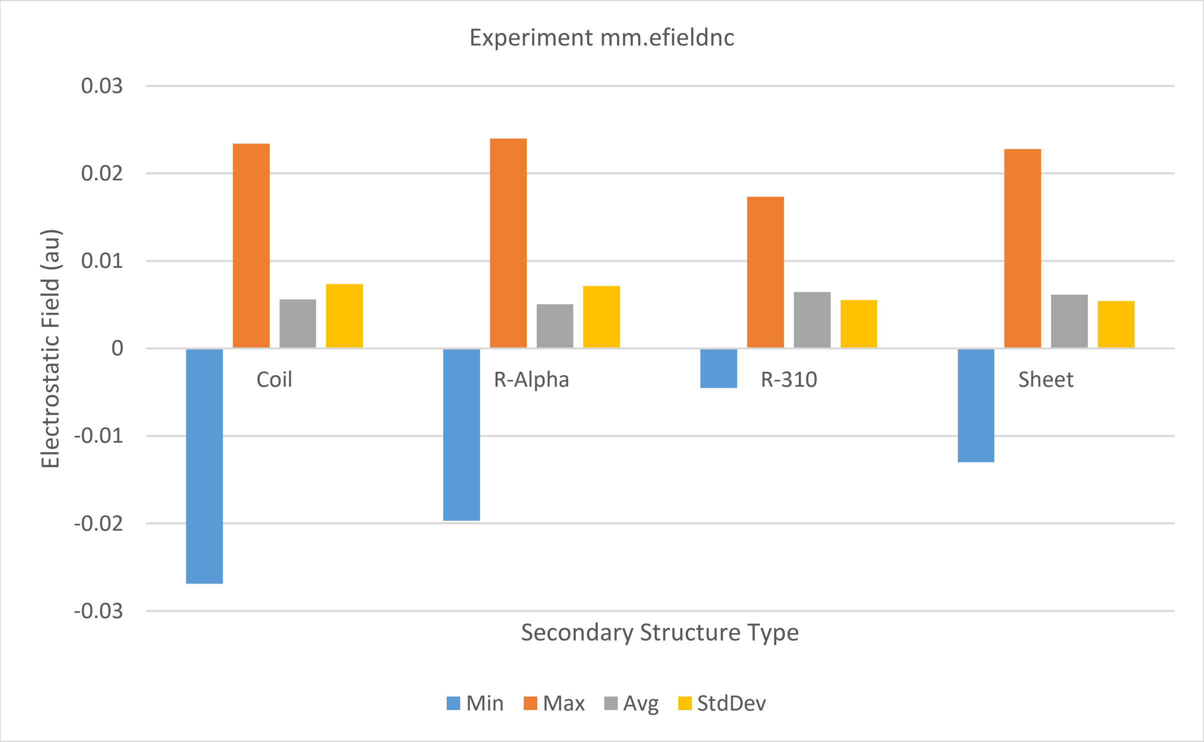


Figure 4.12. C-N Vector Component of Electrostatic Field at N in Backbone Amides of 10 Small Proteins as Calculated with AMOEBAPRO-2013 with Generalized Kirkwood Continuous Solvent

4.9.8 Protein beta sheet

The multi-amide structure described above (Figure 4.9) differs from a chain of backbone amides in a parallel beta sheet only in that the units of the multi-amide structure are N-methylformamide rather than backbone amides. The C-N vectors within a RAHB chain of backbone amides of a parallel beta sheet are all aligned. As for the multi-amide experiments above, the resonance of every backbone amide in a chain in a parallel beta sheet will be varied in a similar manner by a uniform electrostatic field, the varied resonances cooperatively determining the extent of hydrogen bonding in the chain. The results described for the multi-amide structure may be taken as indicative of the behaviour of a backbone amide chain in a parallel beta sheet in an electrostatic field.

The backbone amide C-N vectors of a parallel beta sheet (Figure 4.13) HB chains an even number of chains away from a nominated chain are all aligned. The backbone amide C-N vectors of chains an odd number of chains away from a nominated chain are also all aligned, but these vectors have a component antiparallel to those of those an even number of chains away. A uniform electrostatic field may be applied that increases RAHB in one set of chains, even or odd, but to different extent reduces it in the other set. This leads to a hypothesis given in Section 4.10.1 (Hypotheses/Beta sheet).

An antiparallel beta sheet differs from a parallel beta sheet with respect to patterns of backbone amide C-N vectors (Figure 4.14). In a single antiparallel beta sheet RAHB chain, the C-N vectors are alternately substantially orthogonal. A uniform electrostatic field may be applied such that the resonance of alternate amides is increased or decreased with little to no direct change to resonance of the other amides in the chain. Chains in the sheet an even number of chains away from a nominated chain have a similar pattern of amide C-N vectors, so the RAHB of these two chains may be similarly modulated by a field. C-N vectors of amides on chains adjacent a nominated chain but on the same beta strand are largely orthogonal but on adjacent beta strands they are largely opposed, so RAHB in the chains an even number of chains away from the nominated chain may be varied in opposition to chains an odd number of chains away.

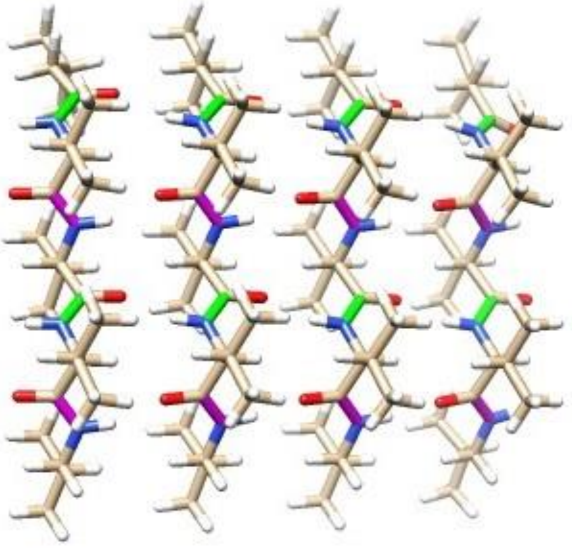


Figure 4.13. 2 sets (magenta, chartreuse) of parallel C-N Bonds in Parallel Beta Sheet

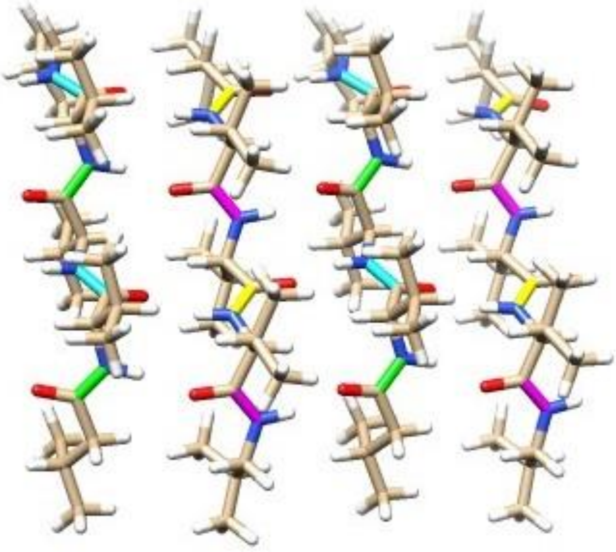


Figure 4.14. 4 sets (yellow, magenta, light blue, bright green) of parallel C-N Bonds in Antiparallel Beta Sheet

4.9.9 RAHB protein helices

In a RAHB protein helix, the helix macro dipole is collinear with the helix axis which is not altogether orthogonal to the backbone amide C-N vectors. The electrostatic field associated with the helix macro dipole [172] will reduce the resonance of the backbone amides and hence cooperatively reduce RAHB in the RAHB chains or spines as they are sometimes referred to in the context of helices. Charge caps, negative charge at the N-terminus of the helix and positive charge at the C-terminus, can compensate for the field associated with the helix macro dipole and hence assist RAHB. Where the charge caps are also involved in charge transfer, they tend to balance the RAHB terminal charge deficits which arise in unidirectional RAHB because charge transfer into and out from these sites does not cancel [6].

The average angle between the helix axis and the C-N vectors differs between the helix types 3-10, alpha and pi and the field associated with the helix macro dipole also differs between these types because of a difference in divergence of average C-O and H-N vectors from the helix axis and differences in hydrogen bonding arising from these geometry differences, and the variation in backbone amide resonance caused by the helix's intrinsic electrostatic field differs between the helix types.

RAHB helices may be less susceptible to destabilization by uniform electrostatic field not collinear with a helix axis than a beta sheet is to field with any direction due to the lower proportion of C-N vectors in a helix backbone chain being aligned. In an alpha helix for example, the backbone amide C-N vectors are approximately aligned only when the helix turn of ~3.6 backbone amides approximates an integer. Uniform electrostatic field not collinear with the helix axis may be more destabilizing for some helix types than others due to their different turns. The different RAHB protein helix types, 3-10, alpha and pi have different numbers of spines. The 3-10 helix has only 2 spines, which may make it more susceptible to RAHB irregularity caused by electrostatic field.

Charge positioned away from the helix axis, such as on an amino acid residue sidechain, varies each backbone amide resonance depending on the vector of the non-uniform electrostatic field at the amide. Reduction of RAHB in a spine requires reduction of resonance in only one amide, and an equal enhancement of resonance in another amide of the spine does not necessarily compensate, depending on the position of the two amides in the spine. A decrease in resonance at one unit can be expected to make RAHB behave more as two separate RAHB chains either side of that unit.

The electrostatic field created by charged sidechains may be modulated by being surrounded by high permittivity medium such as occurs by immersion in water, and by salt-bridging. An absence of these modulations of electrostatic field caused by charged sidechains invokes the energetic cost of reducing helix RAHB.

4.9.10 Proline

Where the C-terminal-side residue in a peptide bond is the imino acid proline, a N(lp)->C-O(p)* interaction still exists. A field parallel to the C-N bond will tend to reduce variation in peptide bond twist away from that optimal for that donor-acceptor interaction and increase the energetic barrier of a cis-trans or trans-cis transition, and a field antiparallel to the C-N bond will tend to increase variation in peptide bond twist and decrease energetic barrier of a cis-trans or trans-cis transition, which is certainly relevant to protein folding and may be exploited by prolyl isomerases [9].

4.9.11 Permittivity

The hydrophobic interior of a protein has a low permittivity and well supports an electrostatic field. The whole of the sidechain of hydrophobic residues has a low permittivity, and the hydrophobic stalk of every residue type other than glycine and proline has the low permittivity of the alkanes. This is much smaller than the relative permittivity of water which at 80 degrees F is ~80 times that of vacuum [173]. Formamide, which is the functional group of sidechain amides, has a yet higher permittivity and so is more capable of damping a field than water.

Where protein-protein binding interfaces are largely hydrophobic, the hydrophobic interiors of the proteins may be effectively joined, with the electrostatic field of one interior extending into the other.

A charged residue without neutralizing salt-bridge partner in an otherwise hydrophobic protein, or similarly unpaired charged residue at a hydrophobic binding interface that excluded solvent on binding to partner protein would exert long-range electrostatic effects. A change to internal protein electrostatic environment may cause change to secondary structure and conformational change.

Conformational change due to unpaired charged lysine was explored in [174].

4.9.12 Protein folding, conformational and allosteric change

The change in protein structure due to change in electrostatic field may be quite extensive and long-ranged, for a change in field may change secondary structure, which may in turn change the position of functional groups of sidechains which then changes the field, changing yet more secondary structure and so on. Electrostatic field variation of backbone amide resonance can be expected to be important in conformational change in proteins. Similar remarks apply to allosteric change [175].

The electric flux local to amide group will vary during protein folding, with the amide groups preferring to orient to the local flux so as to maximize amide resonance and the energy associated with resonance charge transfer and hydrogen bonding. Variation in the total energy associated with backbone amide resonance during folding may prove a revealing metric, as might the sum for all backbone amides of the C-N component of electrostatic field.

In summary, it is being proposed that the amide local electrostatic environment which is substantially determined by residue sidegroup charge, polarity and permittivity is an important factor in protein folding, structure and function. How important it is relative to other factors in protein folding is yet to be determined, but there is an elegance to this mechanism, offering explanation as it does of the fundamental features of nature’s primary structural polymer.

4.9.13 Unstructured regions

Electrostatic field in a region may prevent formation of secondary structure in a length of backbone due to diminishment resonance of backbone amides in secondary structure conformations, and less structured conformations allow larger total resonance.

4.9.14 Sources of electrostatic field variation in proteins

There are many sources of variation of electrostatic field in proteins. The functional groups of amino acid residue sidechains may be charged or polar. Post-translational modification may change charge or polarity. Phosphorylation and de-phosphorylation figure frequently in signal transduction [176]. Forming or breaking of salt bridges, such as between lysine and glutamine, changes the field. Solvation or de-solvation of charged or polar functional groups of sidechains changes electrostatic field. Binding by desolvated ions, such by the divalent calcium cation, Ca^{2+}, to Calmodulin [177] brings large changes to a field. Also, a single solvation shell is unlikely to fully dampen the field created by a divalent cation.

4.9.15 Absence of charged residues

In experiments by Kurnik *et al*. detailed in [178], S6 was deprived of all charged sidechains and neutrally C-terminally capped, and was found to fold normally. The absence of formal charge except N-terminally does not imply there is no electrostatic field environmental to backbone amides, for dipolar residues are present and ions from solution may associate with the protein. There is no threshold for the effect we describe and change of resonance in one backbone amide in an RAHB chain changes resonance and hydrogen bonding throughout the chain.

Charged residues which are salt-bridged or solvated may have modest bearing on electrostatic field at backbone amides, and their removal may make no difference to fold. However, this finding warrants including S6 in any physical survey of the C-N vector electrostatic field variation at C, N and the midpoint of C and N. If a protein's internal electrostatic field at each peptide group is negligible, folding would be possible due to intrinsic amino acid preferences with other causes such as steric interaction with the backbone [179] and steric blocking of backbone hydration [15] and hydrophobic [52] patterning of the amino acid sequence.

4.9.16 Development of methods

Successes in understanding protein folding can be had by modelling pure electrostatics and hydrophobia alone [180, 181]. Classical calculation is oblivious to peptide resonance, EVPR-CN and RAHB, and these omissions will limit what accuracy can be had.

Programs entered into the Critical Assessment of protein Structure Prediction [182] contests might be improved as a result of considering how well these programs capture EVPR-CN and RAHB. Similar remarks apply to Computational Protein Design [183], including de novo protein design.

We await the availability of quantum methods suitable for the quantification of EVPR-CN and its consequences in whole explicitly solvated proteins and protein complexes, noting the large errors of established DFT methods in calculating resonance in beta sheets (Chapter 3).

On the advent of accurate wavefunction methods with analytic gradients that scale to RAHB protein secondary structures and preferably whole proteins with medium and large non-Pople basis sets, observations may be made of the response of RAHB protein secondary structures and amyloid fibril to an electrostatic field. DLPNO-MP2 [184] and the most recent DLPNO-CCSD(T) implementation [185] do not have analytic gradients and need further constant-factor reduction in runtime but are promising. Improvement in non-wavefunction methods due to Density Perturbation Theory [186] seems promising. Accurate, linear scaling and low prefactor methods are highly desirable so that extensive explicit solvent may be used to immerse protein complexes for molecular dynamic calculations. The need for benchmark of the accuracy of all methods applied to proteins in calculating variation of backbone amide resonance is indicated. For this purpose and otherwise, development of physical methods suitable for accurate observation of backbone amide resonance and of the electrostatic field at the mid-point of the backbone amide C-N bond is highly desirable.

4.10 Hypotheses

4.10.1 Beta sheet

We predict that it is the similarity of RAHB of the backbone amide chains of a beta sheet which primarily determines the stability of that sheet and hence what residue sequences form a beta sheet. A variation in the resonance of amides of a backbone strand will cause irregularity in structure and variation in hydrogen bonding between a pair of strands. A hydrogen bond of greater binding energy than immediately adjacent hydrogen bonds in binding two backbone strands will tend to determine local geometry in its favour rather than in that of its adjacent hydrogen bonds.

In an antiparallel beta sheet, there is steric clash between the HA atoms of adjacent strands which is partially resolved by twisting of the strands and offsetting the strands in favour of the HA hydrogen bonding with O on the adjacent strand in competition with the H. If one backbone hydrogen bond has more favourable geometry than those immediately adjacent between the same pair of strands, the cooperative RAHB in its chain is advantaged leading to further geometry improvement in the favour of that chain. The adjacent chains are disadvantaged since their geometry is necessarily de-optimized. With sufficient RAHB mismatch, the beta sheet is unstable.

In a parallel beta sheet, the steric clash is not between HA atoms and is confined to that between HA and the H and O of the adjacent strand. There is an additional consideration being that even in the absence of steric considerations, only every second hydrogen bond between adjacent strands can be optimized due to the alternating distance between amide oxygen and subsequent amide proton of one strand facing the same neighbouring strand.

Greater balance in the hydrogen bonding between two strands will give more regular structure. The hydrogen bonded chains may traverse more than two strands with the hydrogen bonds between each pair of strands balanced, but balance between different pairs of strands is not required.

The RAHB chains of a possible beta sheet might be seen as independently pursuing RAHB maximization where geometry optimization cannot be had by all the chains, and should there be difference in the field at the backbone amides of the chains, the RAHB in some chains will exceed that of others. The RAHB chains are in competition for optimal geometry.

The hypothesis of the present section is irrespective of the means by which amide resonance and RAHB is varied. Amino acid residue sidechains vary the field and permittivity to backbone amides, and in this manner sidechains influence what sequences form beta sheets. A non-uniform electrostatic field could be crafted to particularly challenge or encourage the structural integrity of a beta sheet. Non-electrostatic influences on the formation of beta sheets are known, such as steric conflict with the backbone in other conformations [179] and blocking of backbone hydration [15].

In previous work (Chapter 3), we concluded that established non-double hybrid DFT methods significantly underestimated resonance in amides when the carbonyl bonds engage in hyperconjugative interactions such as occur in beta sheets and geometry optimization using these DFT methods should be taken as significantly underestimating RAHB.

4.10.2 Amyloid fibril

Amyloidogenicity is determined by the balance in hydrogen bonding between two beta strands as in a beta sheet, but requires a yet finer degree of balance. The electrostatic environment provided by a strand’s sidechains is repeated for each strand since the strands have the same residue sequence, so any variation in backbone amide resonance is cooperatively amplified by RAHB. A single backbone amide resonance variation in a beta sheet RAHB chain leads to changes in resonance throughout the chain, and the same inducement to variation at every backbone amide resonance as in the case of amyloid fibrils leads to a more considerable variation in RAHB throughout the chain. Also, the similarity of residue sequence of each strand does not permit any partial normalization of RAHB by the sidechains of adjacent strands.

Layering of the beta sheets of an amyloid filament so as to provide an even electrostatic environment for RAHB in the chains in each sheet will be important in finely balancing the RAHB. There is a tendency for amyloidogenic sequences to be hydrophobic which create little field but are more permissive of electrostatic field than non-hydrophobic residues and so they moderate electrostatic field mostly by increasing the distance between the backbone amides and causes of field.

Also, the chains run the length of the fibril and may exhibit significant RAHB. Elevated RAHB may increase the absolute differences between the RAHB of the chains.

It is noted that there are others factors which may contribute to stability of amyloid fibrils. Hydrogen bonding and dispersion between sheets may exclude water [187]. Where water is excluded, there are prospects for ions also being excluded.

4.10.3 Polyproline helices types I and II

Though there is no peptide group to peptide group hydrogen bonding in polyproline helices, the variation to the primary peptide resonance-type charge transfer in an electrostatic field will be stabilizing or destabilizing, since there is an energy change associated with the change in the primary peptide charge transfer itself. This energy change has almost always been neglected in considering the energetics of protein structure, and hydrogen bonding has been the focus instead. An electrostatic field parallel or antiparallel to the helix axis has a component in common with all C-N vectors in both type I and II polyproline helices, PPI and PPII, and is thus uniformly stabilizing or destabilizing at each peptide bond.

In PPII, the peptide bond dipole is substantially orthogonal to the helix axis. The C-O vector is very close to 90 degrees from the helix axis, with a variation of 120 degrees about the helix axis between subsequent peptide bond dipoles. This means that an electrostatic field parallel or antiparallel to the helix axis is substantially energetically neutral for the carbonyl bond dipoles, unlike alpha helices. The C-N bonds are ~50 degrees from the axis, the cosine being ~0.64 giving substantial component with C-N vector. The variation in peptide bond resonance has an intrinsic associated energy and also varies the propensity of the peptide group to participate in hydrogen bonding with residue side groups or water. Where the helix is solvated, an electrostatic field attenuates quickly with distance and such a helix will be short.

In the much less biologically significant, proline-requiring, right-handed PPI, there is an alignment of carbonyl bonds which resembles alpha helices, and the carbonyl bond dipoles and peptide resonance are both favoured or disfavoured by components of a helix-axis field.

Which type of polyproline helix is preferred may vary with the field strength, since the relation of C-N vectors to the helix axis and psi torsions differ between the two polyproline helix types.

The amide oxygen lone pairs are inequivalent (Chapter 5). In polyproline helices the amide oxygen p-type lone pair is available for hydrogen bonding rather than being protected as in beta sheets. Amide resonance and the binding energy associated with N(lp)->C-O(pi)* is maximised by C-O..H-N hydrogen bonding at C-O..N geometry 75 degrees from linear in the amide plane due to increased charge transfer from a lobe of the p-type lone pair (Chapter 5). Both lobes of the amide oxygen p-type lone pair are available for hydrogen bonding, and both lobes participating in hydrogen bonds will result in notably greater amide resonance than only one lobe. With an electrostatic field component aligned with the helix axis, a hydrogen bond involving one lobe will be favoured and the other disfavoured, since the charge transfer for the hydrogen bonds is varied by electrostatic field.

In summary, we predict that the structure of polyproline helices, lacking backbone RAHB chains, is determined by helix-axis electrostatic field causing increase of energy variation directly associated with the increase in primary peptide charge transfer and by increased RAHB with water or residue sidegroups. Also, greater access to the backbone amide oxygen p-type lone pair for HB by water allows increased amide resonance and hence binding energy associated with N(lp)->C-O(pi)*.

4.10.4 Protein folding

4.10.4.1 Overview

In a variation and extension of the backbone-based theory of protein folding of Rose *et al*. [3], it is proposed that the universal folding mechanism is peptide group resonance, PR, rather than inter-peptide group hydrogen bonding, IPHB. PR is integral to the backbone whereas IPHB is a backbone-backbone interaction, and so this proposal directly associates the universal folding mechanism with the backbone and widens the scope of the backbone-based folding mechanism to include peptide groups not participating in IPHB. The energy internal to a peptide group associated with its PR is variable with PR and is posited to drive protein folding. The causes of variation in PR are not limited to RAHB and EVPR-CN, and include C-N torsion, pyramidalization at N, the busy donor effect at the N lone pair and interactions with the carbonyl orbitals other than N(lp)->C-O(p)*. Since the binding energy of IPHB primarily resides in PR (Section 5.7.1.3), the energetics of PR subsume those of IPTB. PR is primary in determining the HB binding energy of a peptide group to other peptide groups, water or sidechains.

Each peptide group seeks to increase its PR. As peptide groups are distributed along a polypeptide chain and PR influences RAHB and orientation to electrostatic field, PR is a driver of backbone conformational change, in competition with hydrophobic interactions [52]. The difference in energy associated with variation in resonance of a single peptide group under biologically plausible conditions (the present Chapter and Chapter 5) is comparable to the energy of stabilization of proteins in native conformation (Section 4.10.4.2).

The amino acid sequence specifies fold, and the sequence is evaluated by the universal folding mechanism that is PR, whereas the universal mechanism of the Rose *et al*. proposal [3] does not directly heed sequence. In keeping with their proposal, there is no encoding of how to fold in amino acid sequences. Separation of the specification of fold from the procedure for how to fold is expected to confer great evolutionary advantage.

We refer to the present theory as the Resonance Theory of Protein Folding, RTPF. It includes variation of PR due to any cause, including EVPR-CN, RAHB, C-N torsion, pyramidalization at N, the busy donor effect at the N lone pair and interactions with the carbonyl orbitals other than N(lp)->C-O(p)*.

4.10.4.2 Peptide group resonance drives protein folding

In the present proposal, the amino acid sequence is evaluated by each peptide group in terms of its PR. This evaluation is not restricted to amino acids local in the sequence. Each peptide group seeks to increase its PR, though increase in one PR may come at the expense of another PR. This search for increased PR may appear to be cooperative or competitive depending on context, but can be understood as independent search by a peptide group of its changing environment. Conformational change due to hydrophobic interactions [52] or sidechain-sidechain interactions may also prevent any given PR or total PR from being monotonically increasing during folding. Patterning of residues according to hydrophobicity and sidechain-sidechain interactions can change the conformational energy landscape and introduce local free energy minima, thus frustrating folding [33, 34]. Folding frustration would tend to be minimized by evolution so that the driving mechanism of folding is not impeded, but this minimization is not itself the driving mechanism.

The variation in one PR due to biologically plausible electrostatic field is ~7 kcal/mol and variation due to RAHB is similar (Section 5.7), which overshadows the difference between IPHB and peptide group-water hydrogen bonding being ~1 kcal/mol [50, 51] and is comparable to the stabilization of a protein in its native conformation being 5-15 kcal/mol [188].

PR is posited to be primary in determining backbone conformation which it does by controlling RAHB and hence binding energy of IPHB or other HB involving that peptide group. Also, a peptide group will tend to orient its C->N to align with the local electrostatic field so as to its maximise PR. Peptide groups are distributed along a polypeptide chain and are a factor in determining backbone conformation regardless of whether the PR is part of IPHB secondary structure or not.

Chains of IPHB are resonance-assisted, and RAHB chains are cooperative. In the case of IPHB, the resonance is PR. A change in HB or resonance anywhere in an RAHB chain is accompanied by changes throughout the chain. In this sense, an RAHB chain evaluates changes to one of its HBs or resonant units in terms of all the HB and resonant units of the chain. Since the cooperativity of some IPHB chains is greater than others due to this variation of RAHB, IPHB chains of extensible secondary structures have different affinity for extension. Transient secondary structures with IPHB chains of higher RAHB are more likely to extend and those with lower RAHB are more likely to contract in the early stages of folding. Secondary structures are in competition for peptide groups during early folding. This does not mean that a peptide group has moved immediately from one secondary structure to another, rather that extension of one secondary structure can drag the polypeptide chain so that peptide groups are lost from an IPHB chain of another secondary structure.

When the steric exchange energy of an IPHB is deducted from that associated with the charge transfer of that IPHB, the result is very small (Section 5.7.1.3). The binding energy of an IPHB resides primarily in the resonances of the two peptide groups (Section 5.7.1.3). In this sense, PR subsumes IPHB.

The amino acid secondary structure preferences as categorized on the basis of chemical structure by Malkov *et al*. [189] influence PR in IPHB chains in specific backbone conformations in various ways, varying the resonance of the peptide groups throughout the IPHB chain. Strand preferring residues have unfavourable steric interactions with the backbone when in the alpha helical conformation [179], disrupting IPHB. Strand preference correlates with steric blocking by sidechain of backbone hydration in that preferred conformation [15], so alpha helical preferring residues in a beta sheet will weaken IPHB unless the surrounding residues are complementary in blocking backbone hydration. These effects are unfavourable to IPHB and hence RAHB and PR. A study of all the ways amino acid secondary structure preferences influence PR is needed.

4.10.4.3 Specification of fold separated from how to fold

In a purely declarative language [190], what is to be accomplished is described in a manner which says nothing about the means by which it is to be accomplished. This absence of procedure for arrival at solution from statements in the language is greatly simplifying of the language. The existence of a procedure for solution is assumed by the language and is universal to all expressions in the language. Rose *et al*. [3] refer to this separate procedure as a universal folding mechanism. This universal folding mechanism is the means by which a procedure for how to fold is not given in sequences of amino acid residues.

Each amino acid specifies variation in PR via modulation of RAHB of IPHB chains in secondary structure conformations and electrostatic properties including permittivity. All of these have a bearing on PR. The evaluation of a statement in this language, a sequence of amino acids, proceeds primarily by each peptide group searching for maximum resonance, and yields the polypeptide backbone fold.

4.10.4.4 Comparison with earlier backbone-based theory

The present theory, RTPF, can be stated as three principles:

(1) There is a binding energy associated with PR itself. This binding energy is primarily that of the amide/peptide donor-acceptor interaction N(lp)->C-O(p)*. In the absence of factors varying PR and hence the energy associated with PR, different DFT methods calculation of this energy range from about 50 to 100 kcal/mol. This is a large donor-acceptor interaction, and there are a number of known sources of variation of this donor-acceptor interaction: electrostatic field

with C-N component, hydrogen bonding, any degree of pyramidalization at N, any C-N torsion (amide non-planarity), the busy donor effect at N and interactions with the carbonyl orbitals other than N(lp)->C-O(p)*. All of these vary the same quantity - PR. Further, all of these interact through PR. Even though the primary amide donor-acceptor interaction is internal to the peptide group, this energy variation is as significant as energy variation of same magnitude elsewhere.

(2) The NBO interactions directly between the peptide groups engaged in IPHB (direct donor-acceptor interactions minus direct steric interactions) sum to very close to zero, with the IPHB binding energy residing in the (internal) peptide group resonance. This means that PR gives a good account of IPHB, and that most of the IPHB stabilization is subject to all other sources of variation of peptide group resonance.

(3) Sidechains influence all of the sources of variation of PR. The peptide group is distributed uniformly along polypeptide chains and each peptide group integrates the various sources of variation of PR as experienced by that group. The variation in PR energy at different backbone conformations is significant, and a driving role in backbone conformational change is anticipated. Residue sequence specifies PR value in a structure-dependent manner, then PR drives folding from there. How to fold does not need to be evolved for each new sequence.

A feature of the earlier theory is that the number of possible backbone conformations is constrained by the restriction on the number of alpha helices and beta strands with short links between them possible in any given length of amino acid residue sequence, a constraint referred to here as ABSL. ABSL would limit the search for protein fold to far fewer possible backbone conformations than is given by the Levinthal estimate. One of the subset of backbone conformations allowed by ABSL is selected as native structure by means involving sidechains. Assuming the length of alpha helices and beta strands is fixed at the observed average and that residue sequences linking them are short, there are about a thousand possible backbone folds for a residue sequence of length 100 [191]. A sequence length of 100 was used by Levinthal to estimate the number of conformations that need to be searched if independence of every residue’s phi and psi is assumed. Unlike the number estimated by Levinthal as demonstration of absurdity, the previous backbone theory gives that the possible backbone conformations may be exhaustively searched in plausible time.

Principles (1) and (2) provide inter alia a mechanism for secondary structures formation subject to backbone constraints such as ABSL, so no conflict between principles (1) and (2) and ABSL arises, though these principles do not assume a tendency to uniform length secondary structure elements with short links. However, ABSL tends to constrain the possible conformations to which principle (3) may drive the backbone, and the extent to which ABSL does this is likely to be decided on a sequence-by-sequence basis. Note that formation of secondary structure elements follows from RTPF, so ABSL may be viewed as an intermediate-level and approximate result in terms of RTPF. The length of links between alpha helices and beta strands is highly variable from native structure to native structure, and perhaps half the population of links is greater than 5 residues each (Figure 1A of [192]). The lengths of alpha helices and beta strands varies. It cannot be assumed for all sequences that alpha helices and beta strands are stable once formed and thus be fixed constraints during the remainder of folding. The length of alpha helices and beta strands may vary after initial formation during folding, including being reduced to zero so that the secondary structure element is transient. Search through different sets of alpha helices and beta strands may occur during folding and ABSL does not offer an account of this search whereas principle (3) does.

4.10.4.5 Observation

Support for this theory could be gathered by observing PR throughout folding and correlating it with other factors. Ideally, each PR would be observed throughout the folding of single molecules by physical methods. Increasing the demands on physical methods, these observations need to be compared with energy changes due to changes in hydrophobia, sidechain-sidechain interactions and entropy. Meeting these demands requires observing the structure of a single molecule throughout folding, a long sought development.

Classical calculations have no account of PR variation, and Quantum Mechanical Dynamics, QMD, calculations, have simulated times that are far short of the time to fold even the fastest folding protein, but progress in linear-scaling methods [193] is not abating. PR might usefully be observed during binding or limited conformational change by QMD. Attention is drawn to the large errors of established DFT methods in calculating the electron density associated with PR (Chapter 3). These errors are larger than the natural variation in resonance, and these methods are unsuitable for this purpose. Benchmarks are proposed for the development of more accurate DFT methods (Chapter 3).

When observation of the PR and C-N component of electrostatic field of each peptide group throughout folding becomes possible, either by physical experiment or by accurate QMD simulation of folding with quantum mechanical handling of every water molecule, these quantities might be shown on each peptide group in a 3D visualization of atoms and bonds. More immediately, such visualization is needed for QMD simulation of a few picoseconds e.g. a few thousand frames. In this visualization, a peptide group could be selected for a report of the sources of electrostatic field with C-N component at that peptide ordered by descending significance. The total peptide resonance and the Gibbs free energy would be shown per frame. The total of peptide resonance for each secondary structure type including unassigned would also be shown. Where the QMD method used can give resonance in terms of energy, kcal/mol might be used as units.

4.10.4.6 Summary

RTPF proposes that the universal folding mechanism is peptide groups seeking to increase their resonance. This search may have the appearance of cooperative or competitive behaviour depending on context, but can be understood as an independent search by peptide groups in their changing environment. This search does not yield monotonically increasing PR, for increase in some PRs may be associated with diminishment of other PRs, and hydrophobic and sidechain-sidechain interactions may alter the conformational landscape to diminish individual or total PR. Evolution of sequences to minimize frustration of folding via local minima in the folding free energy surface would still occur, but is not primary to folding. The driving mechanism of folding, energetic favour of increase in PR and hence RAHB, has complexity and kinetics unrelated to the Levinthal estimate [27].

The present variation and extension of the backbone-based theory of protein folding [3] moves the universal mechanism of folding from IPHB to the backbone per se and allows the common folding mechanism of IPHB secondary structure, polyproline helices and non-secondary structure peptides in folding to be seen. The energetics of PR subsume those of IPHB, since IPHB binding energy primarily resides in PR (Section 5.7.1.3). A peptide group's PR varies as its binding partners change. Resonance of sidechain groups does not drive folding as these groups are not integral to the backbone.

The amino acid sequence is a language which is interpreted by all peptide groups in terms of PR and backbone conformation. This language separates specification of fold from procedure for how to fold which is given by the universal folding mechanism, PR. This separation would confer great evolutionary advantage, since how to fold does not need to be evolved for each new sequence as how to fold is not encoded in the sequence.

Means of observing the resonance of every peptide group in a single molecule throughout folding is needed. QMD simulations could simulate enough time to capture changes to the resonance of peptide groups during binding or limited conformational change. Established DFT methods are unsuitable for this purpose (Chapter 3).

4.10.5 Molecular chaperones and protein complexes

When one protein binds to another at a largely hydrophobic interface, the hydrophobic interiors of the two proteins may be substantially joined, allowing the electrostatic field of one interior to extend into the interior of the other. Charged residues at the binding interface of either of the proteins will be desolvated on protein-protein binding, and if these are not paired to neutrality after protein-protein binding, will cause electrostatic field in the interiors of both proteins. We predict that hydrophobic protein-protein binding by a protein chaperone allows the chaperone to introduce a defined electrostatic field into the client protein thus inducing reorganization of the client. This may be a multi-step process in which in the first stage the chaperone provides an electrostatic field of sufficient magnitude in the hydrophobic environment that the client is forced to reorganize, a second step in which the chaperone no longer extends a field into the interior of the client thus allowing the client to reorganise according to its own electrostatic field and a third stage in which the chaperone detaches. Both pure electrostatics and electrostatic field variation of backbone amide resonance are effectual in this process.

4.10.6 Nitrogenous base pairing

We note features of nucleic acid bases that are likely also subject to electrostatic field variation of N(lp)->C-X(pi)*, where X is either O or N, charge transfer. In a pair of nitrogenous bases there are a number of bonds that may play the role of amide C-N in electrostatic variation of resonance. In these base pairings, wherever there is a nitrogen lone pair out of the plane of the rings and the atom bearing the lone pair is bonded to a carbon participating in another bond which has double-bond character, significant resonance-type charge transfer can be expected to occur and sensitivity to C-N component electrostatic field is to be anticipated. All of these interactions will modify the properties of the rings. Table 4.1 shows the energetics of the subset of these interactions which are closely involved with the base pairing from NBO's default vantage point of best or nominated Lewis picture.

Bonds 1 and 2 (both of Figure 4.15 and Figure 4.16) in both the guanine/cytosine, GC, pairing and the thymine/adenine, TA, pairing are antiparallel and part of cyclic hydrogen bonding between the bases. Cyclic RAHB will be limited by the least member of the cycle, so an overall weakening of the base-pairing hydrogen bonds could be expected in the presence of electrostatic field with component antiparallel to the C-N of one of these bonds. Also, the N(lp)->C-O(pi)* associated with bond 1 will have different sensitivity to the field than the N(lp)->C-N(pi)* associated with bond 2. The vectors of these bonds with respect to the nucleic acid helix differs between the cases of TA versus AT and similarly for GC versus CG. A molecule which moves along the helix and which has an associated electrostatic field to which TA versus AT and GC versus CG bond 1 and bond 2 vectors are not symmetric will vary inter-base hydrogen bonding differently in accord this asymmetry. Also, electrostatic field aligned with the major groove is reversed by a field is aligned with the minor groove.

The protonated nitrogen on the thymine ring directly involved in inter-base hydrogen bonding has two N(lp)->C-O(pi)* interactions, making it a busy donor. Only one of those interactions is part of cyclic RAHB involving the two base-pairing hydrogen bonds, so an electrostatic field with the C-N vector of the other interaction, bond 3 (Figure 4.16), could be expected to diminish cyclic RAHB, again with the consequence of facilitating base-pair opening.

The nitrogen of cytosine that would connect to the helix backbone has a N(lp)->C-O(pi)* interaction, across bond 4 (Figure 4.15), is involved in non-cyclic inter-base hydrogen bonding and can be expected to vary the ring properties of a paired base.

Variation in electrostatic field in nucleic acid polymers occurs when the negatively charged phosphates of the nucleic acid backbone are permitted to attract positive charge, or a charged molecule binds to the major or minor groove of the nucleic acid helix.

Variation of these resonances by electrostatic field may vary the binding energy of base pairing, and hence the energetic barrier of base pair opening for nucleic acid strand separation.

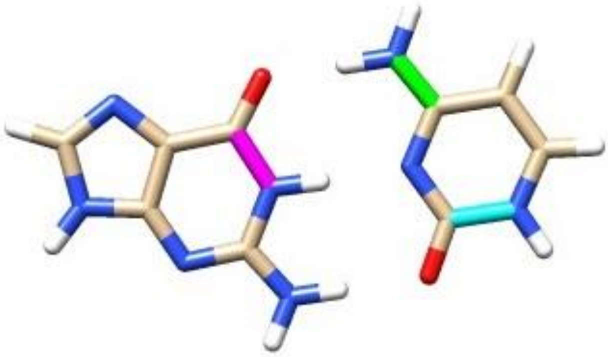


Figure 4.15. Guanine/Cytosine Pairing. Bond 1 magenta, Bond 2 chartreuse, Bond 4 Light Blue. Major Groove Top, Minor Groove Bottom.

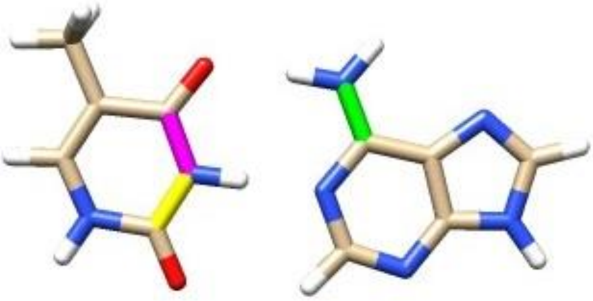


Figure 4.16. Thymine/Adenine Pairing. Bond 1 magenta, Bond 2 chartreuse, Bond 3 yellow. Major Groove Top, Minor Groove Bottom.

Table 4.1. Selected N(lp)->C-X(p)* Interactions in GC and AT Base Pairs at LC-wPBE(w=0.4)/6-311++G**

Base Pair	N-C Bond ID	Type of X	N(lp)->C-X(p)* kcal/mol
GC	1	O	78.38
AT	1	O	87.42
GC	2	N	126.65
AT	2	N	113.26
AT	3	O	88.07
GC	4	O	70.21

4.11 Conclusion

Amide resonance is sensitive to electrostatic field with component parallel or antiparallel to the C-N vector. A field of 0.000238 au is sufficient to vary the primary amide resonance-type charge transfer by 0.001 electron, a level we nominate as the threshold of chemical significance. The charge transfer as calculated by both DFT and wavefunction methods increases linearly with field magnitude in the range -0.005 to 0.005 au without threshold. A variation in field between -0.005 and 0.005 au gives a 0.042 electron variation in primary amide charge transfer in N-methylformamide and a 0.2 electron variation in charge transfer is calculated between the cases when the two monovalent ions, Li+ and F-, are placed on the N-C line with one 4 angstroms from C further from N and the other 4 angstroms from N further from C as compared with the ions being swapped in those positions to reverse the field. This arrangement is not offered as biologically plausible but shows that a 0.2 electron variation in the primary amide charge transfer is physically possible though is not the full extent of what is physically possible. Electronic structure of the amide group is not particularly multi-reference (Chapter 3) and a field of magnitude 0.005 au causes negligible change to this (data not shown).

If an amide is participating in a protein secondary structure RAHB chain, change in the amide's resonance causes variation in hydrogen bonding in the chain. Electrostatic field with direction that increases the resonance of an amide in the chain without directly diminishing hydrogen bonding stabilizes the RAHB chain. This hydrogen bonding also involves charge transfer from lone pairs, and its charge transfer is varied less in absolute terms at a given field magnitude than that primary to amide resonance. Also, with the electrostatic field orthogonal to H-N so the hydrogen bonding resonance-type charge transfer is not varied with O..H-N angle constrained, the field has a substantial component parallel or antiparallel to the C-N vector.

A consequence of EVPR-CN is the change in energy associated with change in the primary amide charge transfer itself which is irrespective of the amide bonding context, which may be considerable and must then be taken into account when studying the energetics of protein structure.

We discussed stability of beta sheets and non-polyproline helices due to electrostatic field variation of backbone amide resonance, mentioned a connection to conformational and allosteric change and implications for unstructured regions, and offered hypothesis concerning stability of beta sheets, amyloid fibrils and both types of polyproline helices, and of the function of protein chaperones. We predict analogous stability considerations for nitrogenous base pairing in an electrostatic field, and offer a hypothesis in this regard.

It may be that other factors such as hydrophobia, entropy and pure electrostatics are sufficient to determine a fold of some proteins. However, it would be remarkable if there were any protein that provided an electrostatic environment for all its backbone amides such that there was negligible field component parallel or antiparallel to the C-N vector environmental to the amide group. Wherever such a field exists, it necessarily changes amide resonance. The significance of the effect and its consequences for protein folding and conformational change need to be placed in relation to that of other effects influencing protein folding and conformational change. However, this variation of backbone amide resonance is directionally sensitive to the electrostatic field created and permitted by amino-acid residue sidechains, and offers a novel mechanism for the relationship of amino-acid residue sequence and protein fold [22]. This mechanism is a more precise specification of structure than less directionally-sensitive hydrophobia, but is facilitated by hydrophobia's creation of a low permittivity protein interior.

Rose et al. [3] proposed that protein folding is backbone-based and that backbone hydrogen bonding is a universal folding mechanism, but do not propose direct variation of this mechanism by residue sidechains. Variation of backbone amide resonance by electrostatic field is a backbone-based mechanism, with electrostatic properties of residue sidechains directly varying backbone amide resonance, hence varying backbone hydrogen bonding. We proposed the Resonance Theory of Protein Folding (Section 4.10.4) in which protein folding is driven by PR, where PR is varied by any effect including EVPR-CN, RAHB, C-N torsion, pyramidalization at N, the busy donor effect at the N lone pair and interactions with the carbonyl orbitals other than N(lp)->C-O(p)*.

4.12 Acknowledgements

Prof. John A. Carver is acknowledged for reading this manuscript and offering editing suggestions.

eResearch South Australia is acknowledged for hosting and administering machines provided under Australian Government Linkage, Infrastructure, Equipment and Facilities grants for Supercomputing in South Australia, directing funds to the acquisition of Nvidia Tesla GPU nodes and allocating 64 CPU cores and 256 GB RAM of the NeCTAR Research Cloud (a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy) to the present work.

4.13 Appendix 1

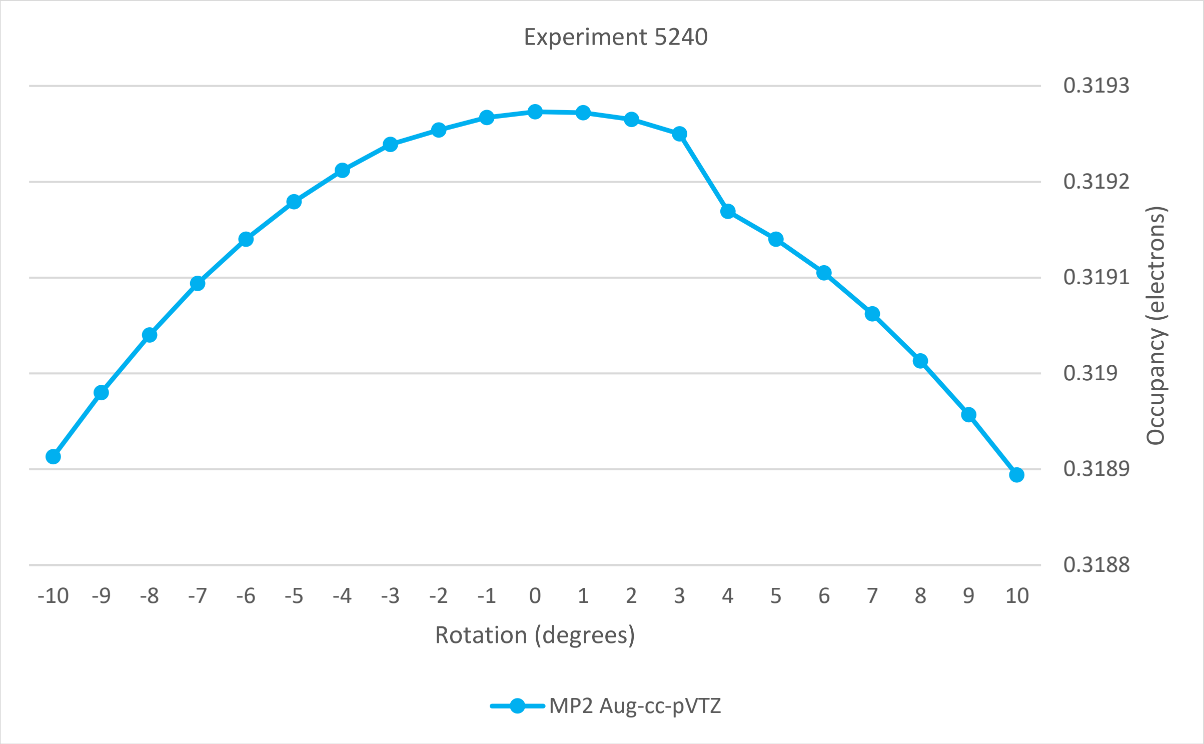


Figure 4.17. C-O(pi)* NBO Occupancy in N-methylformamide in 0.005 au Uniform Electrostatic Field Rotated in O-C-N Plane Starting from C-N Vector

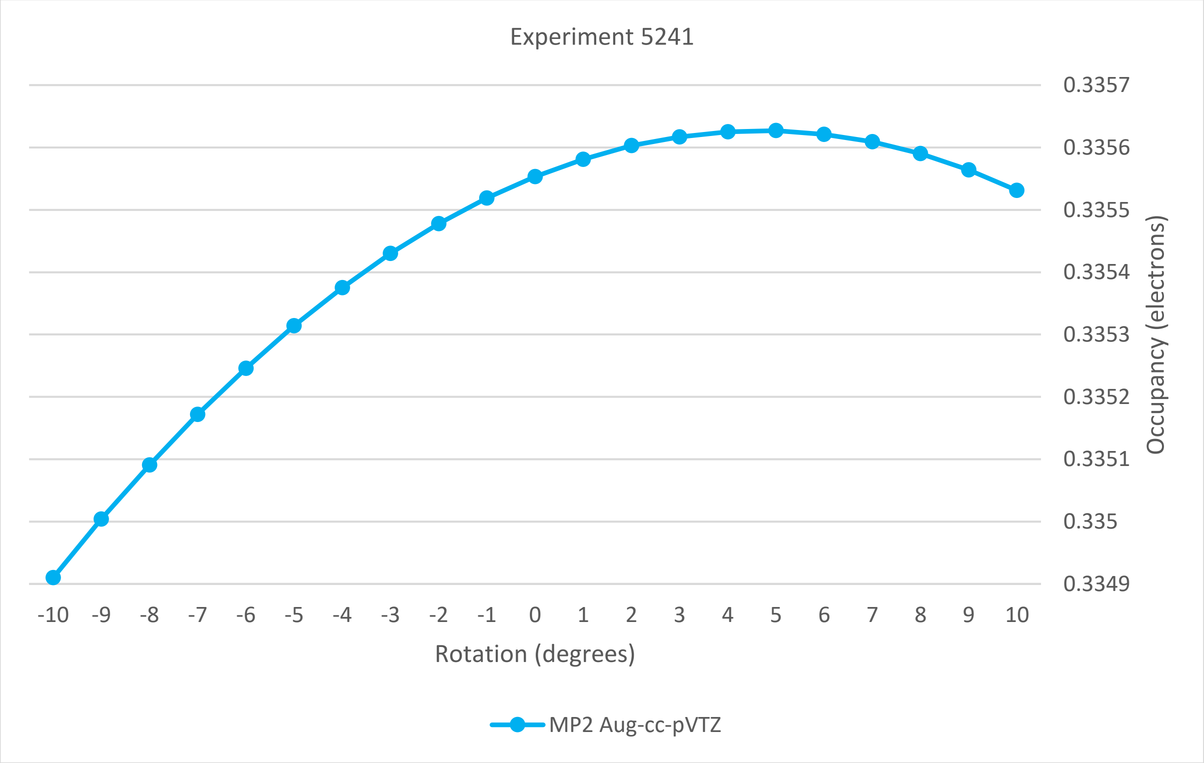


Figure 4.18. C-O(pi)* NBO Occupancy in N-methylethanamide in 0.005 au Uniform Electrostatic Field Rotated in O-C-N Plane Starting from C-N Vector

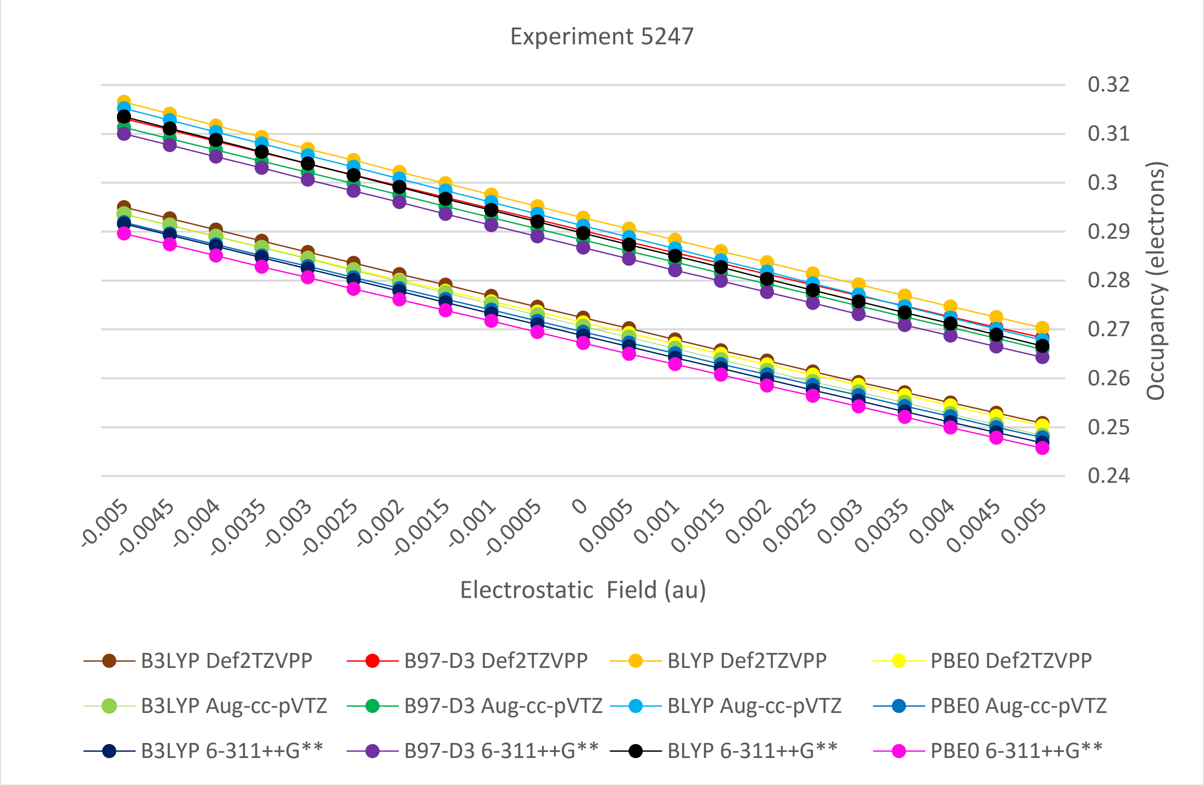


Figure 4.19. C-O(pi)* NBO Occupancy in N-methylformamide in Uniform Electrostatic Field with N-C Vector

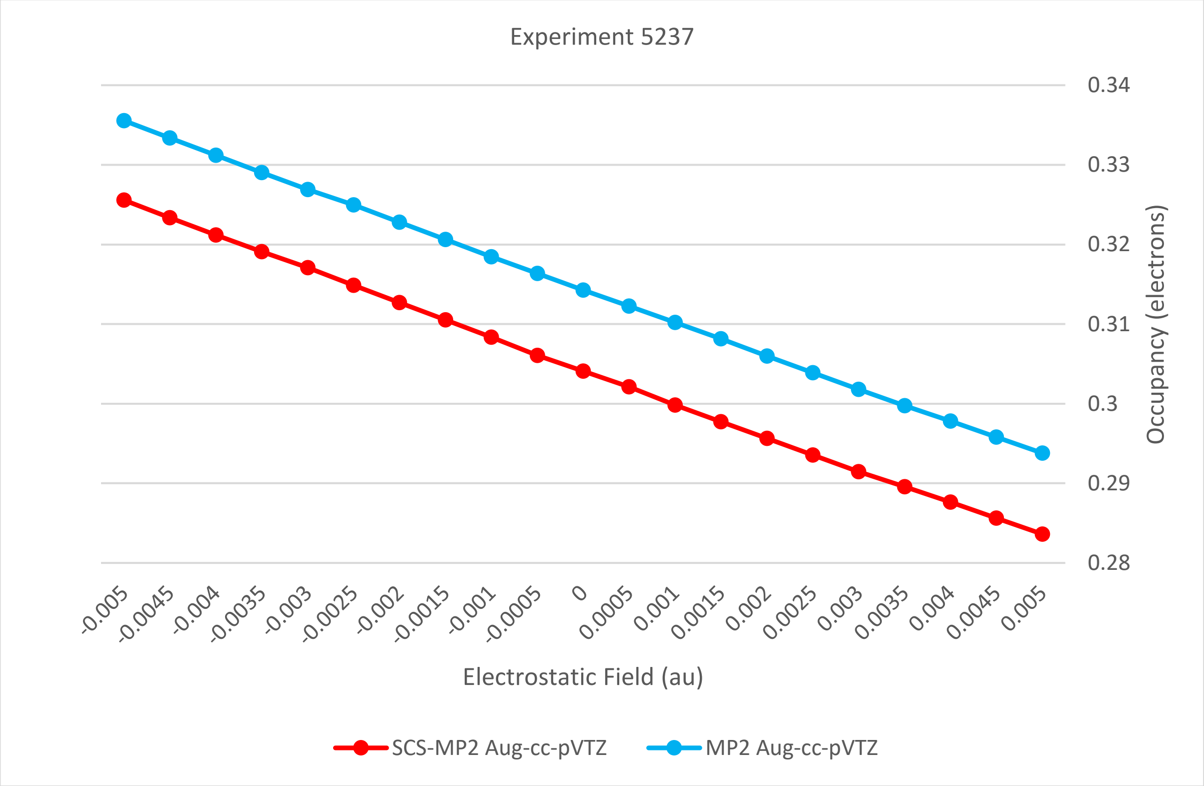


Figure 4.20. C-O(pi)* NBO Occupancy in N-methylethanamide in Uniform Electrostatic Field with N-C Vector

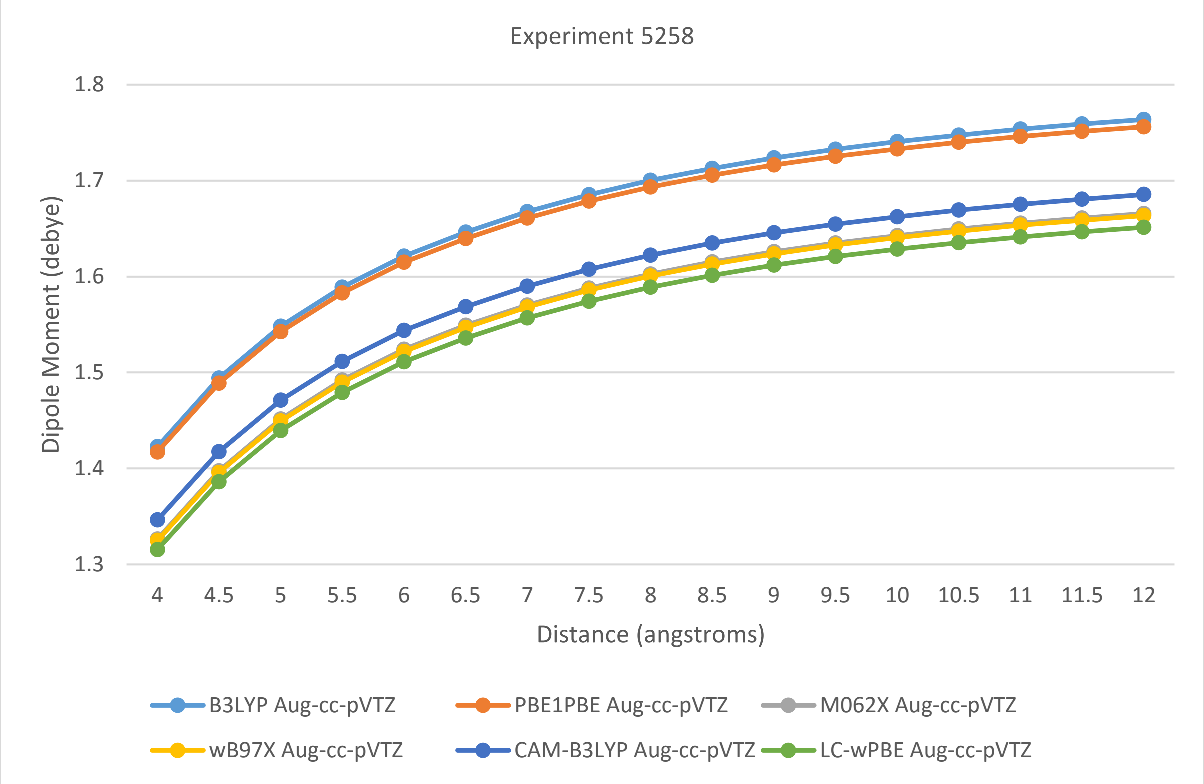


Figure 4.21. Dipole Moment of N(lp) NLMO in N-methylformamide with Li+ on C-N Line at Distance from N

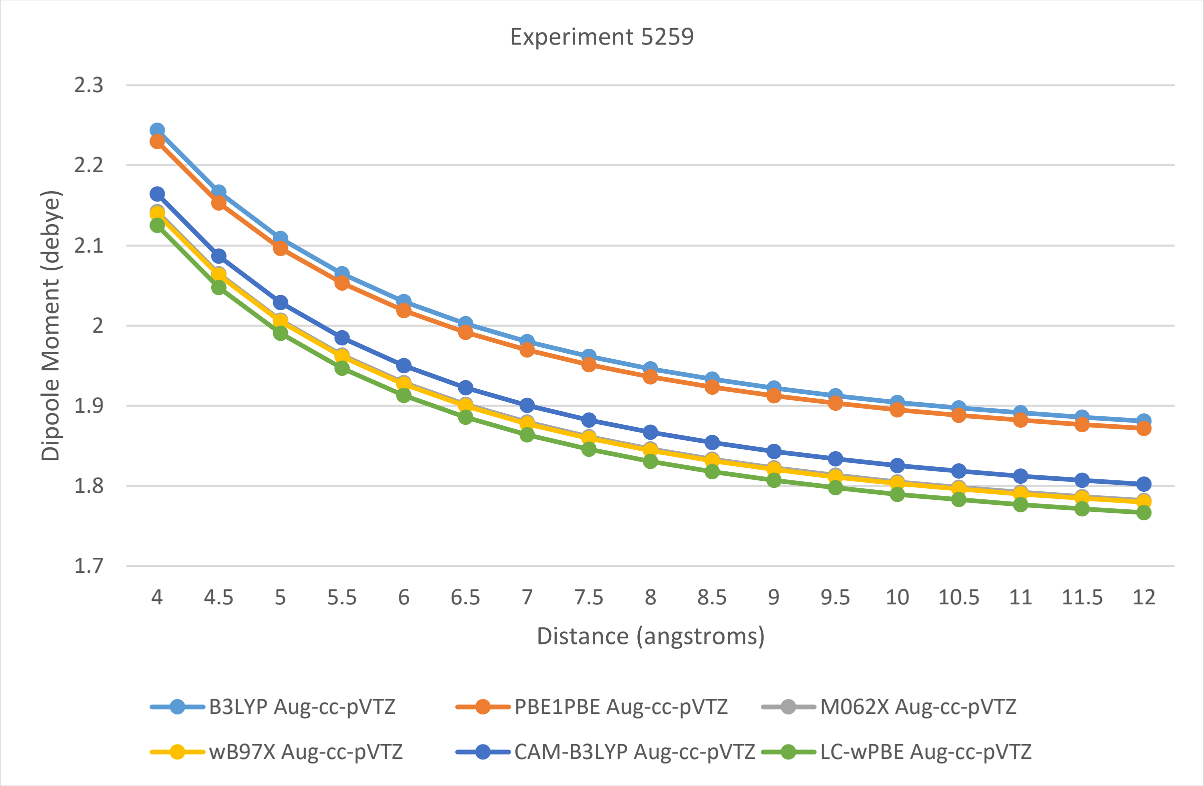


Figure 4.22. Dipole Moment of N(lp) NLMO in N-methylformamide with F- on C-N Line at Distance from N

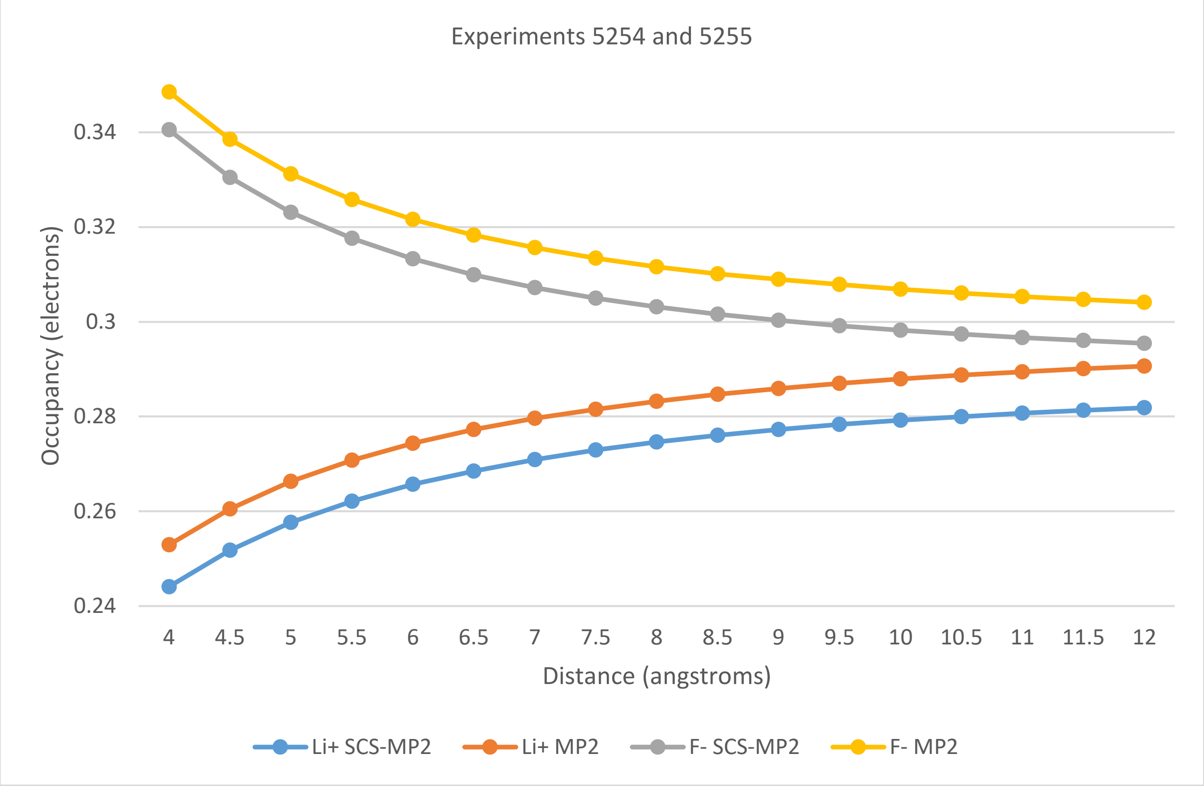


Figure 4.23. C-O(pi)* NBO Occupancy in N-methylformamide with Ion on C-N Line at Distance from N with aug-cc-pVTZ

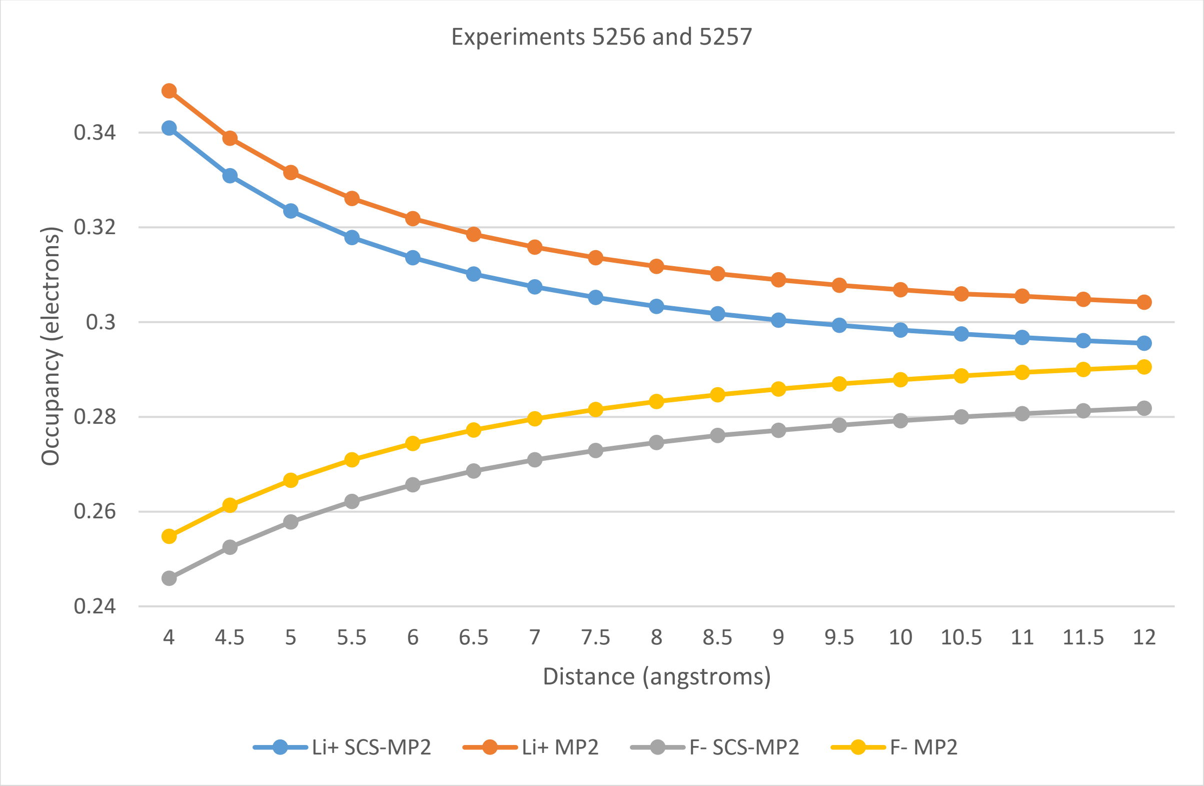


Figure 4.24. C-O(pi)* NBO Occupancy in N-methylformamide with Ion on N-C Line at Distance from C with aug-cc-pVTZ

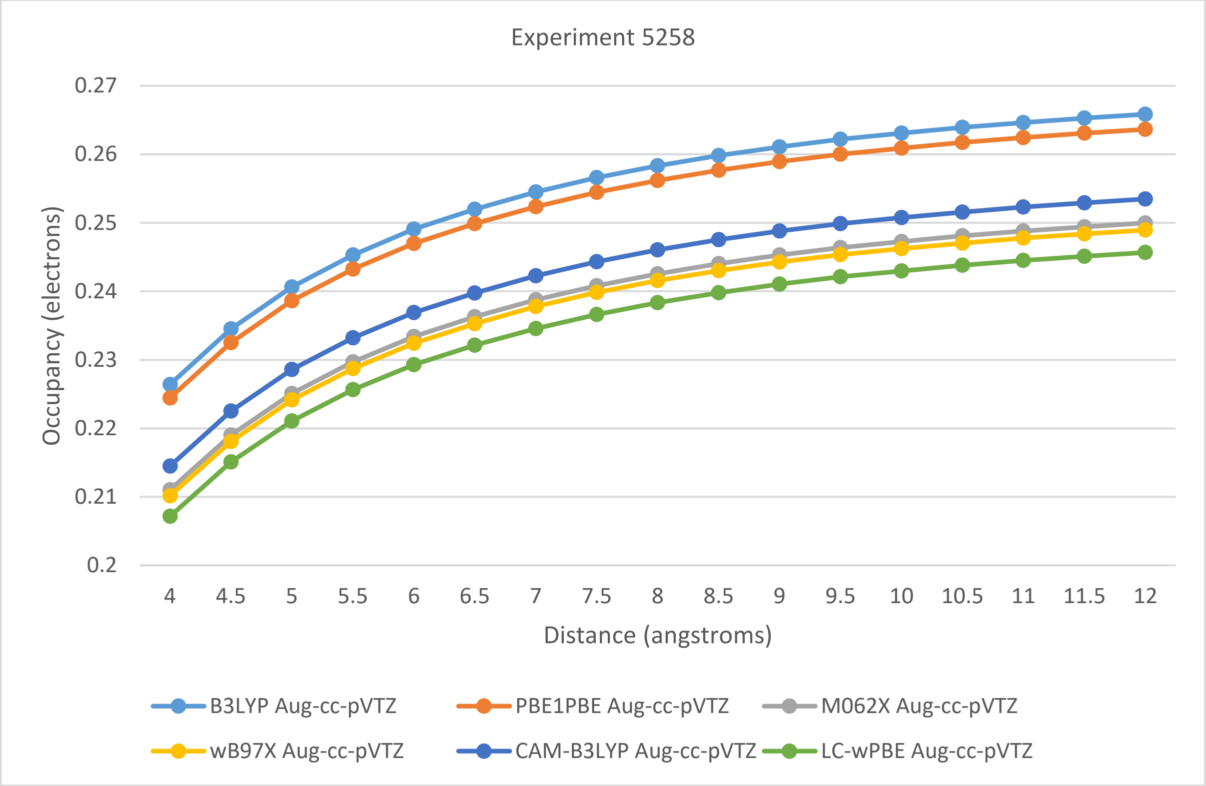


Figure 4.25. C-O(pi)* NBO Occupancy in N-methylformamide with Li+ on C-N Line at Distance from N using Gaussian SP over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

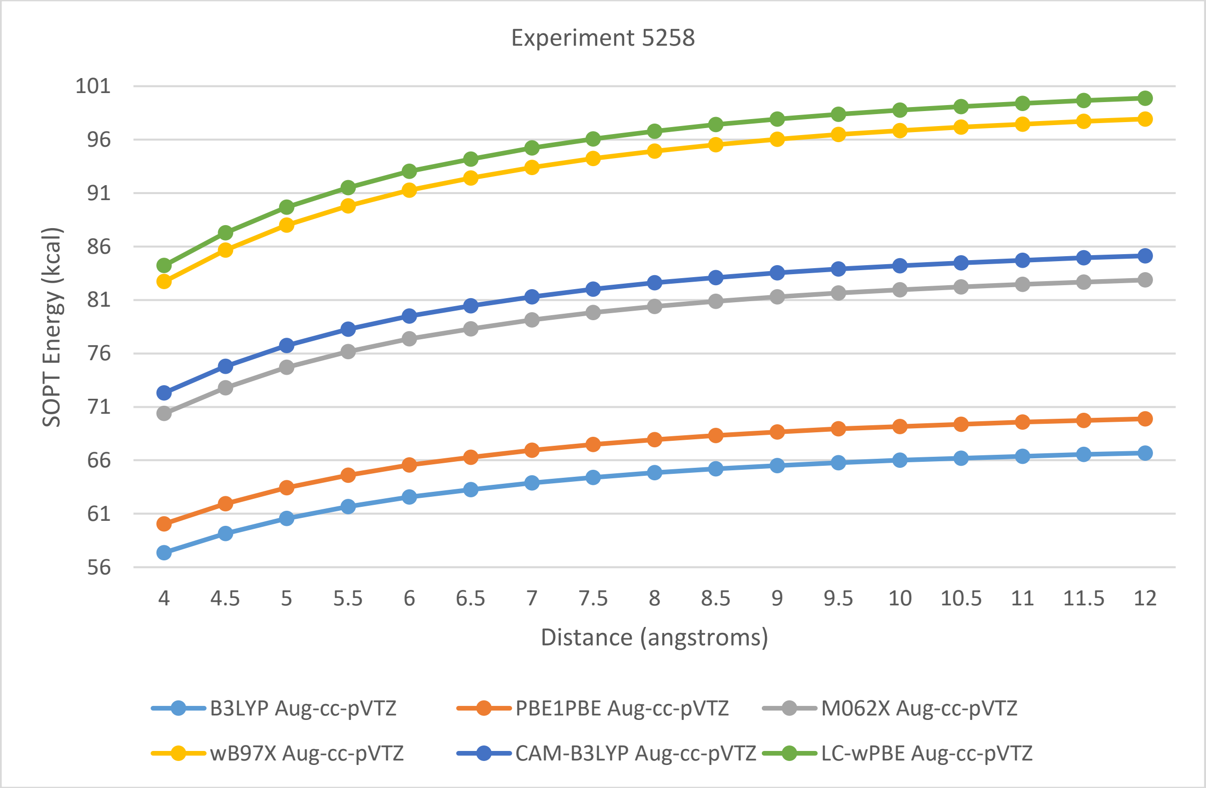


Figure 4.26. N(lp)->C-O(pi)* SOPT Energy in N-methylformamide with Li+ on C-N Line at Distance from N using Gaussian SP over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

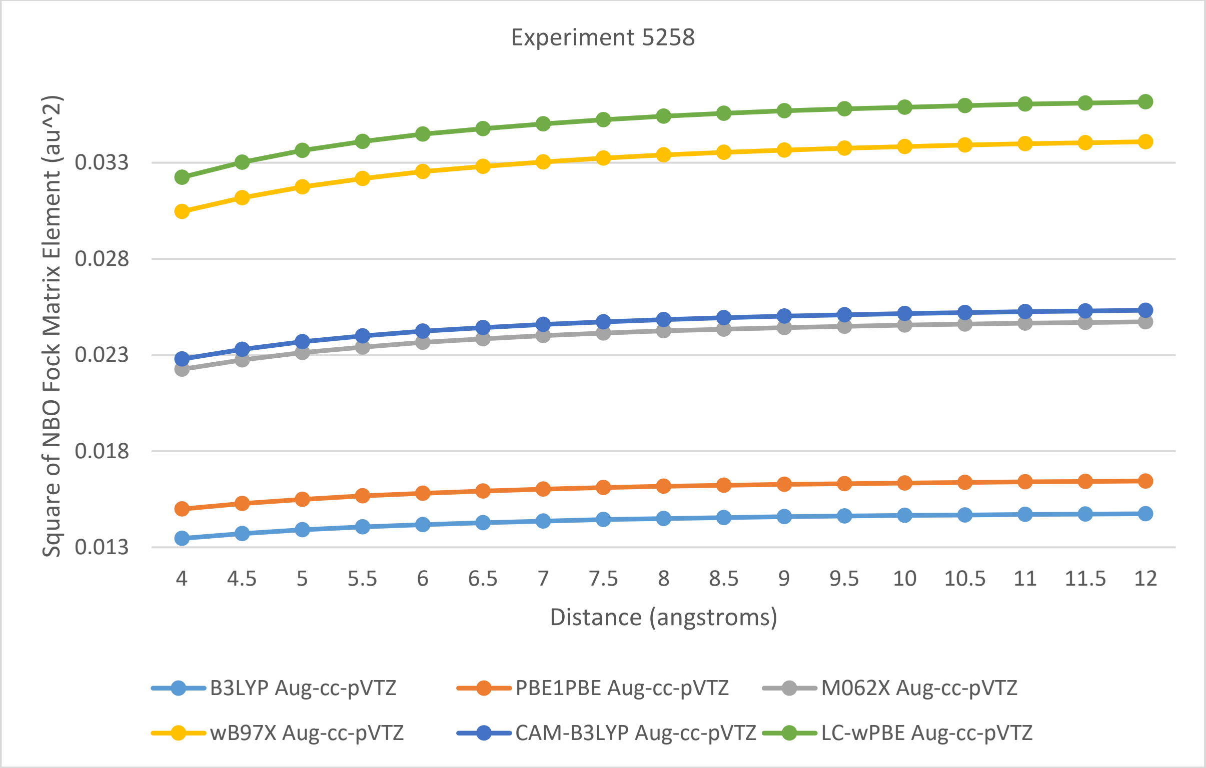


Figure 4.27. Square of NBO Fock Matrix Element for N(lp)->C-O(pi)* in N-methylformamide with Li+ on C-N Line at Distance from N using Gaussian SP over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

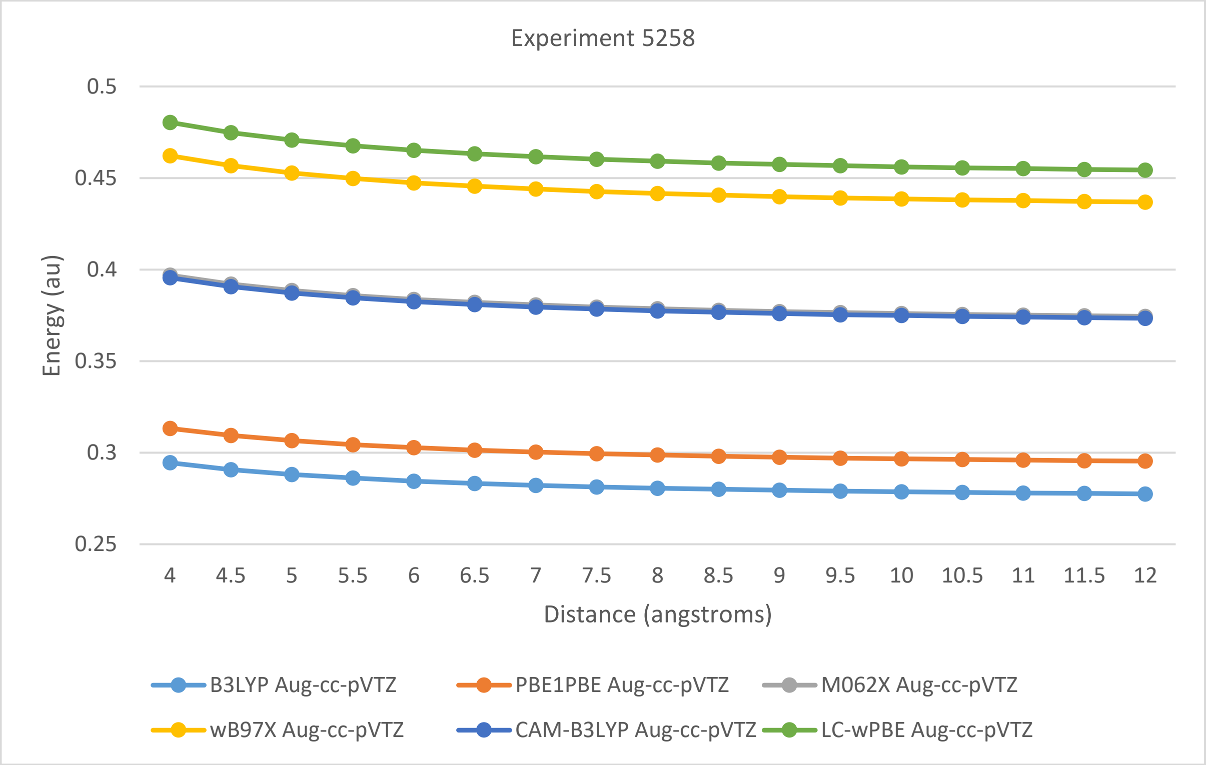


Figure 4.28. Energy Level of C-O(pi)* Minus Energy Level of N(lp) in N-methylformamide with Li+ on C-N Line at Distance from N using Gaussian SP over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

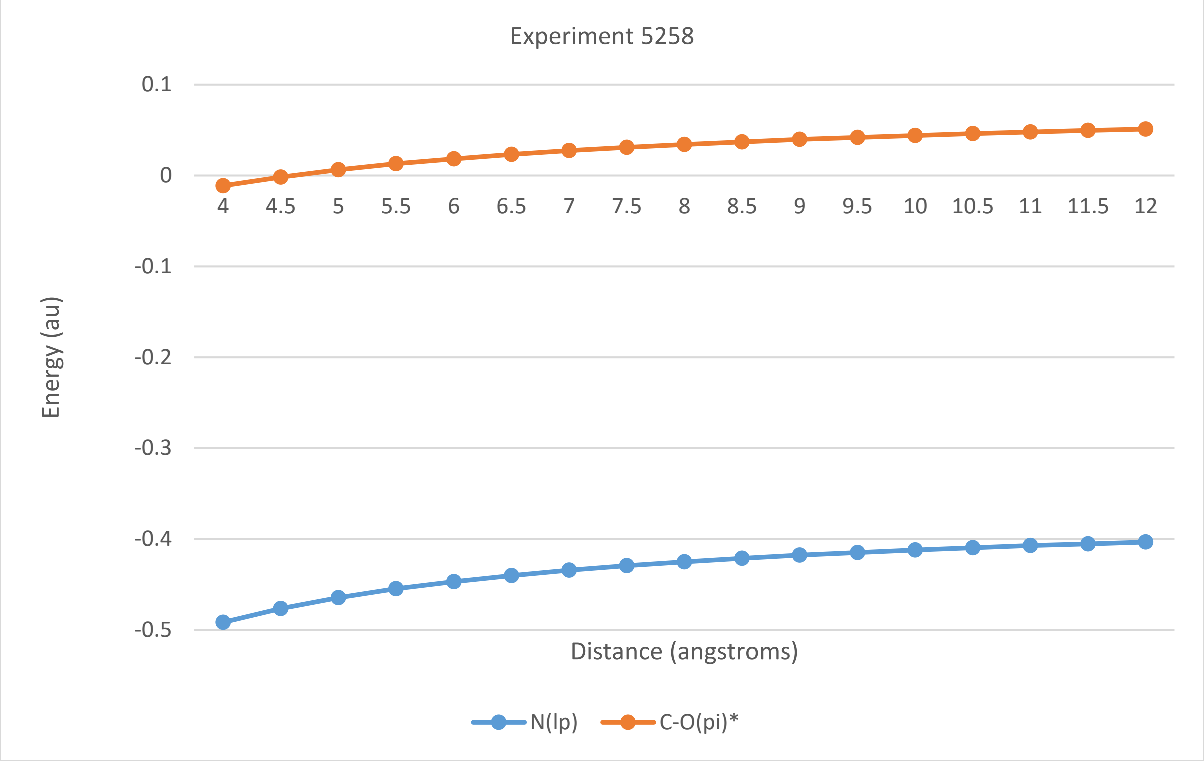


Figure 4.29. Energy Level of NBOs in N-methylformamide with Li+ on N-C Line at Distance from C using Gaussian SP over Coordinates Optimized with Orca at SCS-MP2/aug-cc-pVTZ

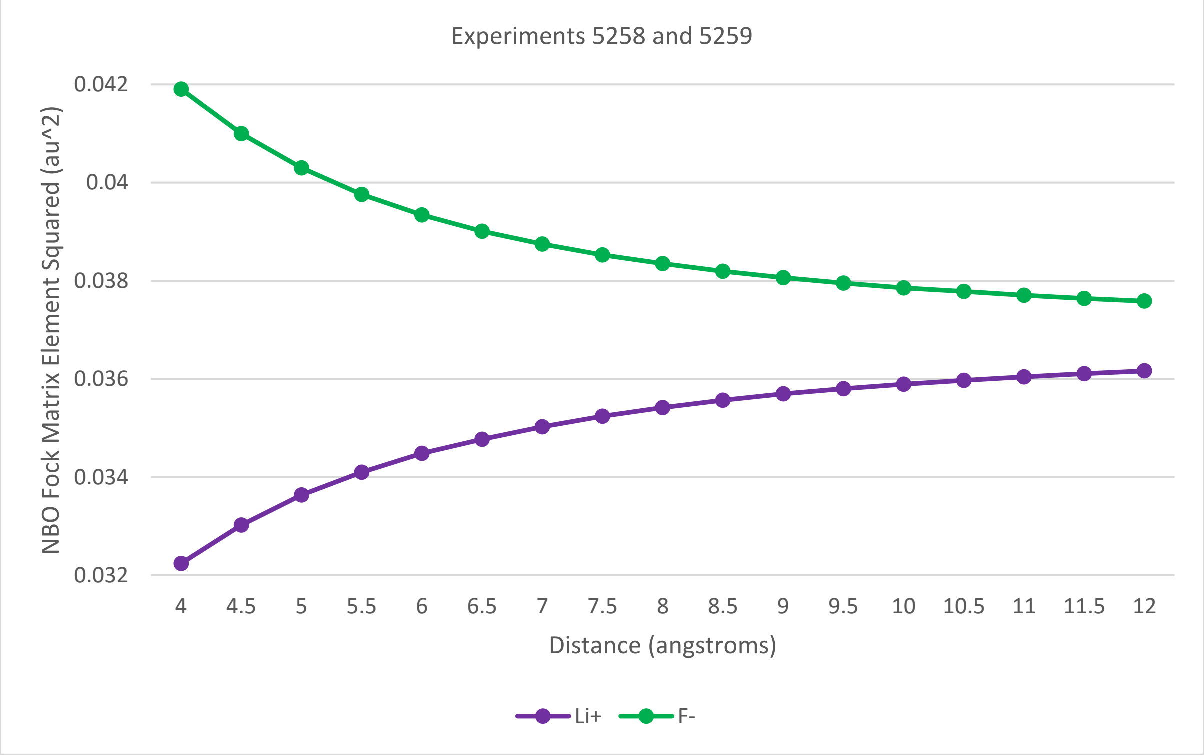


Figure 4.30. Square of NBO Fock Matrix Element for N(lp)->C-O(pi)* in N-methylformamide with Ion on C-N Line at Distance from N with Gaussian SP LC-wPBE/aug-cc-pVTZ over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

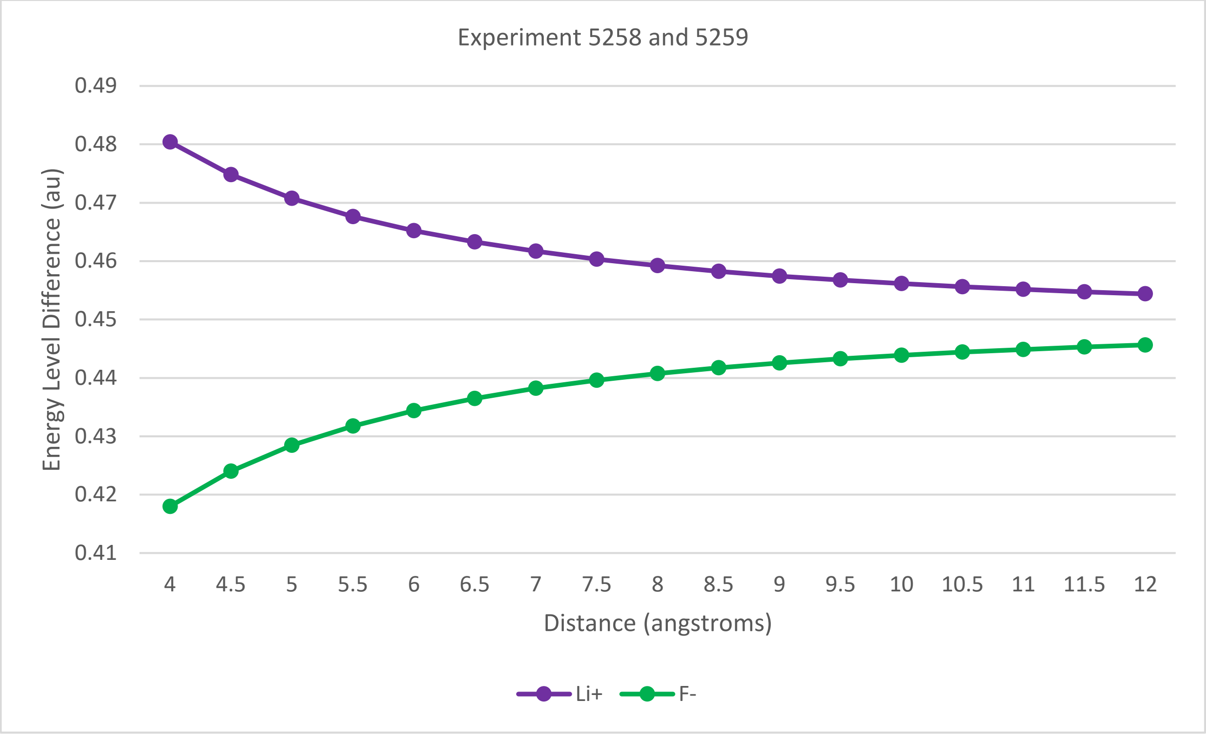


Figure 4.31. Energy Level of C-O(pi)* Minus Energy Level of N(lp) in N-methylformamide with Ion on C-N Line at Distance from N using Gaussian SP LC-wPBE/aug-cc-pVTZ over Coordiantes Optimized by Orca at SCS-MP2/aug-cc-pVTZ

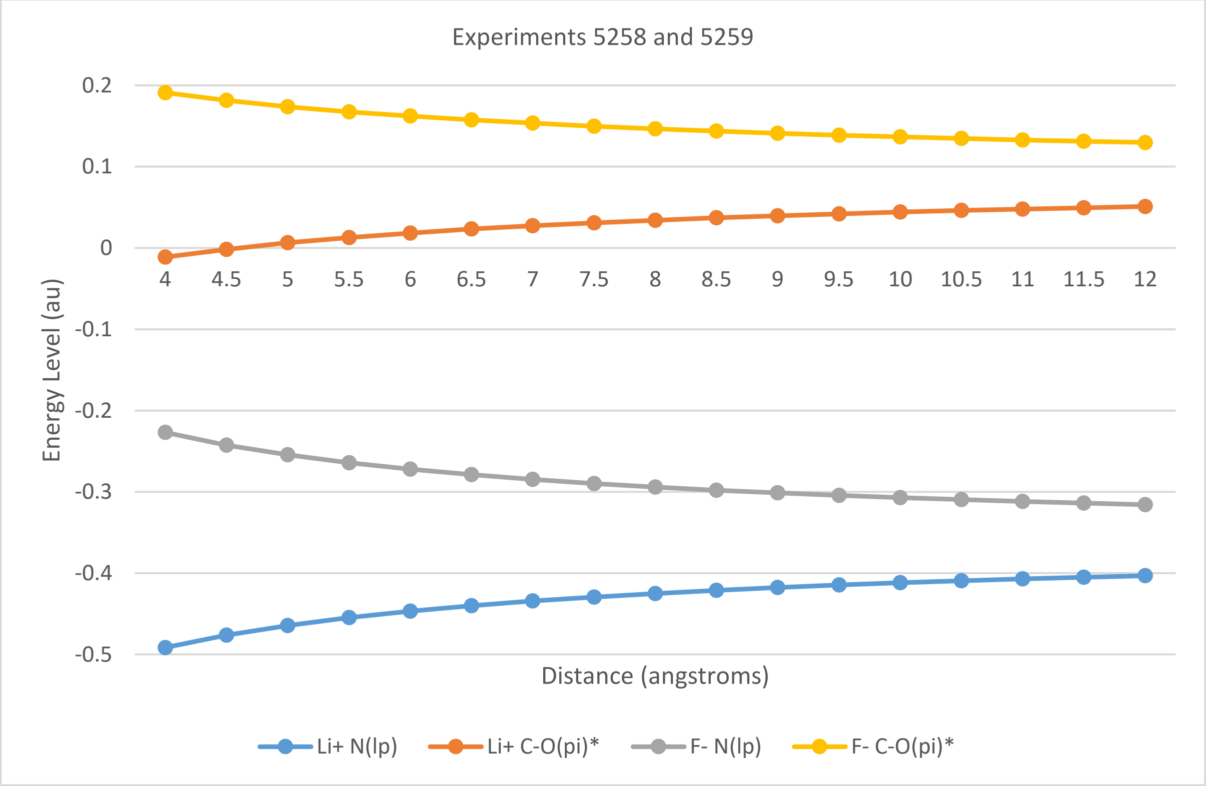


Figure 4.32. Energy Level of NBOs in N-methylformamide with Ion on C-N Line at Distance from N with Gaussian SP LC-wPBE/aug-cc-pVTZ over Coordinates Optimized by Orca at SCS-MP2/aug-cc-pVTZ

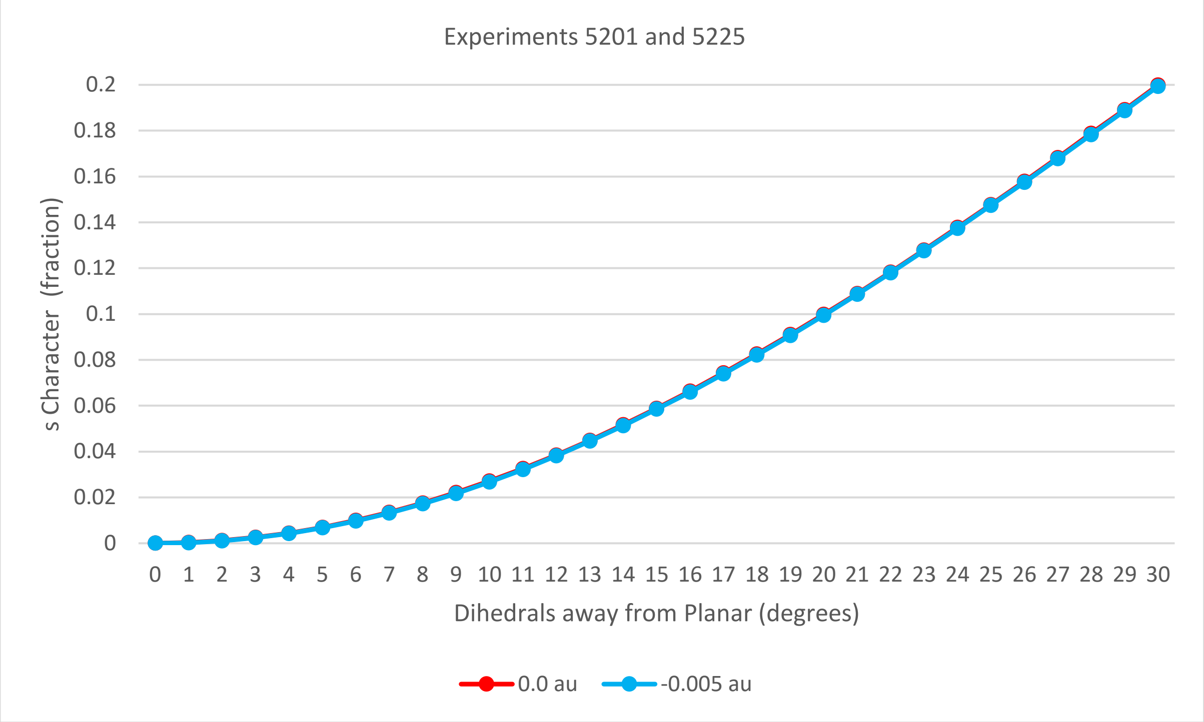


Figure 4.33. N(lp) NBO s Character in N-methylformamide with O-C-N-CA and O-C-N-H Dihedrals Constrained to Angles Away from Planarity toward sp3 Geometry at N in O-C-N Normal Uniform Electrostatic Field 0.0 and -0.005 au at SCS-MP2/aug-cc-pVTZ

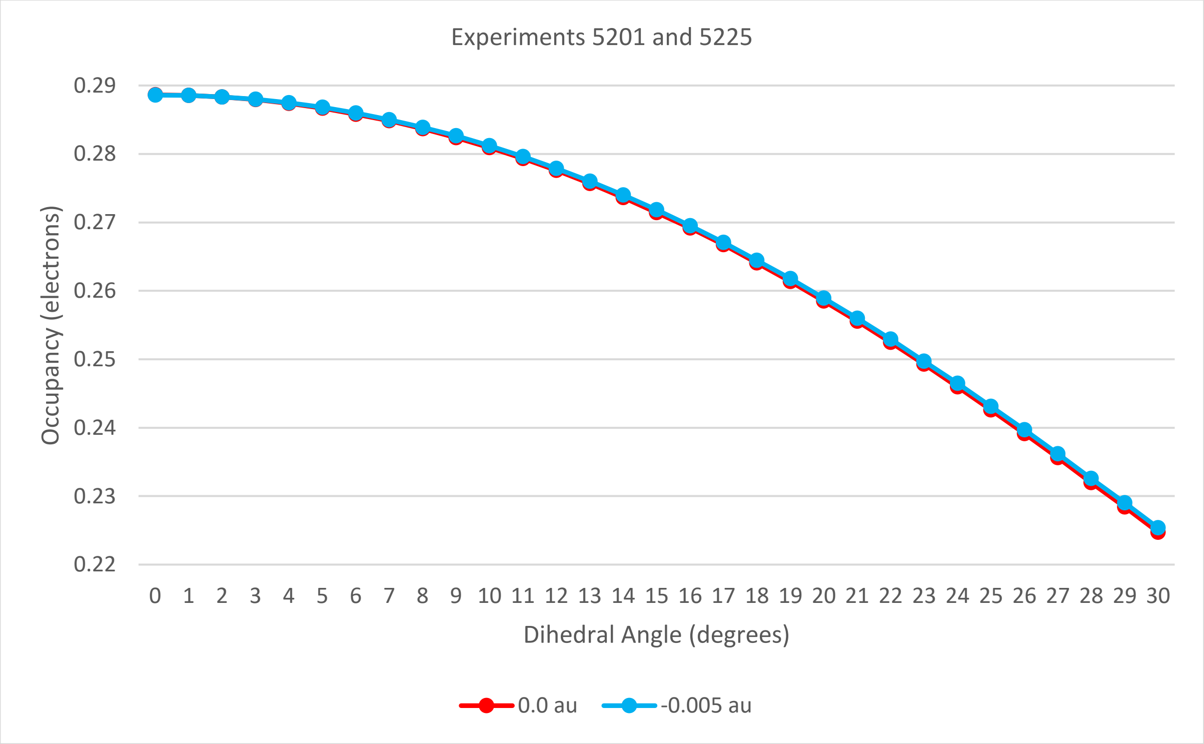


Figure 4.34. C-O(pi)* NBO Occupancy in N-methylformamide with O-C-N-CA and O-C-N-H Dihedrals Constrained Away from Planar Toward sp3 Geometry at N in O-C-N Normal Uniform Electrostatic Field of 0.0 and -0.005 au at SCS-MP2/aug-cc-pVTZ

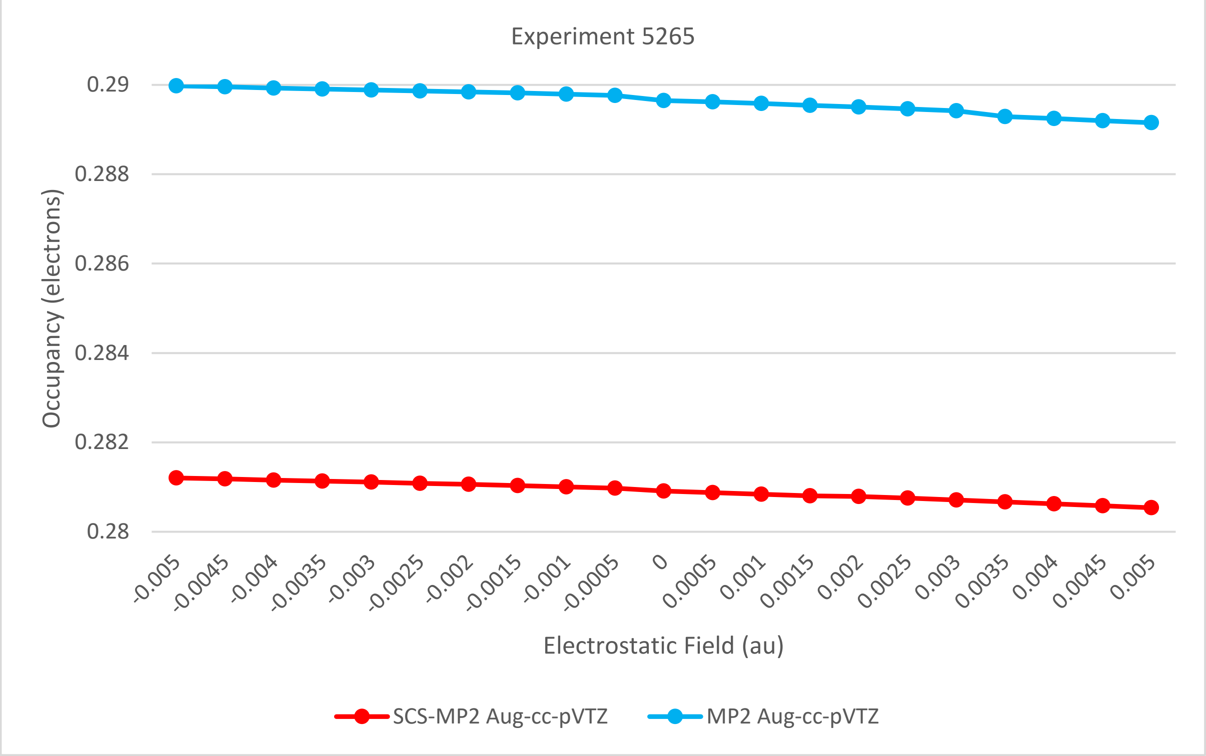


Figure 4.35. C-O(pi)* NBO Occupancy in N-methylformamide with O-C-N-CA and O-C-N-H Dihedrals 10 Degrees from Planar Toward sp3 geometry at N in O-C-N Normal Uniform Electrostatic Field of 0.0 and -0.005 au at SCS-MP2/aug-cc-pVTZ

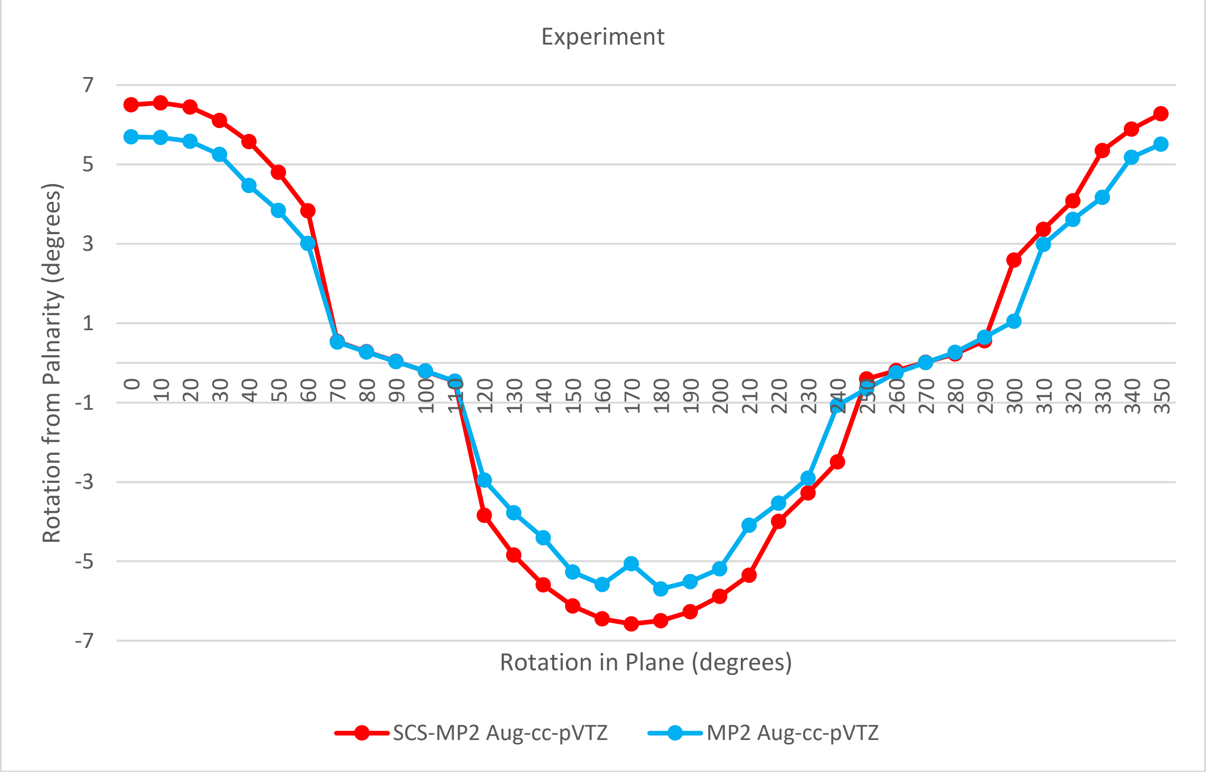


Figure 4.36. O-C-N-H Dihedral from Planarity in N-methylformamide in 0.005 au Uniform Electrostatic Field Rotated in Plane Containing N-C and O-C-N Normal Starting from O-C-N Normal

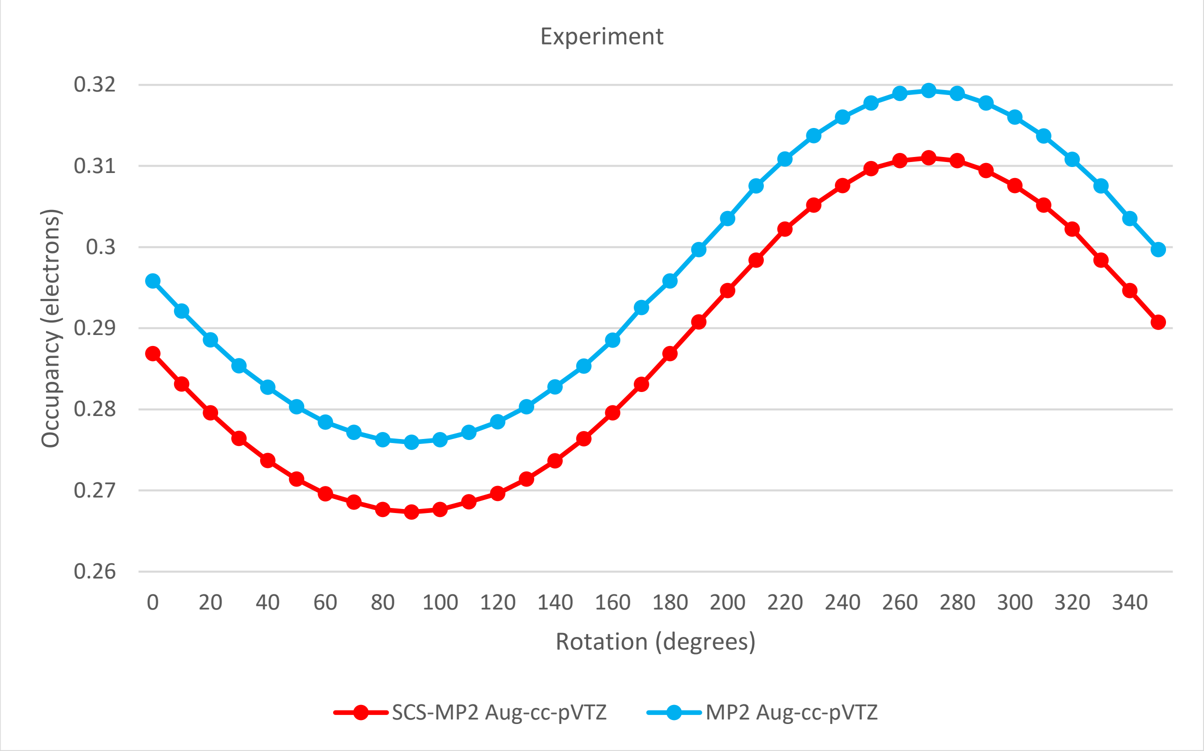


Figure 4.37. C-O(pi)* NBO Occupancy in N-methylformamide with Unconstrained H, CA and HA in a 0.005 au Uniform Electrostatic Field Rotated in Plane Containing N-C and O-C-N Normal Starting from O-C-N Normal

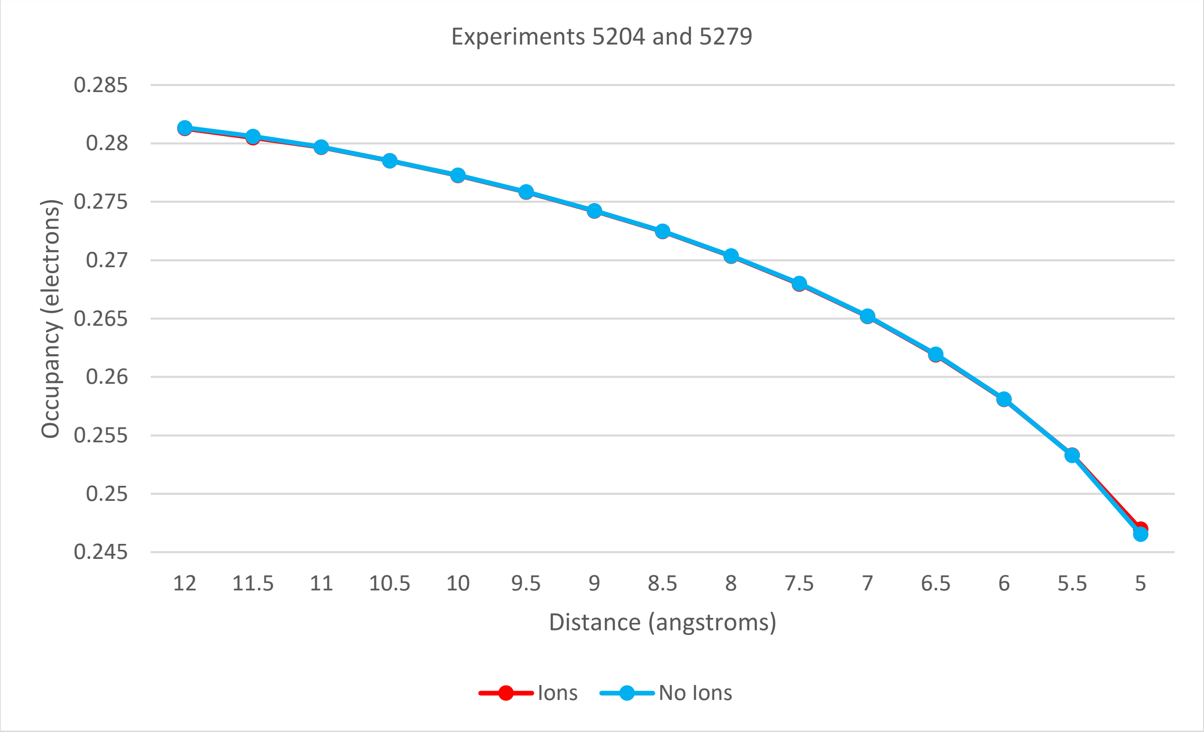


Figure 4.38. C-O(pi)* NBO Occupancy in N-methylformamide at O-C-N-CA and O-C-N-H Dihedrals Resulting from Li+ and F- On Opposite Sides Equidistant from N on O-C-N Normal with Ions Present and Ions Removed at SCS-MP2/aug-cc-pVTZ

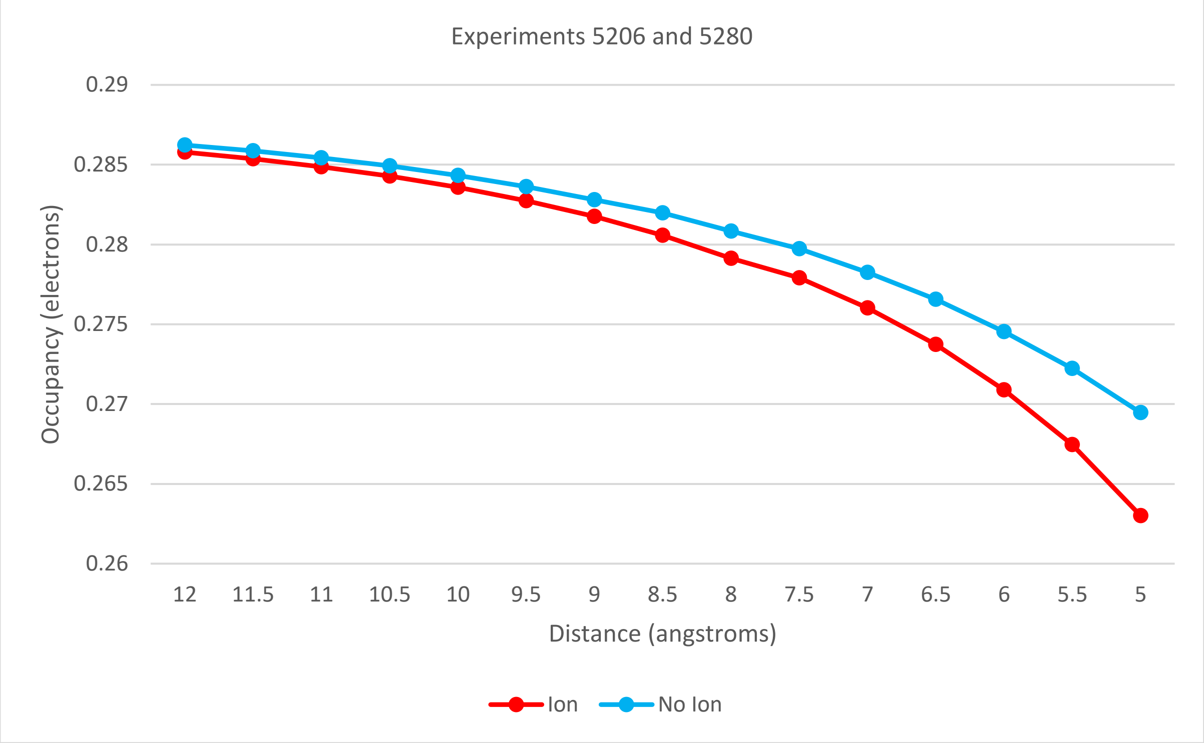


Figure 4.39. C-O(pi)* NBO Occupancy in N-methylformamide with O-C-N-CA and O-C-N-H Dihedrals Constrained to Result of Li+ on O-C-N Normal at Distance from N with Ion Present and Ion Removed at SCS-MP2/aug-cc-pVTZ

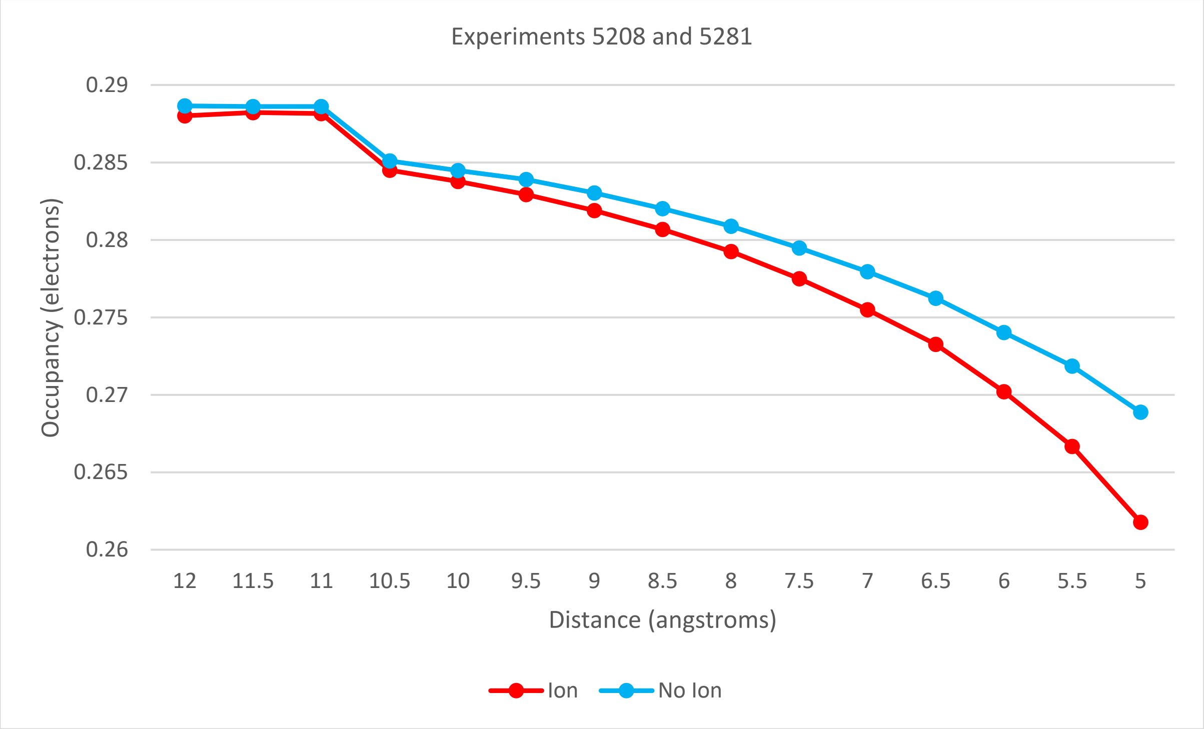


Figure 4.40. C-O(pi)* NBO Occupancy in N-methylformamide with O-C-N-CA and O-C-N-H Dihedrals Constrained to Result of Ammonium Ion on O-C-N Normal at Distance from N with Ion Present and Ion Removed at SCS-MP2/aug-cc-pVTZ

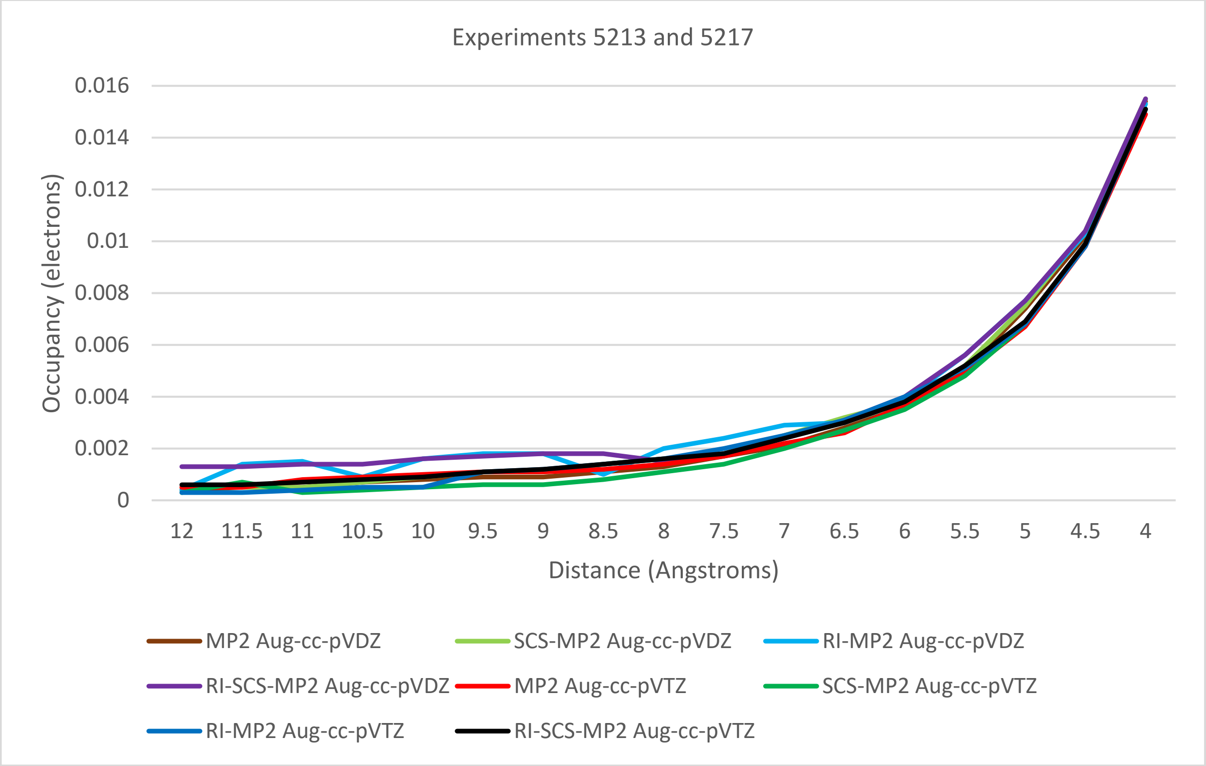


Figure 4.41. C-O(pi)* Occupancy in N-methylformamide with Ammonium Nitrogen at Distance from N Normal to O-C-N Plane Subtracted from that when Ammonium Not Present but O-C-N-H and O-C-N-CA Dihedrals Retained

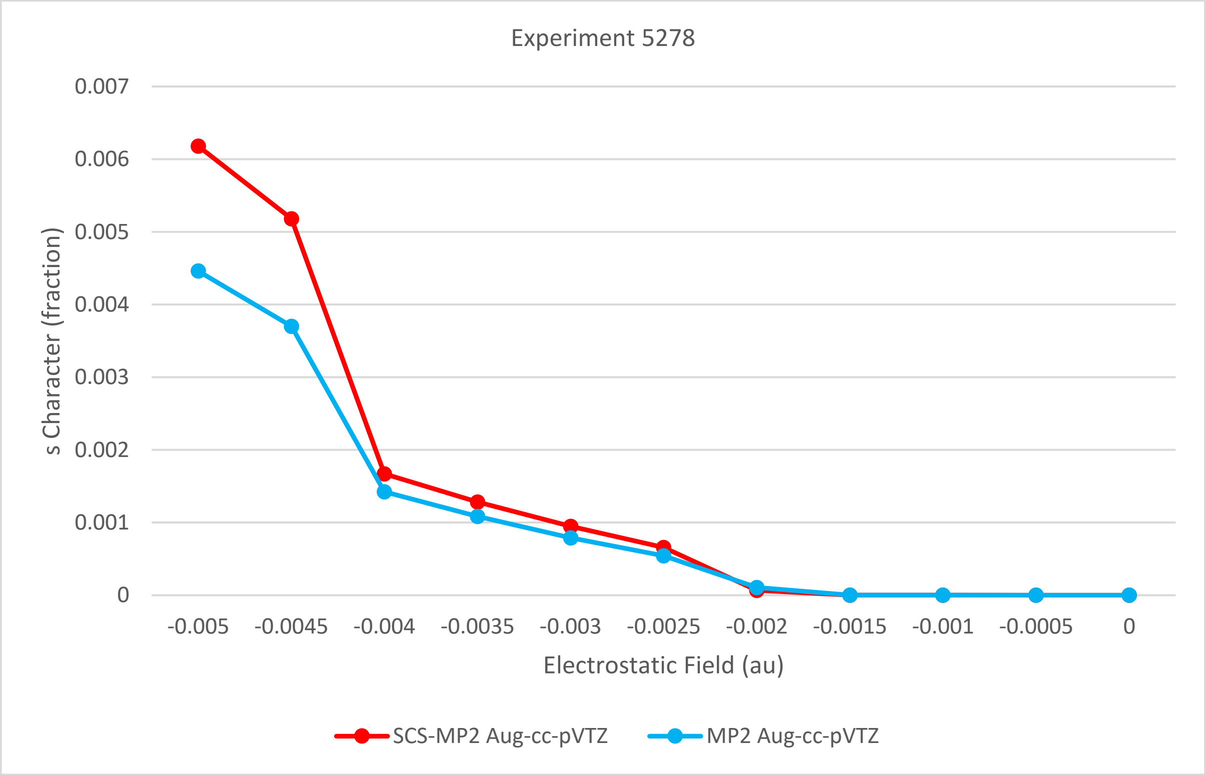


Figure 4.42. N(lp) NBO s Character in N-methylformamide with Unconstrained O-C-N-CA and O-C-N-H in O-C-N Normal Uniform Electrostatic Field

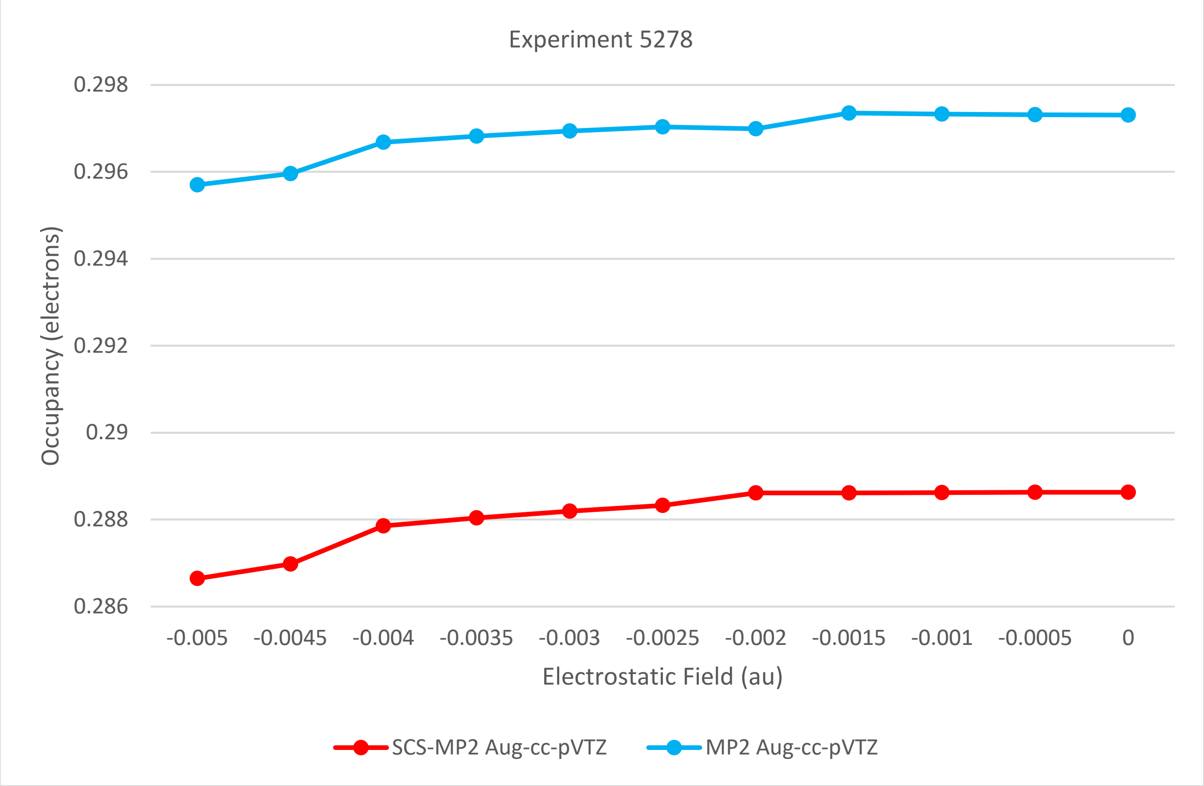


Figure 4.43. C-O(pi)* NBO Occupancy in N-methylformamide with Unconstrained O-C-N-CA and O-C-N-H in O-C-N Normal Uniform Electrostatic Field

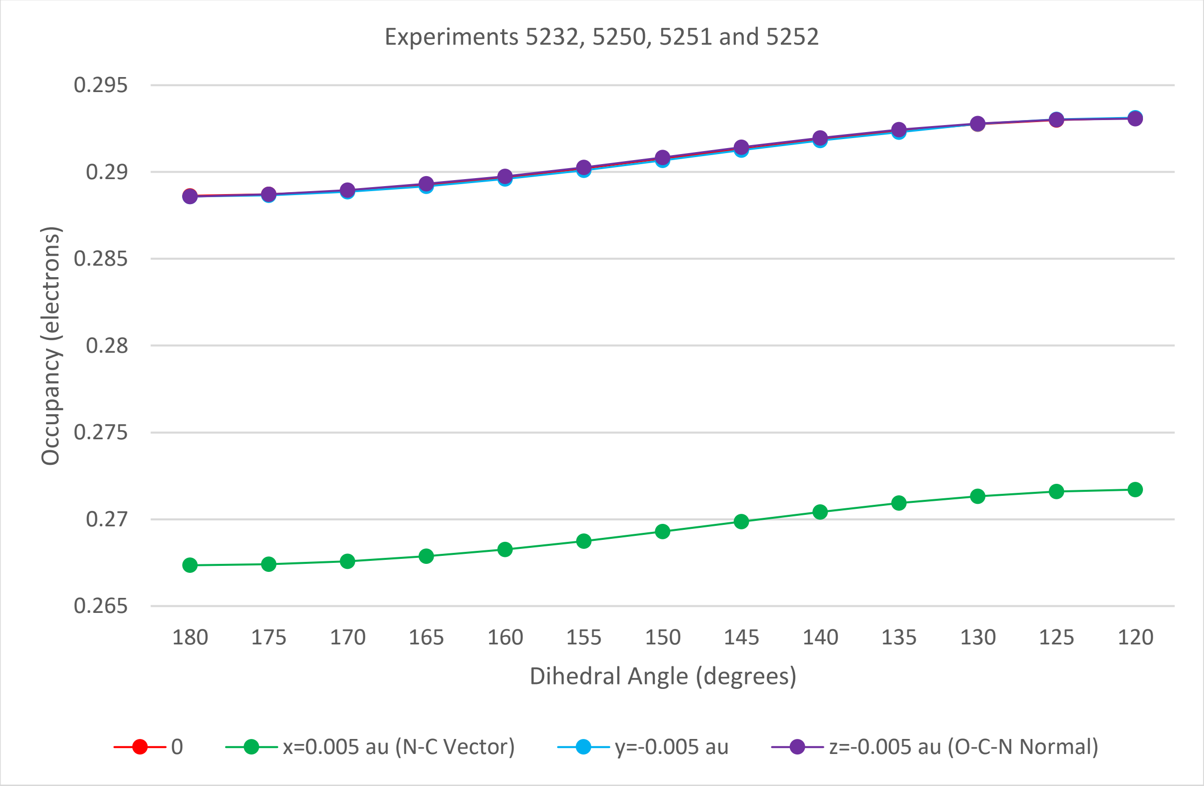


Figure 4.44. C-O(pi)* NBO Occupancy in N-methylformamide at C-N-CA-HA Dihedral in Uniform Electrostatic Field at SCS-MP2/aug-cc-pVTZ

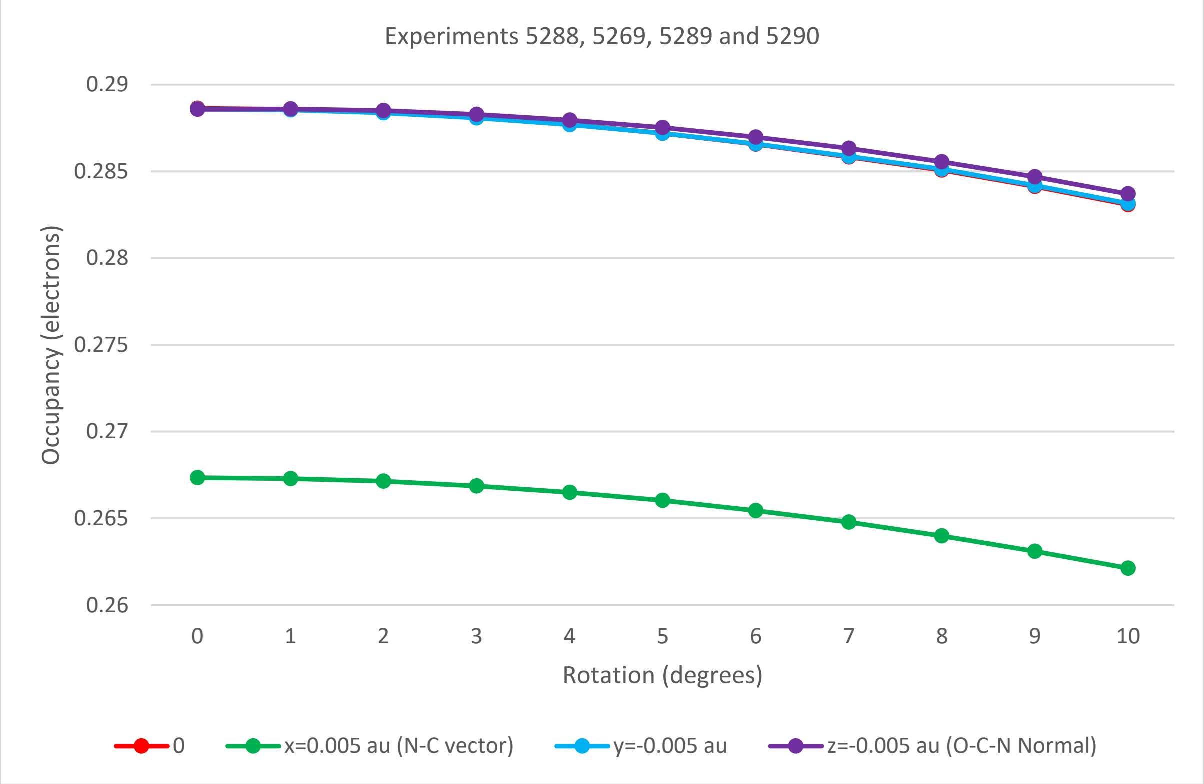


Figure 4.45. C-O(pi)* NBO Occupancy in N-methylformamide with Rotation About C-N Bond in Uniform Electrostatic Field at SCS-MP2/aug-cc-pVTZ

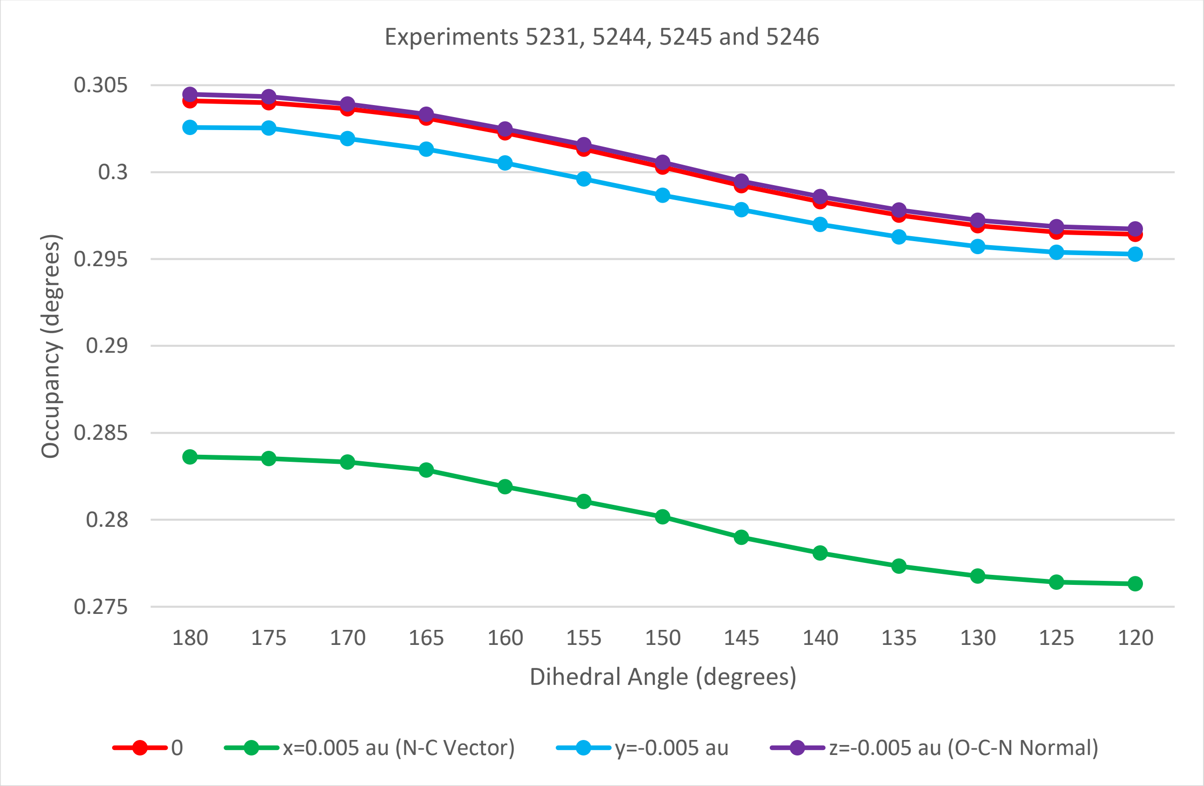


Figure 4.46. C-O(pi)* NBO Occupancy in N-methylethanamide with O-C-CA-H Dihedral Rotated in Uniform Electrostatic Field at SCS-MP2/aug-cc-pVTZ

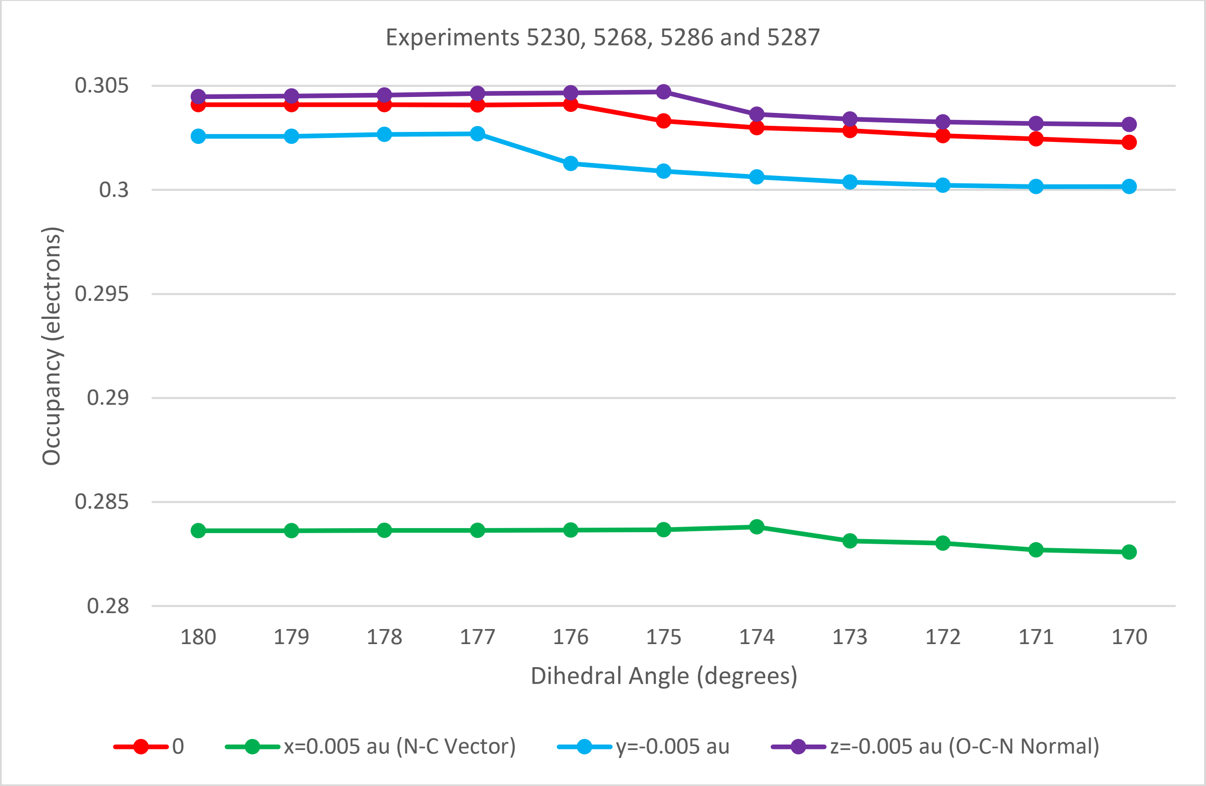


Figure 4.47. C-O(pi)* NBO Occupancy in N-methylethanamide Pyramidalized by Rotating CA-N-C-CA Dihedral Angle while Maintaining CA-N-C-O at Zero in Uniform Electrostatic Field at SCS-MP2/aug-cc-pVTZ

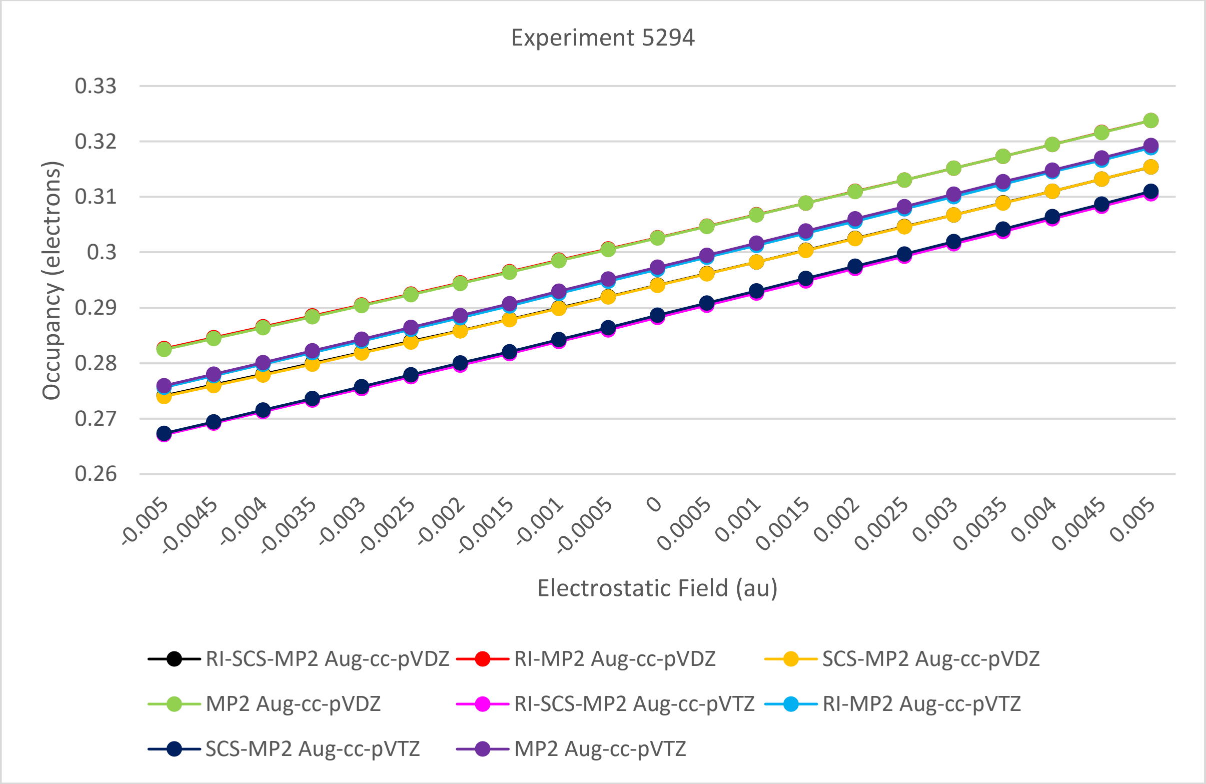


Figure 4.48. C-O(pi)* NBO Occupancy in N-methylformamide in Uniform Electrostatic Field with C-N Vector

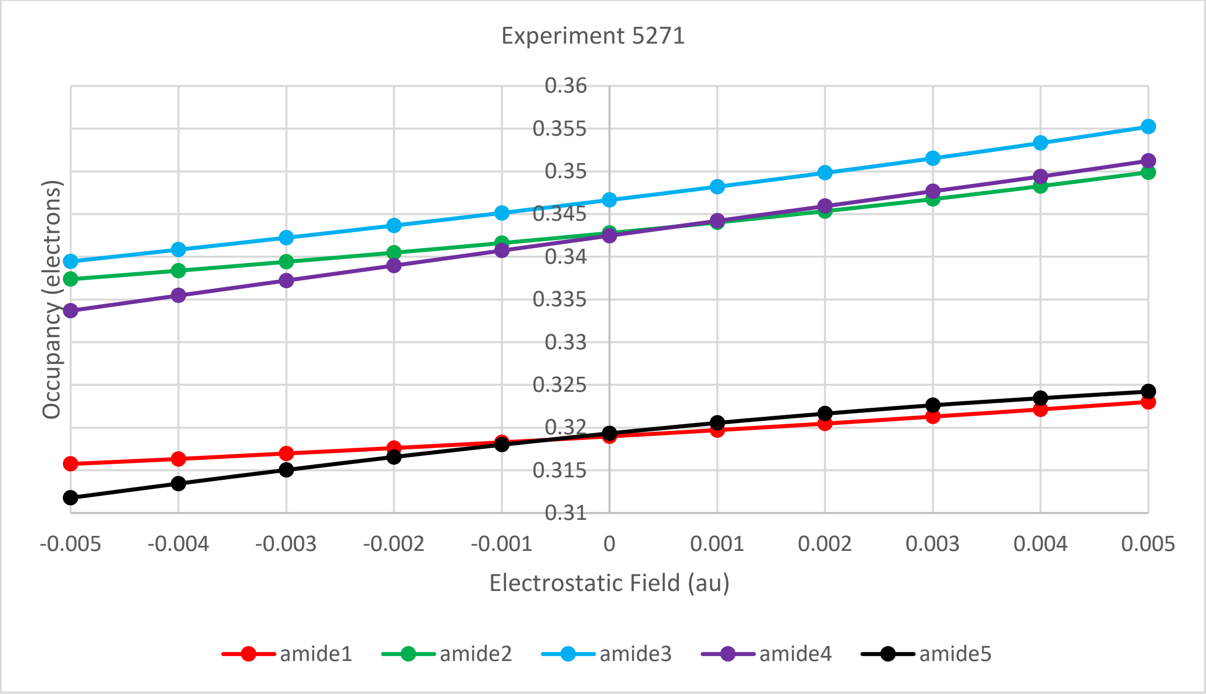


Figure 4.49. C-O(pi)* NBO Occupancy in Hydrogen Bonded N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to N-C Vector in the O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

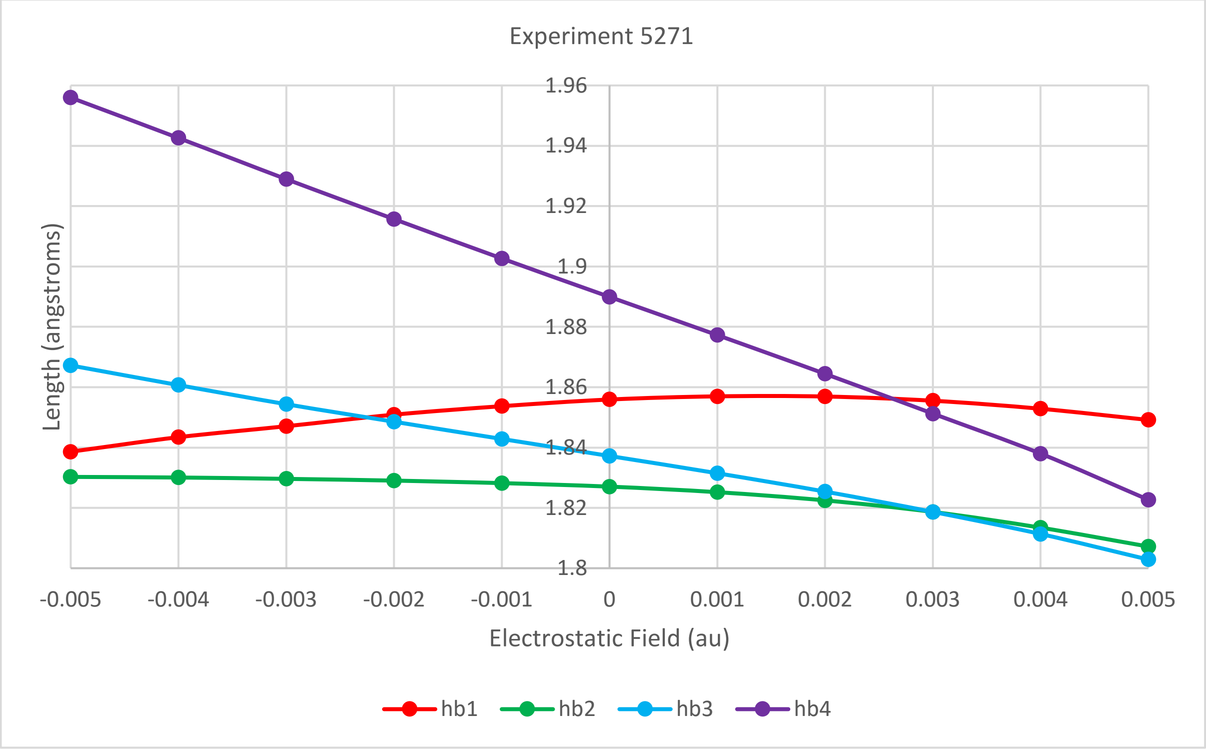


Figure 4.50. Hydrogen Bond Length in a Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to N-C Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

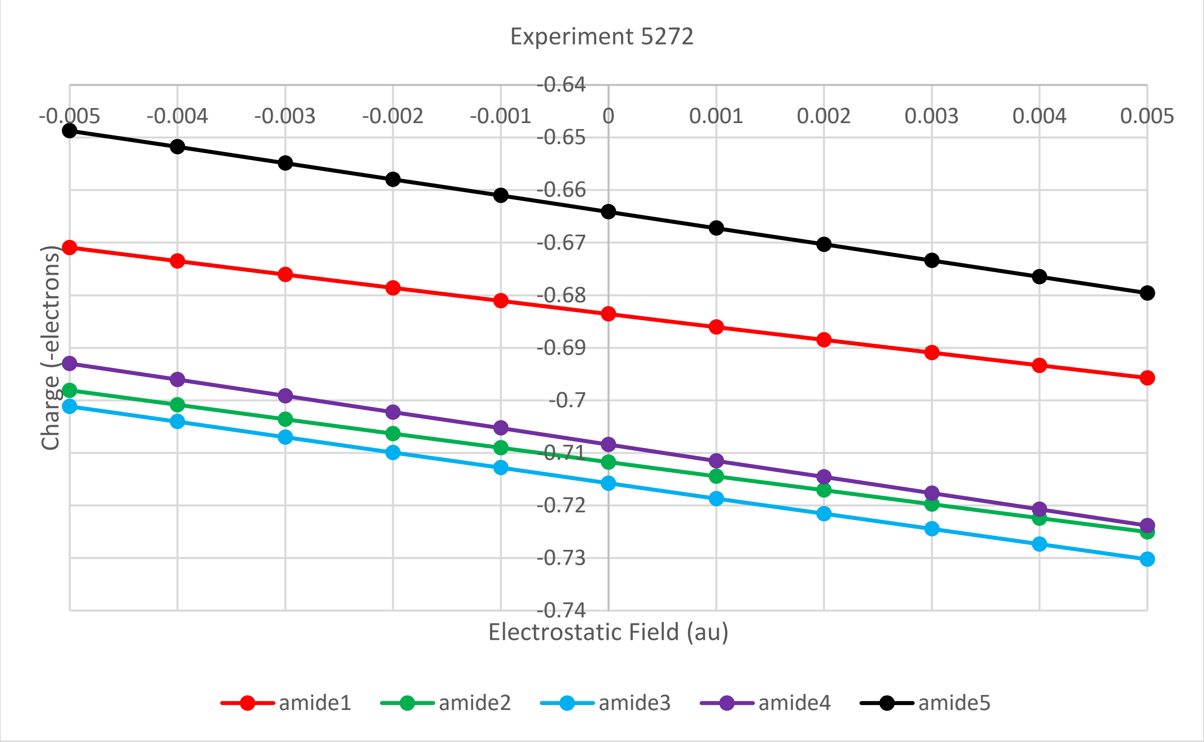


Figure 4.51. Oxygen Natural Atomic Charge in Hydrogen Bonded Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to H-N Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

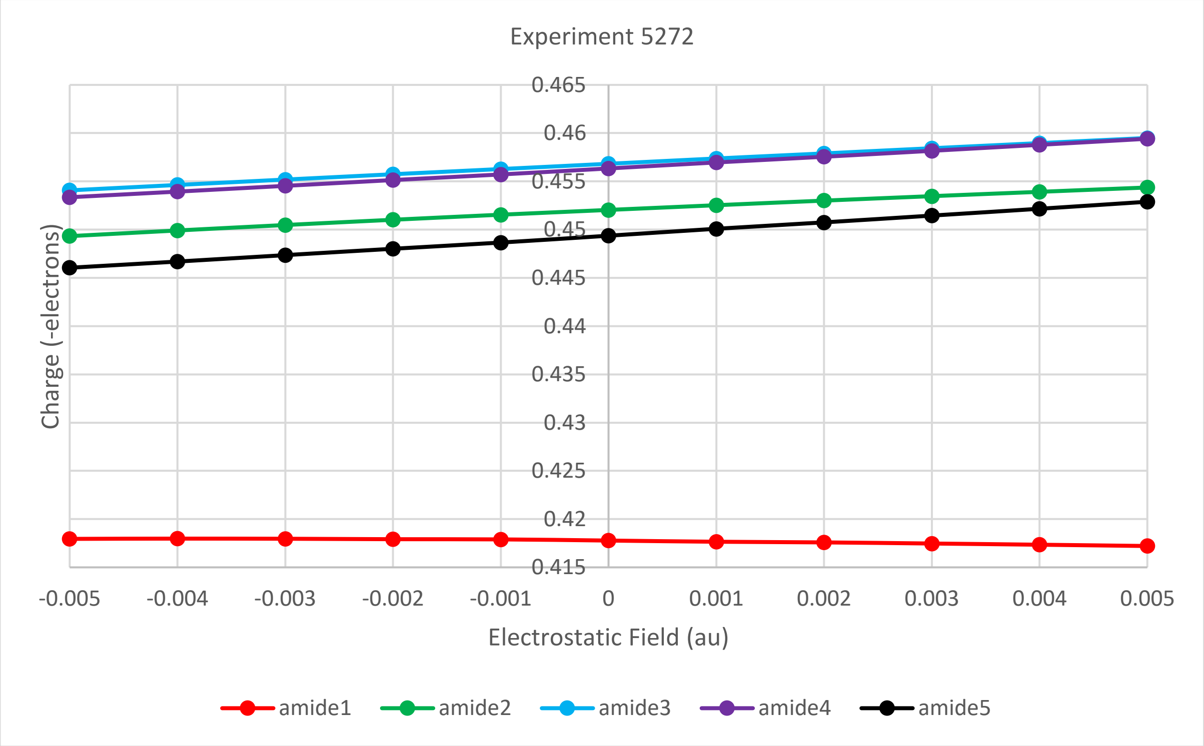


Figure 4.52. Amide Proton Natural Atomic Charge in Hydrogen Bonded Chain of N-methylformamides with Common N-C Vectors in Uniform Electrostatic Field Orthogonal to H-N Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

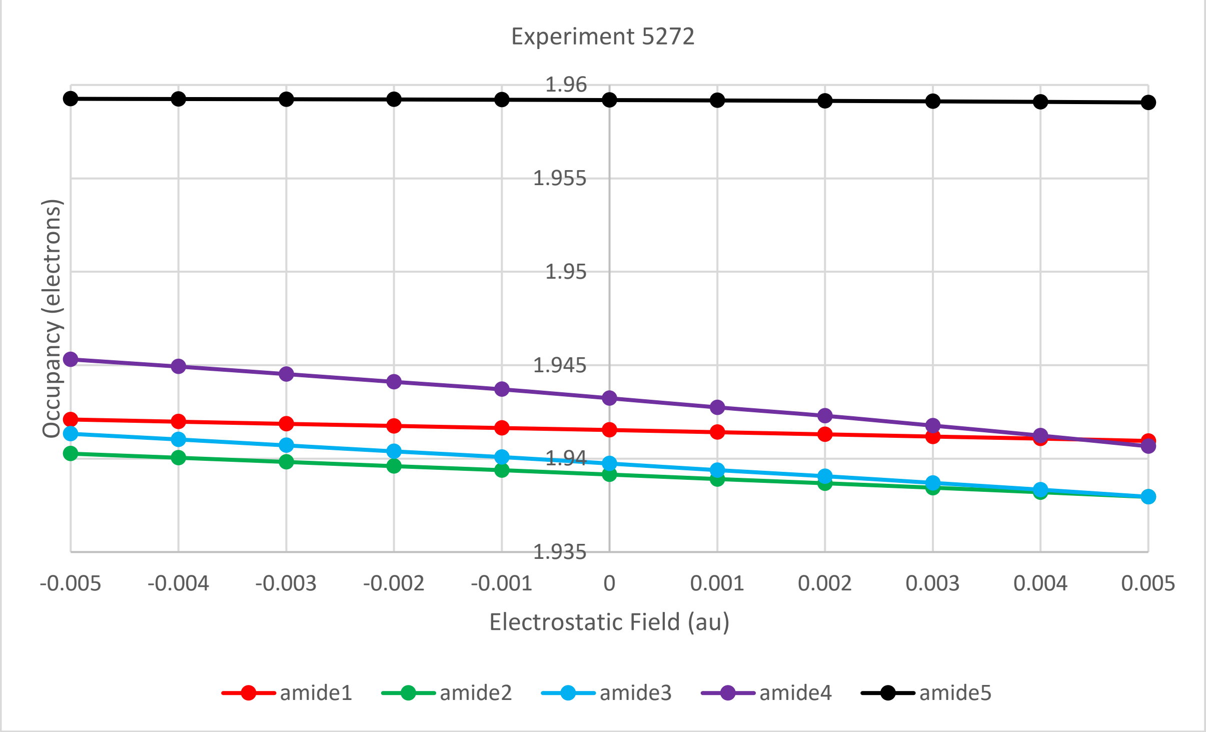


Figure 4.53. O(sigma-lp) NBO Occupancy in Hydrogen Bonded Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to H-N Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

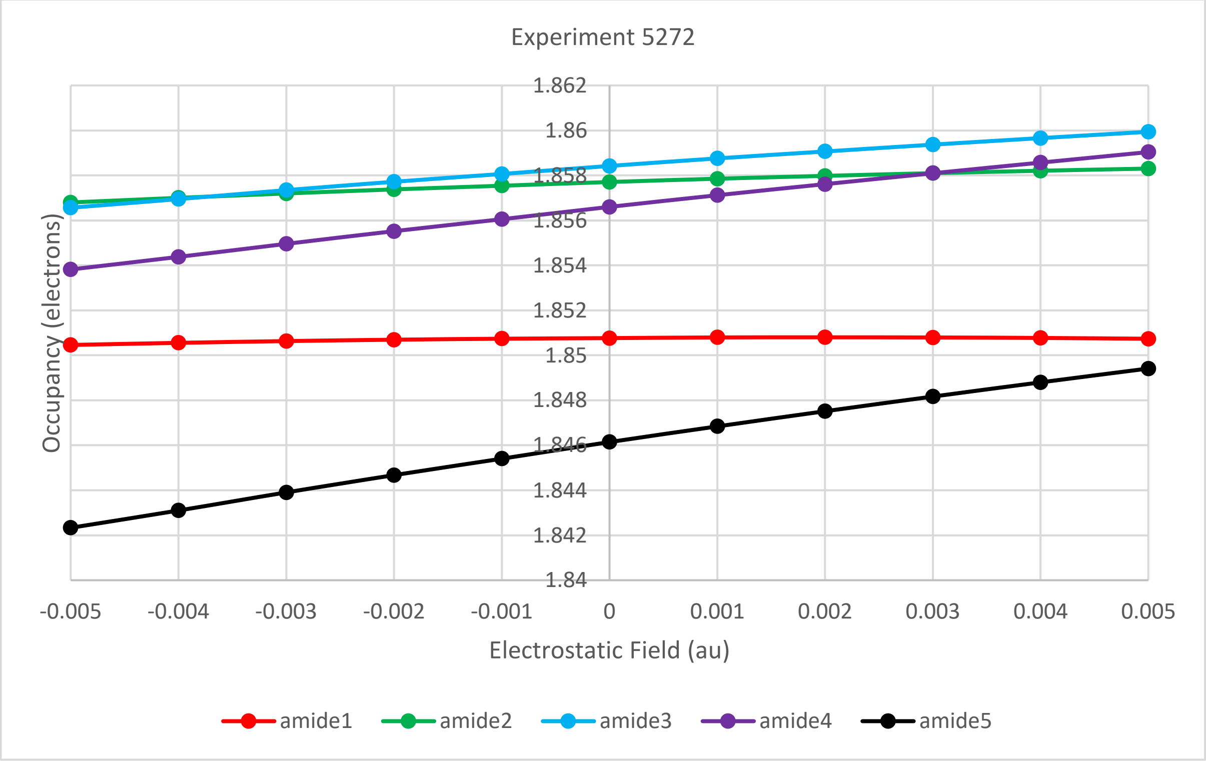


Figure 4.54. O(pi-lp) NBO Occupancy in Hydrogen Bonded Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to H-N Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

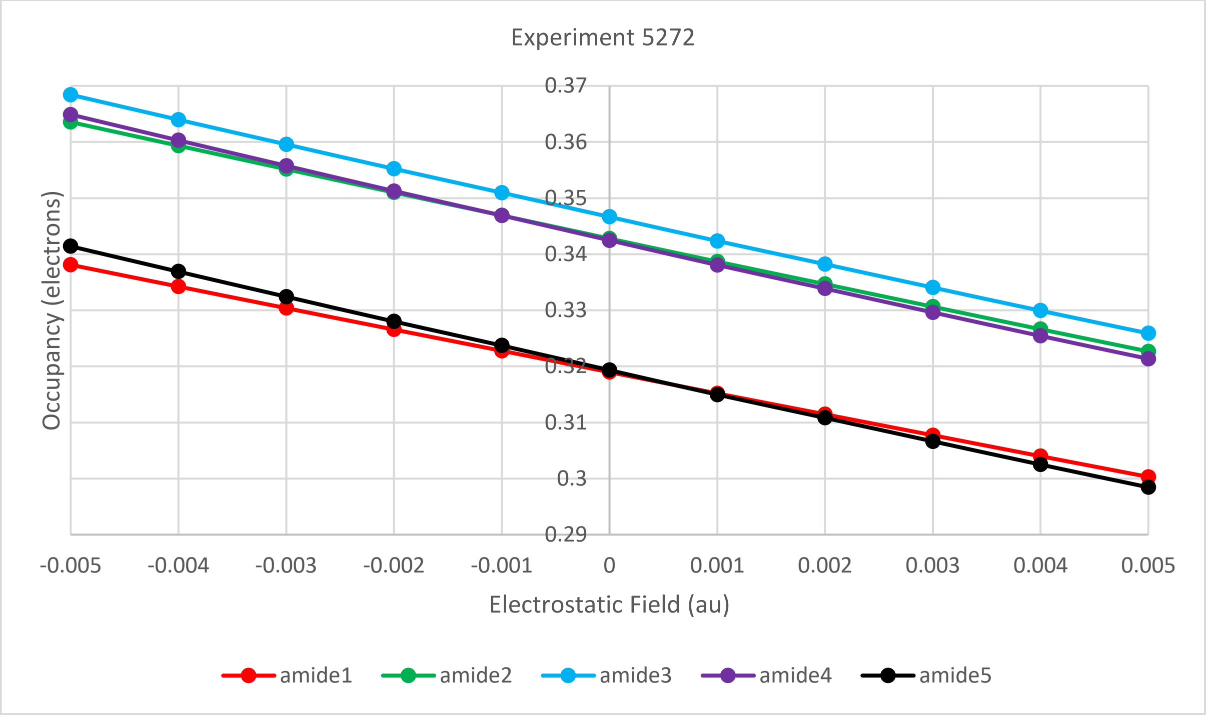


Figure 4.55. C-O(pi)* NBO Occupancy in Hydrogen Bonded N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to H-N Vectors (28.861 degrees from N-C Vectors) in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

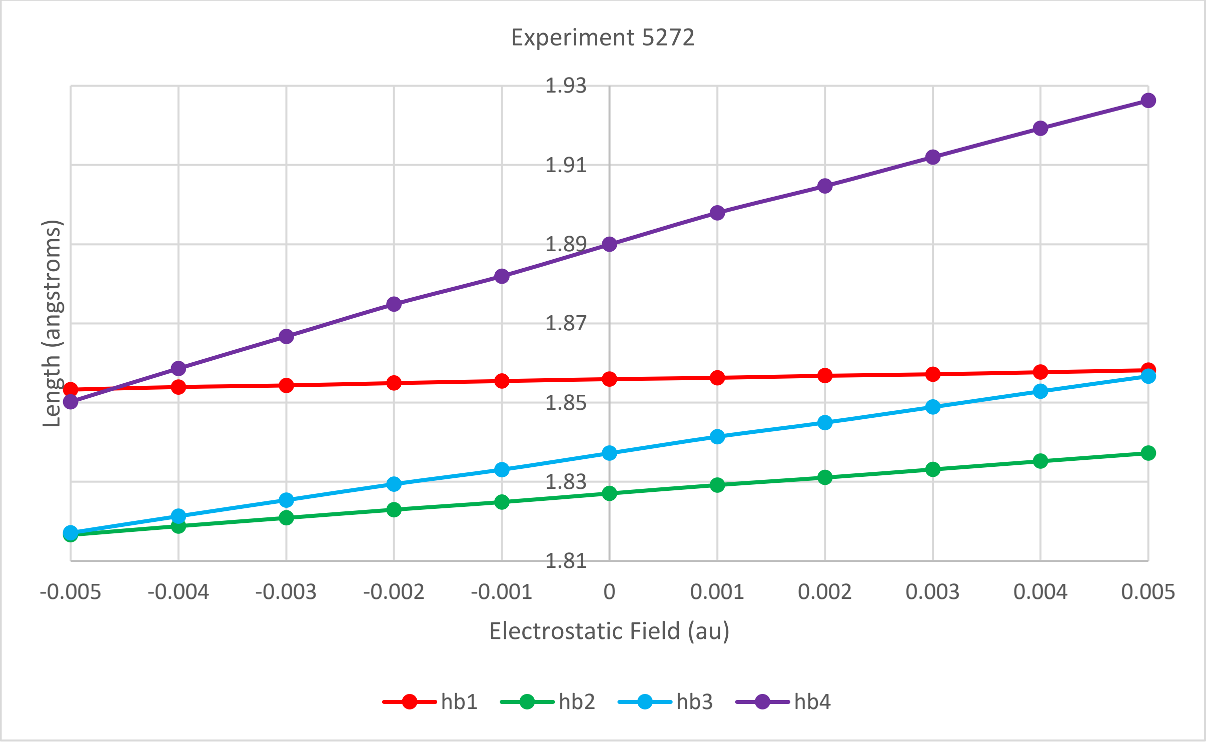


Figure 4.56. Hydrogen Bond Length in Chain of N-methylformamides with Common N-C Vectors and O-C-N-H Planes in Uniform Electrostatic Field Orthogonal to H-N Vectors in O-C-N-H Plane at RI-SCS-MP2/aug-cc-pVDZ and Coulomb and Correlation Auxiliary Basis Sets

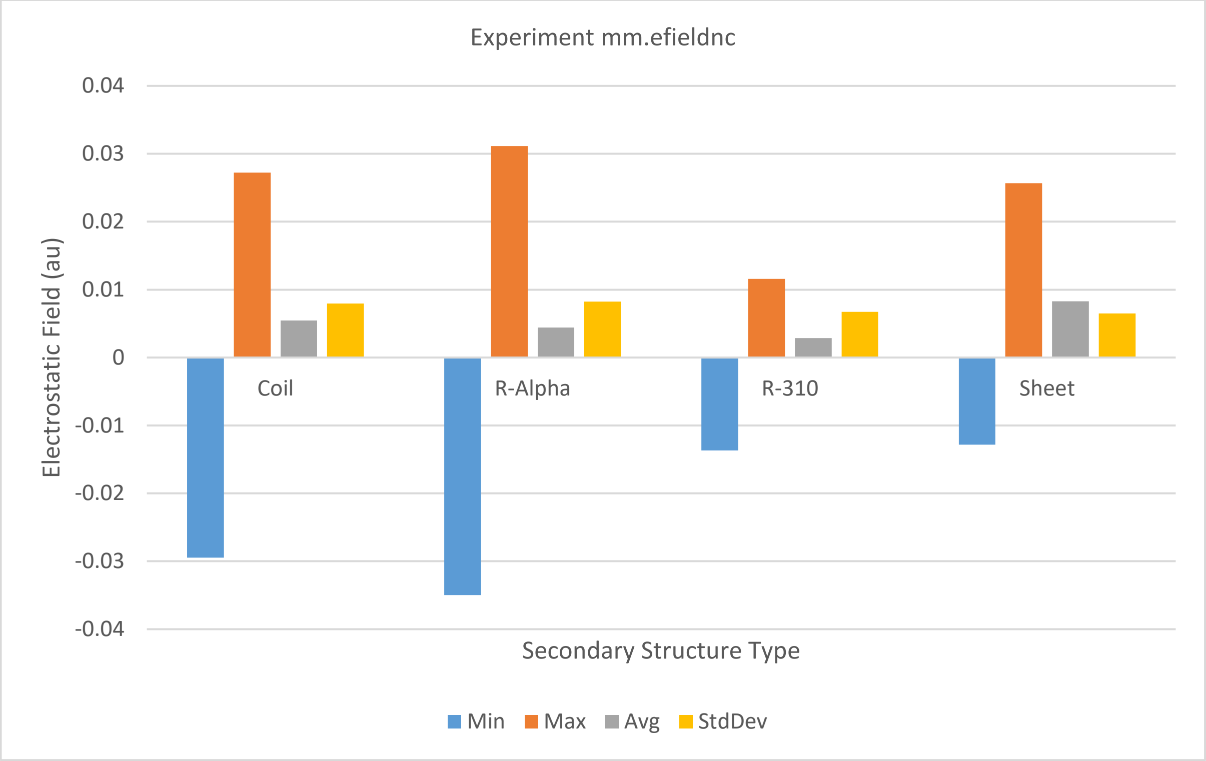


Figure 4.57. C-N Vector Component of Electrostatic at C in Backbone Amides of 10 Small Proteins as Calculated with AMOEBAPRO-2013 and Generalized Kirkwood Continuous Solvent

4.14 Appendix 2

Table 4.2. Hydrogen Bond Length (Angstroms), Dihedral Angles (Degrees) and C-O(pi)* NBO Occupancy (Electrons), with Central Atom of Molecule Constrained to O-C-N Normal from N in N-methylformamide at SCS-MP2/aug-cc-pVTZ

Molecule	Length	O-C-N-H	O-C-N-CA	Occupancy
methane	2.678	179.941	0.652	0.28791
ammonia	2.4	-170.687	-4.424	0.28121
hydrogen sulfide	2.357	-170.353	-5.043	0.27617
water	2.213	-164.869	-8.225	0.26296
ammonium	1.812	-153.949	-19.062	0.1961

5 Do cooperative cycles of hydrogen bonding exist in proteins?

5.1 Statement of authorship

Title of Paper	Do cooperative cycles of hydrogen bonding exist in proteins?	
Publication Status	☐ Published ☐ Submitted for Publication	☐ Accepted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	arXiv:1601.01792 First version appeared on arXiv (arxiv.org) on 2016-01-08.	

Name of Principal Author (Candidate)	John Neville Sharley		
Contribution to the Paper	all		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	2016-09-26

5.2 Author contact

john.sharley@pobox.com

5.3 Abstract

The closure of cooperative chains of Hydrogen Bonding, HB, to form cycles can enhance cooperativity [6]. Cycles of charge transfer can balance charge into and out of every site, eliminating the charge build-up that limits the cooperativity of open unidirectional cooperative chains. If cycles of cooperative HB exist in proteins, these could be expected to be significant in protein structure and function in ways described below. We find no mention of an example of this kind of cycle in the literature. We investigate whether cooperative HB cycles not traversing solvent, ligand or modified residues occur in proteins by means including search of Nuclear Magnetic Resonance [166], NMR, spectroscopy entries of the Protein Data Bank [164], PDB.

For the direct interactions of inter-amide HB, when the energy associated with Natural Bond Orbital [6], NBO, steric exchange is deducted from that of NBO donor-acceptor interactions, the result is close to zero, so that HB is not primarily due to the sum of direct inter-amide NBO interactions. The NBO binding energy is primarily associated with the increase in primary amide resonance of the amides, a consequence of which is that the majority of the NBO binding energy is susceptible to variation by electrostatic field with component parallel or antiparallel to an amide C-N bond (Chapter 4).

The question of what geometry most favours HB in amides is revisited with emphasis on the inequivalence of amide/carbonyl oxygen lone pairs.

A possible avenue for the design of HB-chaining polymers with improved stability is discussed.

5.4 Introduction

5.4.1 Review of Resonance-Assisted HB

The hydrogen atom is unique in that it has no core electronic shell. A donor orbital can overlap most of the H of an H-X antibonding orbital rather than being limited to the region outside of a nodal boundary [6]. A lone pair of electrons is particularly suitable as donor in this donor-acceptor interaction. An unusually strong donor-acceptor interaction, the HB, arises in this manner. That the HB is primarily resonance-type covalency or charge transfer rather than electrostatic in nature is most recently supported by evidence of anti-electrostatic HB [194].

An HB, nominated HB1, induces repolarization of the H-X acceptor antibonding orbital and its corresponding bonding orbital, resulting in elevated partial negative charge on atom X. Lone pair orbitals on the X atom become of higher energy and more diffuse. If one of the X atom lone pair orbitals donates charge for another HB, HB2, then HB2 will be of greater binding energy than in the absence of the repolarization of H-X induced by HB1. HB1 is also of greater binding energy than in the absence of HB2, since HB2 transfers charge from X allowing further repolarization of H-X and better overlap with the HB1 donor. The charge transfer, CT, of both of HB1 and HB2 is greater when both exist, that is, they are cooperative.

Resonance-Assisted HB [5, 78], RAHB, occurs when resonance gives the lone pair donor anionic character or the acceptor orbital cationic character. RAHB differs from Charge Assisted HB [6], CAHB, in that the assisting charge of CAHB is not varied by resonance. The resonance of RAHB is greater again where the resonant group can both accept and donate hydrogen bonds and both hydrogen bonds are present. In the case of the amide group, highly cooperative HB results. Amide resonance features in RAHB protein secondary structures such as alpha helices [8] and beta sheets [7].

A donor-acceptor interaction is a partial CT from a donor orbital to an acceptor orbital. The donor assumes a more cationic character and the acceptor a more anionic character as a result of the CT, which tends to oppose further CT. In an arrangement of donor-acceptor interactions such that the CT from a site is balanced by other CT to the site, the magnitude of the CT is not limited by charge imbalance [195]. In an open chain of CT between similar units at similar successive orientation, the charge imbalance is lower closer the middle of the chain for it is there that CT more closely balances charge at each site, and so CT peaks there. When the chain of CT is closed, the charge is balanced at all sites and the magnitude of each CT is the same. Unless varied by the necessary geometry change between open and closed conformations of chain, each CT in the closed conformation is at least that of the peak of the open conformation. Unless the number of units in a chain is such that an asymptotic limit of cooperativity has been closely approached, each CT in the closed conformation will exceed rather than equal that of the peak of the open conformation. With reference to hydrogen fluoride clusters, Weinhold and Landis [1] remark that “the strong preference for cyclic clusters is quite perplexing from a classical dipole-dipole viewpoint”. In investigating cyclic cooperativity in proteins, we take the NBO view that HB has partially covalent nature. This nature is most pronounced when charge transfer is largest, which tends to be when the HB is short.

5.4.2 Phenomena anticipated if cyclic cooperative HB exists in proteins

If cyclic cooperative HBs were to exist in proteins, the following phenomena might be anticipated, further motivating study of the possibility that such cycles exist.

If backbone amide HB chains in secondary structure were connected by additional cooperative units such that cooperative cyclic HB existed, additional stabilization of the secondary structures involved would result, since the sum of free energy of HB in the secondary structure part of the cycle would be higher. Various means of cooperatively cyclizing secondary structure HB chains are conceivable. For example, one spine from each of multiple alpha helices might be connected into a cycle by sidechain amides. An alpha helix might participate in three such arrangements, one for each of its spines. A beta sheet might have pairs of HB chains adjacent in the sheet connected at each end to form cycles, one cycle per pair of HB chains. A beta sheet might have its HB chains connected to form a longer chain, with a spine of an alpha helix diagonally across the sheet to form a cycle.

Binding specificity might be increased by a cooperative cycle which passes through both binding partners. This difference in binding energy between the complete cycle and incomplete cycle would give increased binding specificity. If the cycle traversed the binding interface more than twice, further specificity would result.

A cooperative cycle might be viewed as a spatially distributed store of binding energy. If energy is supplied to break the cycle at one point, the binding energy of all HBs in the cycle is decreased. Closing the cycle increases the binding energy of all HBs in the cycle. This is not dependent on the number of units in the cycle provided that number is under a unit limit of cooperativity. We showed that estimations of this asymptotic limit in beta sheet using established DFT methods and a range of basis sets must be set aside (Chapter 3). The making and breaking of cycles may result in allostery [175]. Two cooperative cycles might be mutually exclusive, with each associated with a conformation.

Some molecular chaperones [13] or other binding partners might form cooperative cycles which include cooperative units in their client, changing conformation in the client, and upon input of energy such as from hydrolyzing ATP [196], break this cooperative cycle in which they are involved, allowing formation of a cooperative cycle internal to the client.

Cooperative cycles might enhance the stability of amyloid fibrils [10] which have long cross-strand chains of inter-peptide HBs. The stacking of beta sheets might give more opportunities for closure of cycles, with sidechains completing cycles with backbone amide chains in the adjacent sheets.

5.5 Notation

“->” denotes NBO resonance-type charge transfer and “|” denotes NBO steric exchange repulsion. “(“ and “)” enclose specification of an orbital type and follow an atom name for single-center NBOs and a pair of atom names separated by “-“ for two-center NBOs.

Examples: N(lp) for the amide nitrogen lone pair NBO, O(lp-p) for the oxygen p-type lone pair NBO, O(lp-s) for the s-rich lone pair NBO, C-O(p)* for the pi carbonyl antibonding orbital NBO and N(lp)->C-O(p)* for the primary amide resonance type charge transfer.

5.6 Methods

5.6.1 Counterpoise correction

Mentel and Baerends [197] found that the use of the Counterpoise Correction [89] for Basis Set Superposition Error [198, 199], BSSE, was not justifiable. In accordance with this finding, we do not use this correction in these experiments.

5.6.2 Dispersion correction

We found that the D3 correction [93, 94] decreased amide carbonyl sigma/pi separation with the three method tested. Since we are primarily concerned with resonance, we broadly avoid this correction, using it only for comparison in one experiment.

5.6.3 Software packages

Methods used in experiments are as implemented by Gaussian 09 D.01 [75], Orca 3.0.3 [72-74] and TeraChem 1.5K [91-94]. Unless otherwise stated, default grids and optimization and SCF convergence limits were used, except that the Orca option VeryTightSCF was used throughout as were cartesian coordinates for geometry optimization with TeraChem.

A pre-release version of NBO [95] was used for its XML [59] output option. The XML was queried with XQuery 3.0 [65] or XSLT 3.0 [63] as implemented by Saxon-PE 9.6.0.4 [69], and the results imported into Excel 2013 [96].

Jmol 14.2.2_2014.06.29 [97] was used for visualization of orbitals.

Molecular coordinates are depicted by UCSF Chimera 1.10.2 [111].

5.6.4 Haskell

As detection of cooperative hydrogen bonding cycles in the PDB is dependent here on program correctness without other confirmation except when a potential example is flagged for investigation by quantum chemical means, some emphasis was placed on high probability of this correctness. The programming language Haskell [58] was used in an attempt to address the problem of program errors remaining undiscovered in all scientific codes. It is common that errors in scientific codes are discovered long after calculations have been performed by such codes. This problem increases with the size of the code base, and scientific programs tend to become large. This problem will not be fixed until scientific codes are formally proven [200] to solve highest-level equations, but until then pure functional languages [66] such as Haskell represent progress which is practicable. In pure functional languages, there are no variables, merely labels immutably bound to the results of function evaluation. By default, the programmer does not control the flow of program execution, and execution follows the necessary data dependencies. Haskell users often remark that if program code passes the compiler checks, it is likely right first time [201]. In summary, Haskell was used in an endeavour to improve reliability of results beyond that likely to be achievable with the most diligent use of imperative languages [202].

Detection of cycles was first implemented here in Haskell without a list comprehension [203], but the results were sufficiently surprising that a simpler list comprehension implementation was written. The different implementations returned the same results.

5.7 Results and Discussion

5.7.1 HB angle

5.7.1.1 Experiments involving hydrogen bonding between an amide group and hydrogen fluoride

It has long been appreciated that C-O..H-N linearity is not optimal for amide-amide hydrogen bonding, though this is usually ascribed to carbonyl lone pairs being equivalent sp2 hybrids having trigonal planar geometry [204]. However, these lone pairs are far from equivalent. As for water oxygen lone pairs when not engaged in intermolecular bonding [83, 205], carbonyl oxygen lone pairs are distinctly inequivalent, but unlike water, are exceedingly reluctant to become more equivalent when engaged in bonding. The morphology of the amide oxygen lone pair NBOs is shown in Figure 5.1 and Figure 5.2. This electron density is not equivalent to two similar hybrids, and NBOs are not unitarily equivalent to canonical molecular orbitals [83, 205]. Substantial maintenance of inequivalence for the case of Hydrogen Fluoride HB with N-methylformamide oxygen at given C-O-F angles in the amide plane with F distal to N as in Figure 5.3 is shown in Ap1:Figure 5.26 and Ap1:Figure 5.28.

With hydrogen fluoride as a probe of HB with N-methylformamide oxygen, and with the C-O-F angle constrained to given angles in the amide plane distal to N, a range of observations are made. Angles are given as deviation from linear, so that collinear C-O-F is given as 0 degrees rather than 180 degrees. This angle corresponds to the angle between the vectors C-O and O-F. Figure 5.4, Figure 5.5 and Ap1:Figure 5.23 contrast observations of H-F* NBO occupancy, HB length and C-O(p)* NBO occupancy with and without the extra constraint that the C-O-H (H of HF) angle is the same as the C-O-F angle. To the degree precision shown, the maximum H-F* occupancy occurs at 75 degrees when the C-O-H constraint is used and 80 degrees when it is not. HB length minimum in the 0 to 90 degree range considered occurs at 70 degrees with the C-O-H constraint and 75 degrees without. The maximum HB length is seen at 0 degrees i.e. with collinear C-O..H-F. The maximum amide resonance as given by C-O(p)* NBO occupancy occurs at 85 degrees with the C-O-H constraint and 75 degrees without. The minimum amide resonance occurs at 5 degrees with the C-O-H constraint and 0 degrees without.

Further figures refer to the case without C-O-H constraint. Ap1:Figure 5.24 shows that the H-F* NBO occupancy is similarly calculated by SCS-MP2/aug-cc-pVTZ and DLPNO-CCSD(T)/aug-cc-pVTZ with coulomb and correlation auxiliary basis sets at SCS-MP2/aug-cc-pVTZ optimized geometry, indicating these occupancy figures do not arise of a unique property of SCS-MP2.

Ap1:Figure 5.25 shows the F-H-O angle at given C-O-F angle with 4 wavefunction methods each with 2 correlation consistent basis sets. There is consensus at C-O-F angle of 55 degrees that the F-H-O angle is zero i.e. F, H and O collinear, and less consensus that at C-O-F angle of 0 degrees that the F-H-O angle is again zero.

Ap1:Figure 5.26, Ap1:Figure 5.27 and Ap1:Figure 5.28 show the fraction of p character of O(lp-p), O(lp-s), C-O(p) and C-O(s) at varying C-O-F angles at both SCS-MP2 and DLPNO-CCSD(T). The p character of C-O(p) does not vary, C-O(s) varies by ~0.5 percent, O(lp-p) and O(lp-s) vary by ~1.6 percent, so that O(lp-p) and O(lp-s) remain largely inequivalent.

Ap1:Figure 5.29 shows the variation of NBO donor-acceptor Second-Order Perturbation Theory [6], SOPT, and NBO Steric Exchange Energy [6] between the amide oxygen lone pair NBOs and the H-F and H-F* NBOs at C-O-F angle. For a balanced view of energetics of an interaction, donor-acceptor interactions must be considered against steric interaction of the donor and the acceptor's associated bonding orbital e.g. O(lp-p)->H-F* and O(lp-p)|H-F must be considered together. Notable features are that O(lp-p)->H-F* minus O(lp-p)|H-F, referred to here as p delta, exceeds O(lp-s)->H-F* minus O(lp-s)|H-F, referred to here as s delta, meaning the energy gradient is determined by the p-type lone pair. The s delta declines only slightly with C-O-F angle. At 75 degrees the p delta is 13.08 kcal/mol and the s delta is 2.98 kcal/mol.

Figure 5.6 shows the p delta plus s delta for a range of DFT methods at SCS-MP2/aug-cc-pVTZ optimized geometry. Ap1:Figure 5.30 shows this value when the geometry optimization is also done at each method, though due to Gaussian's aversion to near-linear ModRedundant constraints, the range of angles starts at 15 degrees. Ap1:Figure 5.31 shows this value for LC-wPBE(w=0.4) at a range of basis sets with geometry optimization at the same method and basis.

Figure 5.7 shows p delta plus s delta for formaldehyde, formamide and N-methylformamide, revealing the enhancing role of amide resonance in these HB.

Ap1:Figure 5.32 shows selected steric interactions to be considered in explaining why strongest HB does not occur at 90 degrees. Two of these rise sharply between 75 and 90 degrees. F(lp-p-2)|C-H is 0.6 kcal/mol at 75 degrees and 1.72 kcal/mol at 90 degrees. H-F|C-H is 0.23 kcal/mol at 75 degrees and 1.10 kcal/mol at 90 degrees. Together, these interactions offer explanation that it is steric interaction between HF and the hydrogen of the carbonyl carbon that prevent strongest hydrogen bonding occurring closer 90 degrees.

Ap1:Figure 5.33 shows H-F* NBO occupancy as C-O-F angle is varied from 45 to -45 degrees, revealing imperfect symmetry about 0 which is presumably due to interactions between HF and the methyl group attached to the amide N and the difference in electrostatic field caused by HF.

Ap1:Figure 5.34, Ap1:Figure 5.35, Ap1:Figure 5.36 and Ap1:Figure 5.37 show the results of unconstrained geometry optimization from a range of initial C-O-F angles for different methods, basis sets and with dispersion corrections. At least at the 1 degree gradations used for initial angle, there is no initial angle close to 0 where the optimized angle is also close to zero. If a range of initial angles exists that balances the tendency of HF to tip one way versus the other exists, this range must be less than 1 degree. In being free from constraint, this set of experiments considers all factors, including electrostatic interactions and dispersion such as captured by respective methods, and is consistent with experiments that focus on donor-acceptor and steric interactions alone. Neither electrostatic interactions nor dispersion at the applied correction result in linear or near linear C-O-F geometry being preferred.

In Figure 5.8 and Ap1:Figure 5.38, the C-O-F angle is fixed and the rotation is away from the amide plane. Ap1:Figure 5.38 shows H-F* NBO occupancy at a wavefunction method, and Figure 5.8 shows p delta plus s delta, necessarily at a non-correlated method, and is more revealing of the relationship between each given C-O-F angle. Moving out of the amide plane reduces p delta plus s delta due to reduced interaction with the p-type lone pair.

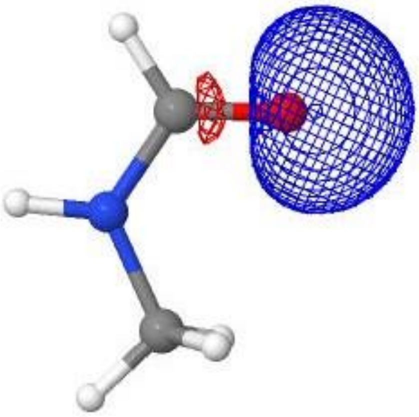


Figure 5.1. s-rich Amide Oxygen Lone Pair NBO

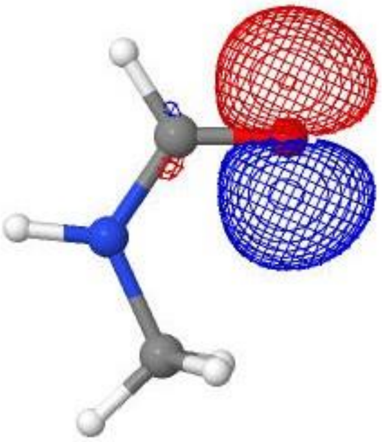


Figure 5.2. p-type Amide Oxygen Lone Pair NBO

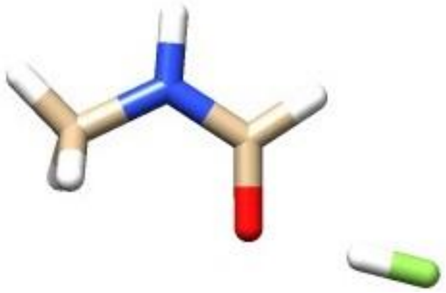


Figure 5.3. HF Hydrogen Bonded to N-methylformamide O at C-O-F Divergence from Linear of 75 degrees in Amide Plane with F Distal to N

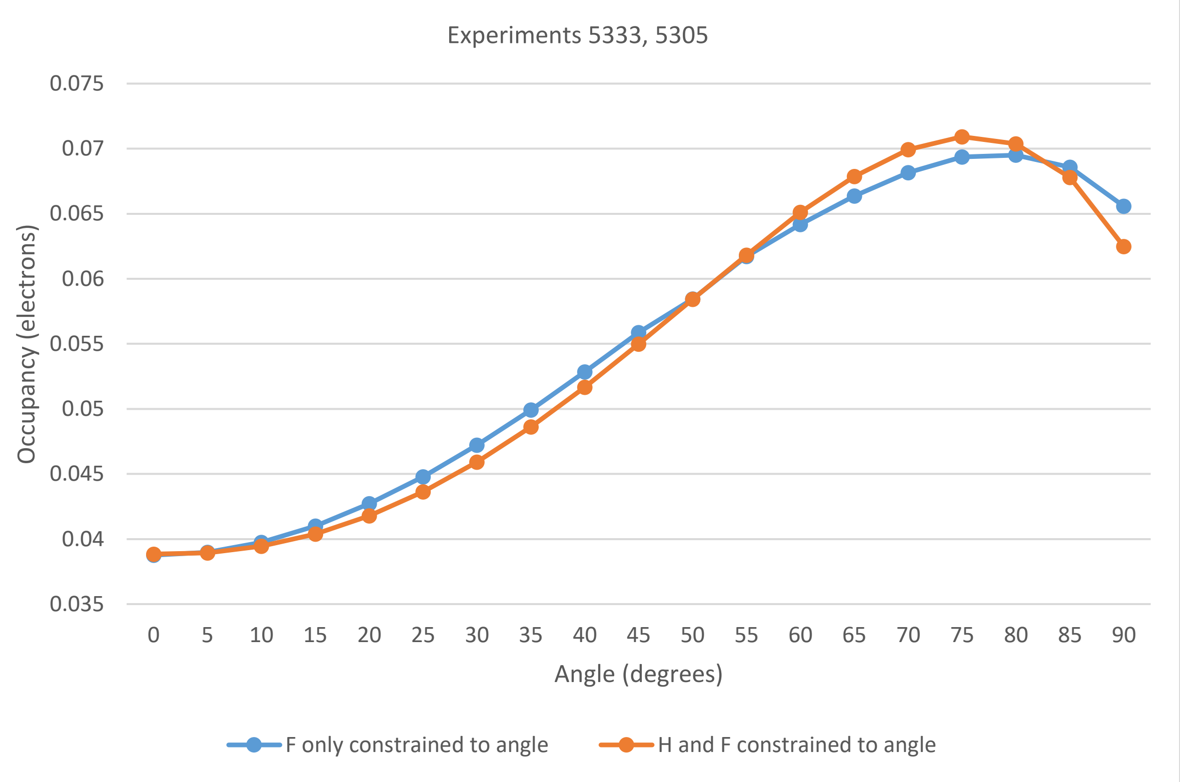


Figure 5.4. H-F* NBO Occupancy with F or H and F Constrained to Angle from C-O at O of N-methylformamide with HF Distal to N and in Amide Plane at SCS-MP2/aug-cc-pVTZ

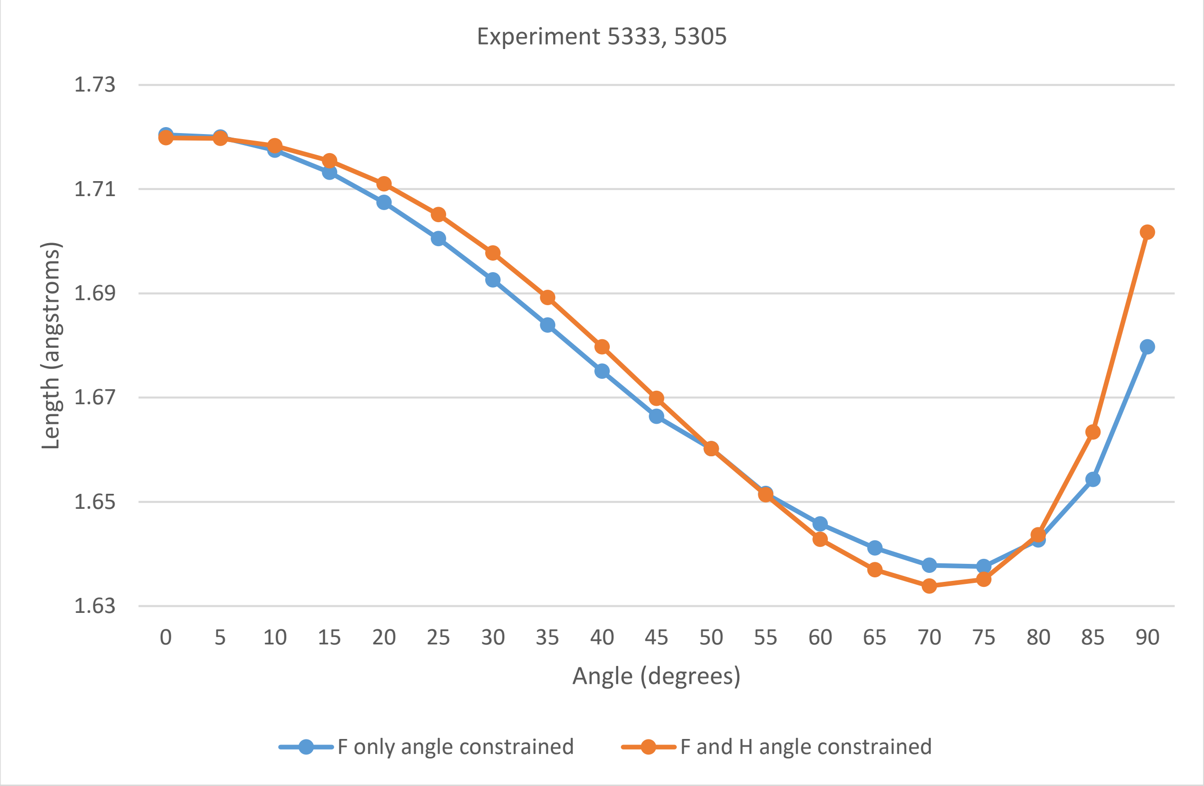


Figure 5.5. N-methylformamide/HF Hydrogen Bond Length with F Only or F and H Constrained to Angle From C-O at O with HF Distal to N and in Amide Plane at SCVS-MP2/aug-cc-pVTZ

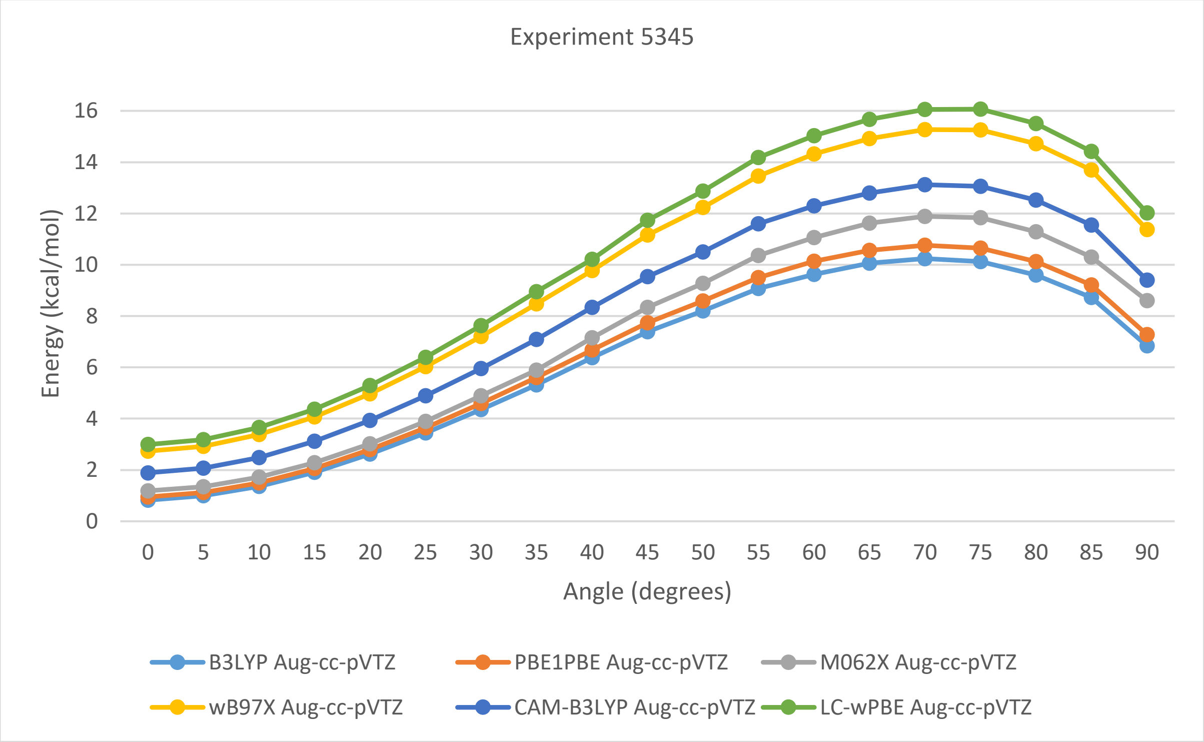


Figure 5.6. Donor-Acceptor SOPT Energy Minus Steric Exchange Energy for Interactions Between N-methylformamide O Lone Pairs and H-F and H-F* at C-O-F Angle in Amide Plane with Geometry Optimized at SCS-MP2/aug-cc-pVTZ

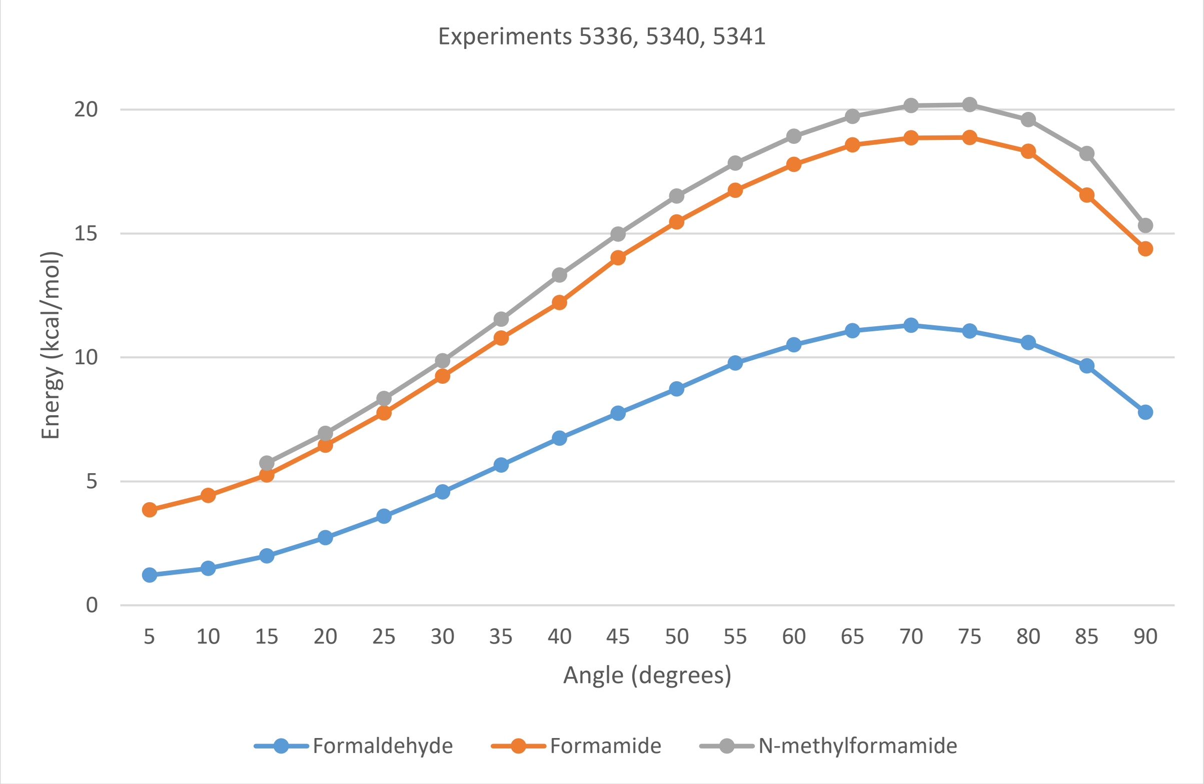


Figure 5.7. Donor-Acceptor SOPT Energy Minus Steric Exchange Energy for Interactions of O Lone Pairs and H-F and H-F* at C-O-F Angle in Aldehyde or Amide Plane at LC-wPBE(w=0.4)/aug-cc-pVTZ

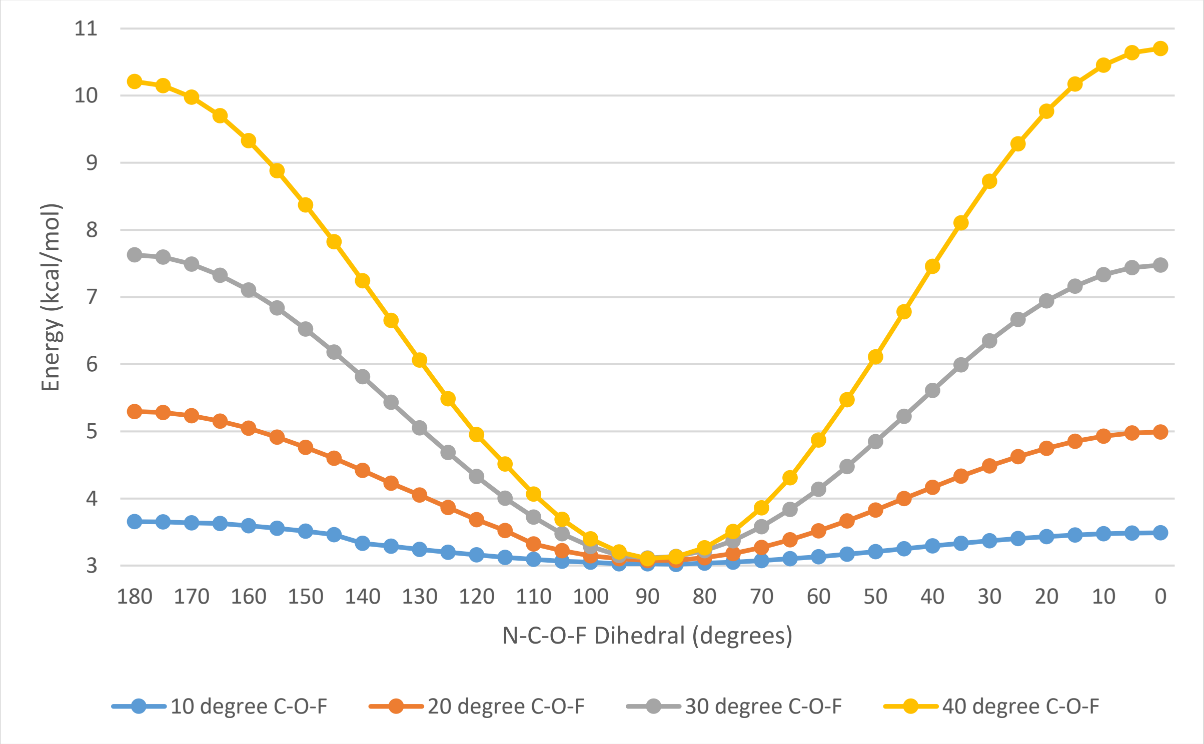


Figure 5.8. Donor-Acceptor SOPT Energies minus Steric Exchange Energies for O Lone Pair Interactions with H-F and H-F* with HF Hydrogen Bonded to N-methylformamide O at Constant C-O-F Angle with F Rotated about C-O Axis at LC-wPBE(w=0.4)/aug-cc-pVTZ over SCS-MP2/aug-cc-pVTZ Optimized Coordinates

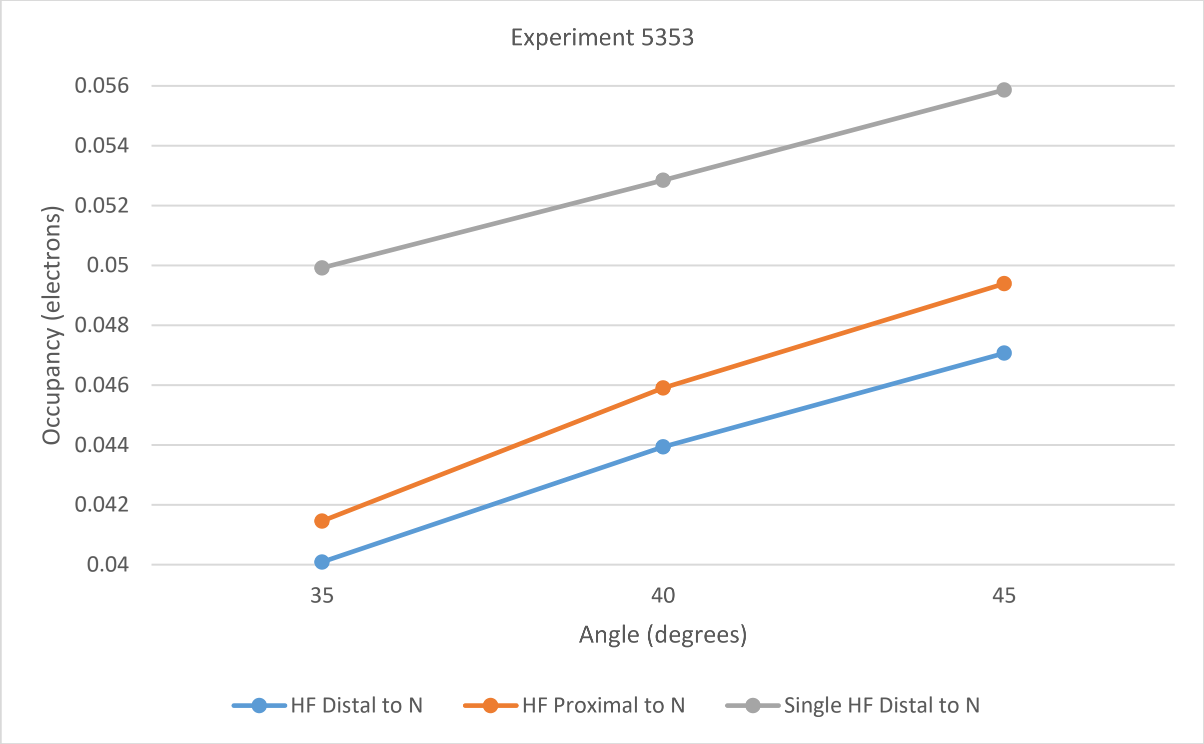


Figure 5.9. H-F* NBO Occupancies with 2 HF Hydrogen Bonded to N-methylformamide O with Both HF in Amide Plane at Same C-O-F Angle Each Side of C-O Compared with Single HF at SCS-MP2/aug-cc-pVTZ

5.7.1.2 *Use of hydrogen bonding to carbonyl oxygen where C-O-N is far from linear*

Figure 5.9 shows H-F* NBO occupancy for the case of two HF molecules in the amide plane, one distal to N, one proximal to N, both at the same C-O-F angle. The range of C-O-F is restricted to minimize the interactions between the HF molecules and the interactions of one of them with the N methyl group. The H-F* occupancy of the distal HF is about 25% less than that of the corresponding single distal HF case, and the proximal HF case has about 22% less occupancy of the single distal case. Reduction due to the busy donor effect [6] is to be expected, but it is noteworthy that the sum of the occupancies of the two HF case is considerably more than the occupancy of the single HF case. The HF interactions in the 2 HF case are anti-cooperative in keeping with the busy donor effect, but the total charge transfer and hence resonance in moieties mediating RAHB suggests an opportunity to develop new HB polymers with greater stability than polyamides [206] and polyurethanes [207] by improving on the RAHB of polymers through a bifurcation geometry more favourable than that of urea-based polymers [208]. In urea-based polymers, two H-N bonds parallel to the C-O of the next chain, giving bifurcated HB at O but at distinctly sub-optimal geometry (Ap1:Figure 5.39), whereas optimal geometry is to be found by the H-N bonds pointing in to the O in the manner of 2 HF HB to amide oxygen experiment above (Figure 5.9). This bifurcation geometry improvement requires that the nitrogens be further apart than they are in urea. One HB to urea oxygen at large C-O..N angle from linear will transfer substantially more charge to H-N* than an HB at shallower angle from linear, and two at large angle will deliver substantially more in total again though less than double. The total resonance of the resonant moieties connecting the HBs will be markedly increased in the progression through the three cases: one HB with linear C-O-N, one HB with large C-O-N divergence from linear, two HBs with large C-O-N divergence from linear. The design problem is to have near-optimal bifurcation geometry and still have a highly resonant moiety present, and a question is at what atom count such a solution might be found should such solutions exist. Evolutionary algorithms [209] might be used to search the space of possibilities.

Large C-O-N divergence from linearity by non-bifurcated hydrogen bonds is seen in nature. There are examples of this in nitrogenous base pairing [12] such as in guanine/cytosine pairing. HB of the carboxyl group to guanidinium demonstrates C-O-N at ~63.5 degrees from linear (Figure 5.40). Replacing formate with carbamate (Figure 5.41) gives yet closer hydrogen bonding, interpretable as being due to 2 charge transfers from carbamate N to its carboxyl group. When guanidinium is replaced with urea (Figure 5.42 and Figure 5.43), bond lengths increase due to urea not having the positive charge of guanidinium to neutralize the charge of the carboxyl group. The overall charge of a unit of the polymer must be zero. These examples of large C-O-N divergence for a single HB might serve as a starting point for design of optimally bifurcated HB at double bonded oxygen or sulfur.

Smart rubbers [210] make no use of covalent chains for their assembly, and entirely rely on HB for their properties. If the design of smart rubber emphasized the HB geometries discussed here, strength of the material might be improved such that the strength of ordinary rubber might be reached.

Much interest is now focussed on development of materials based on covalently bonded sheets and cylinders for nanomaterials applications [211], but desirable properties might still be found with HB polymers, particularly in view of the substantial increase in total resonance and hence stabilization potentially available with more favourable HB bifurcation geometry at double bonded oxygen or sulfur.

5.7.1.3 Experiments involving hydrogen bonding between two amide groups

A pair of N-methylformamides is used to investigate hydrogen bonding between a pair of amide groups. Figure 5.10 shows the sum of the inter amide donor-acceptor SOPT energies minus the sum of the inter-amide steric exchange energies. It is remarkable how low the total is, particularly with the aug-cc-pVTZ basis set. As the angle approaches 0, for some method/basis combinations the result become negative, but otherwise they are quite modest positive figures. That inter-amide donor-acceptor minus steric energetics for HB are minor between 2 co-planar amides and even negative close to C-O-H linearity depending on method/basis used is a surprising result. Ap1:Figure 5.44 separates the SOPT and steric exchange figures for the inter-amide interactions for 1 method and 2 basis sets. Ap1:Figure 5.45 shows select SOPT and steric energies internal to one amide and Ap1:Figure 5.46 internal to the other. Ap1:Figure 5.47 shows the N(lp)->C-O(p)* SOPT energy minus steric exchange energy for each amide with 1 method and 2 basis sets.

Ap1:Table 5.6 gives these internal amide figures for one N-methylformamide in isolation. The N(lp)|C-O(p) steric exchange differs little between this case or either of the bound amides, but the N(lp)->C-O(p)* is ~112 kcal/mol in each of the bound amides at 0 degrees, but in the isolated amide is about 102 kcal/mol, so that the hydrogen bonding at 0 degrees is about 20 kcal/mol due to amide resonance alone. It is not charge transfer minus steric interactions between amides that predominate in hydrogen bonding between amides, it is the change in the resonance of each amide. The interactions between the amides serve to increase the amide resonance in each amide, rather than having direct bonding energetic significance of their own. This is at odds with any view that the direct inter-amide interactions are energetically dominant with the increase in resonance of each amide relegated to the role of bonus. That the majority of bonding energetics reside in the increase of the amide resonance has particular significance in terms of the findings of Chapter 4. Inter-amide hydrogen bonding is largely rather than partially susceptible to variation of amide resonance by electrostatic field with component parallel to the amide C-N bond, increasing the likely significance of this variation relative to other factors in protein folding such as pure electrostatics, hydrophobia and entropy.

Notes for Figure 5.10

Amide 1 has the oxygen participating in the HB, Amide 2 the hydrogen of the same HB

The donor-acceptor (SOPT) interactions between the amides summed are:

Amide 1 -> Amide 2

O(lp-p) -> H-N*, C-O(p)*, C-O(s)*

O(lp-s) -> H-N*, C-O(p)*, C-O(s)*

C-O(p) -> H-N*

C-O(s) -> H-N*

Amide 2 -> Amide 1

H-N -> C-O(p)*

H-N -> C-O(s)*

The steric interactions between the amides summed are:

Amide 2 | Amide 1

N(lp) | C-O(p), C-O(s), O(lp-p), O(lp-s)

H-N | C-O(p), C-O(s), O(lp-p), O(lp-s)

C-O(p) | O(lp-p), O(lp-s)

C-O(s) | O(lp-p), O(lp-s)

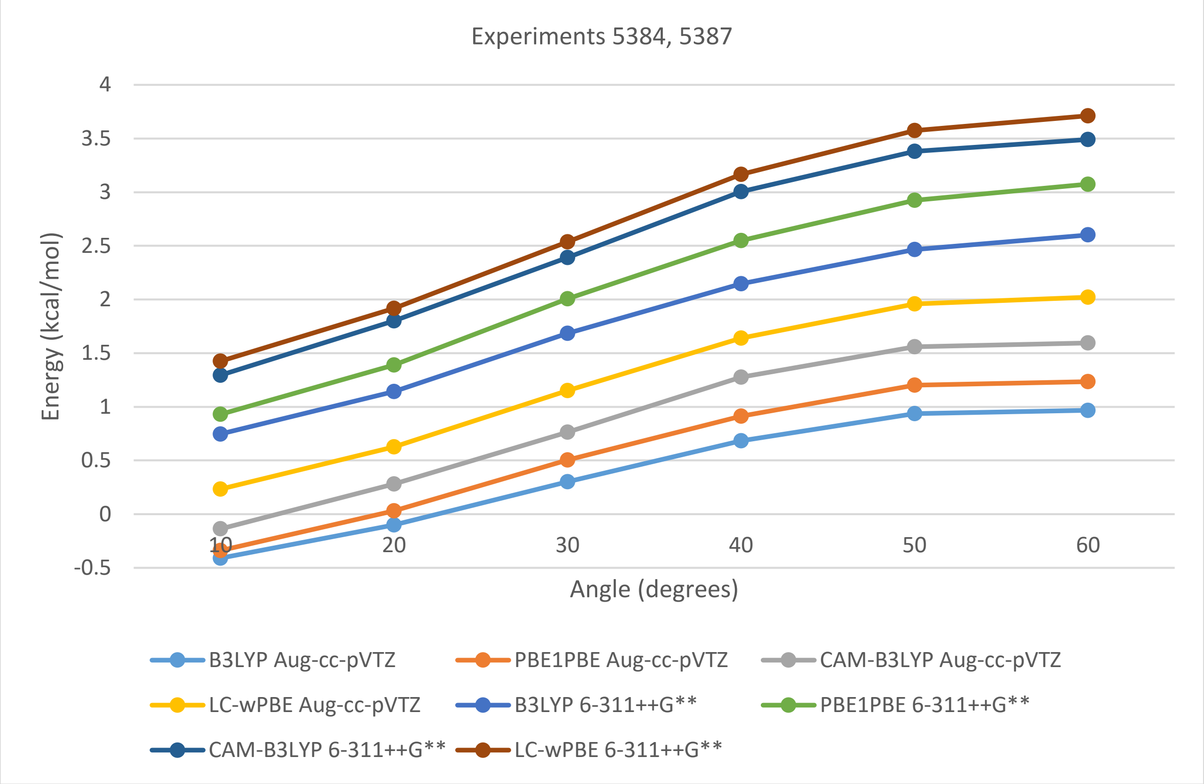


Figure 5.10. Total SOPT Energy Minus Steric Exchange Energy Between 2 Coplanar Hydrogen Bonded N-methylformamides at Given C1-O1-N2 Angles with N1-C1-O1-N2 Dihedral 180 Degrees

5.7.2 Inter-backbone amide hydrogen bonding C-O-N angles in the Protein Data Bank

The Protein Data Bank was queried by the means described in Chapter 4 for backbone amide-backbone amide hydrogen bonding with geometry restricted as defined in the notes for the resulting Ap1:Table 5.7. This data suggests that C-O-N angles with the large non-linearity investigated above are sterically disfavoured in proteins. The method for querying the PDB is further described in Section 5.7.7 (Cyclic HB in the Protein Data Bank).

5.7.3 Linear chain of formamides

A linear chain of 8 coplanar formamides was geometry optimized at LC-wPBE(w=0.4)/6-311++G** without constraint. The chain remained coplanar (Figure 5.11). Figure 5.12 shows that the SOPT energy associated with the primary amide charge transfers follows the expected pattern, peaking in the middle of the chain. Ap1:Figure 5.48 shows that interactions between the formamide units that may be responsible for maintaining approximate linearity of the chain remain minor, though O(lp-s)->H-C* is most notable. Figure 5.49 shows that O(lp-p)->H-N* exceeds O(lp-s)->H-N* by the second hydrogen bond, and the latter declines more than the former in the final hydrogen bond which is in keeping with the increased C-O-N (less linear) angle shown in Figure 5.13. The resonance of the preceding amides is related to the C-O-N angle, except where the resonance has increased O(lp-s)->H-C*.

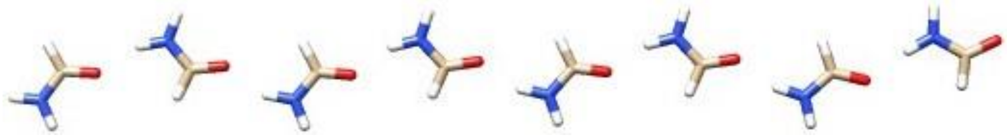


Figure 5.11. Chain of 8 Formamides after Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

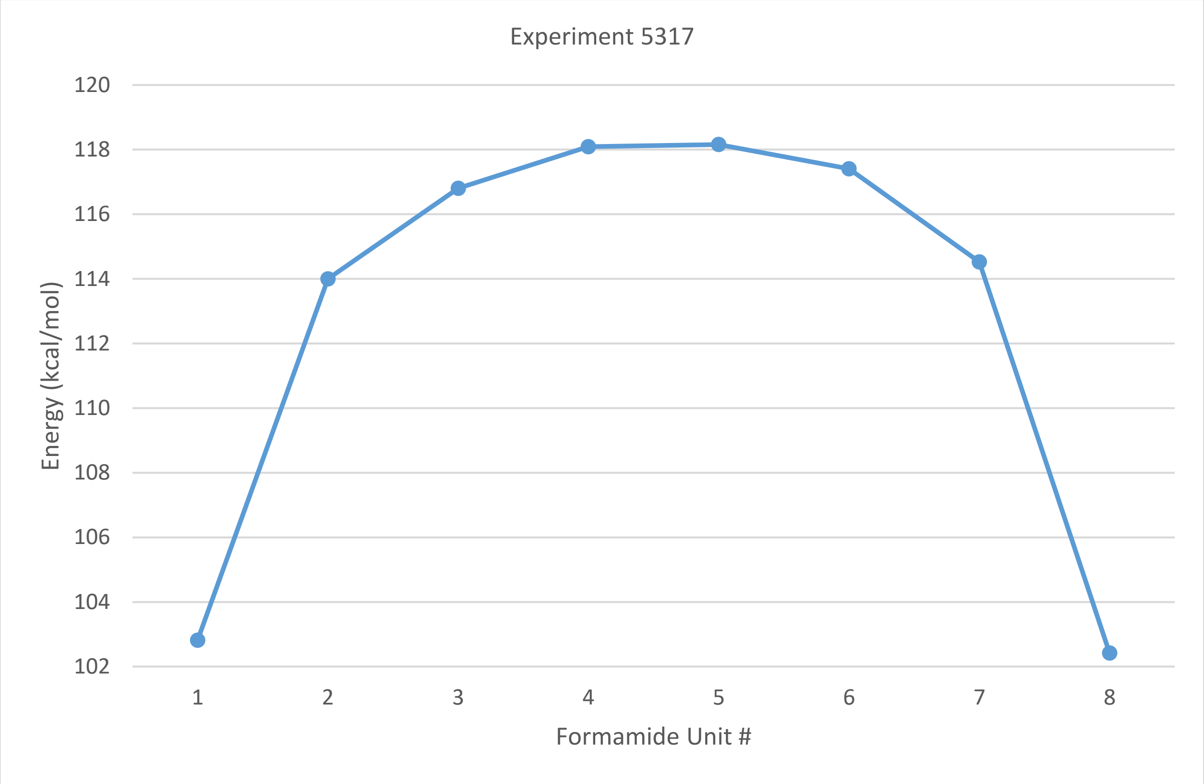


Figure 5.12. N(lp)->C-O(p)* SOPT Energy in Hydrogen Bonded Chain of 8 Formamide Units with Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

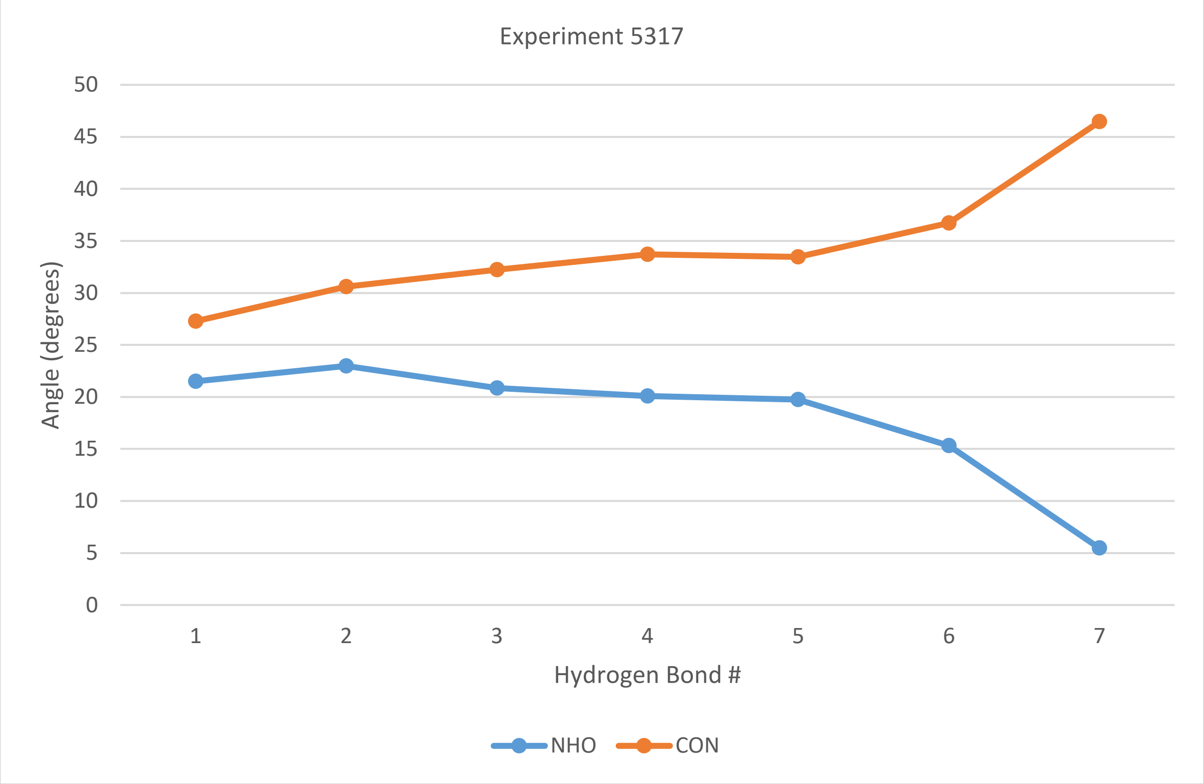


Figure 5.13. Hydrogen Bond Angles in Chain of 8 Formamide Units with Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

5.7.4 Formamide cycles

Cycles from 6 to 12 coplanar units of formamide such as in Figure 5.14 were geometry optimized without constraint at LC-wPBE(w=0.4)/6-311++G**, and these remained coplanar. It can be seen in Figure 5.15 that hydrogen bond length is a minimum at 12 units, though is not uniformly declining from 6 units due to an increase at 11. In Figure 5.16 it can be seen that the primary amide charge transfer energy peaks at 8 units and has a downward spike at 11 units. This is in keeping with the SOPT minus steric exchange energy line seen in Ap1:Figure 5.50, though the slope of the line is shallow. As seen in Figure 5.17, the maximum variation in C-O-N angle with a cycle is at 11 units and the minimum is at 8 units. Also, the minimum variation in N-H-O angle within a cycle occurs at 8 units.

There are two limitations to cyclic hydrogen bonding seen in this section. The first is that when coplanar, the average C-O-N angle necessarily declines with increase in the number of units in the cycle, and eventually amide resonance declines, and with it hydrogen bonding. Coplanarity does not favour large cooperative cycles. The second is the sensitivity to uniformity of C-O-N angle in a cycle, since the least amide resonance in the cycle is limiting of cyclic RAHB.

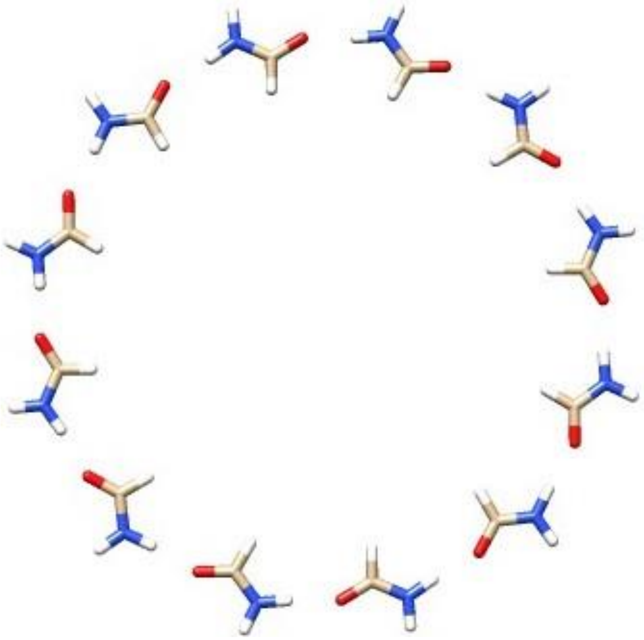


Figure 5.14. Cycle of 12 Formamides after Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

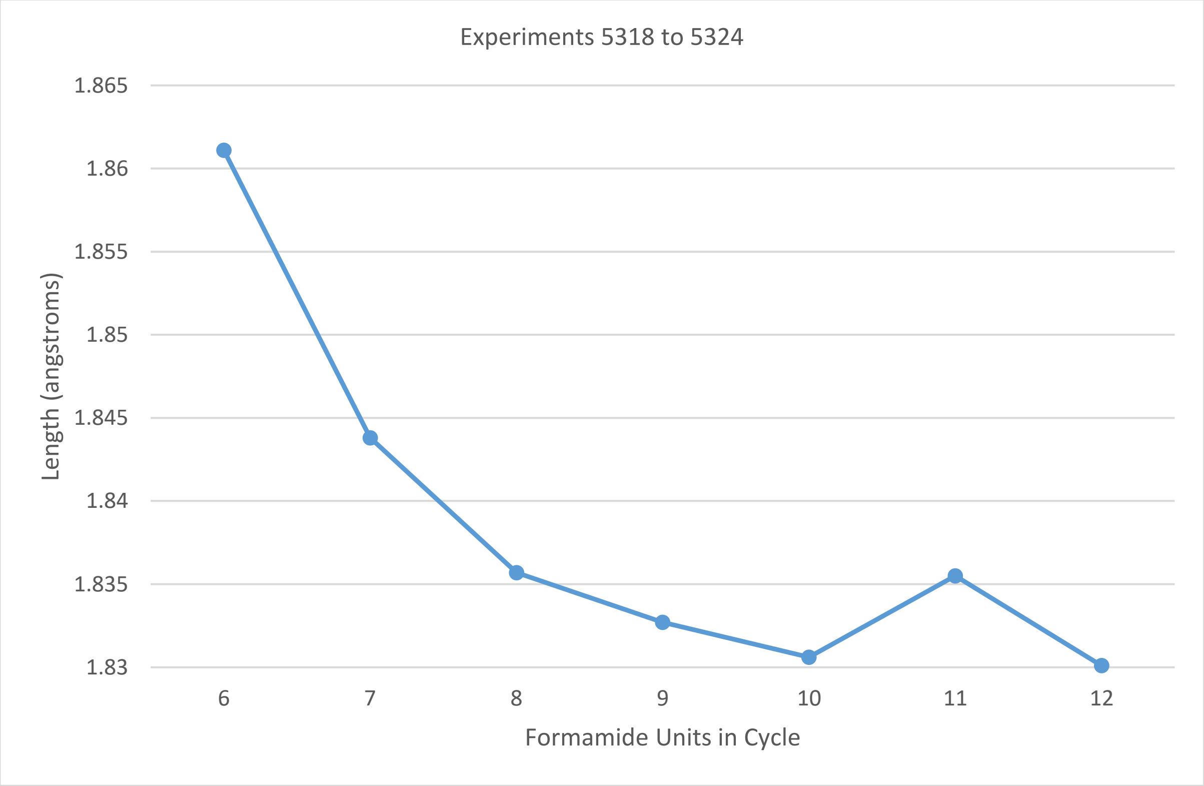


Figure 5.15. Hydrogen Bond Length in Planar Cycles of Formamide Optimized at LC-wPBE(w=0.4)/6-311++G**

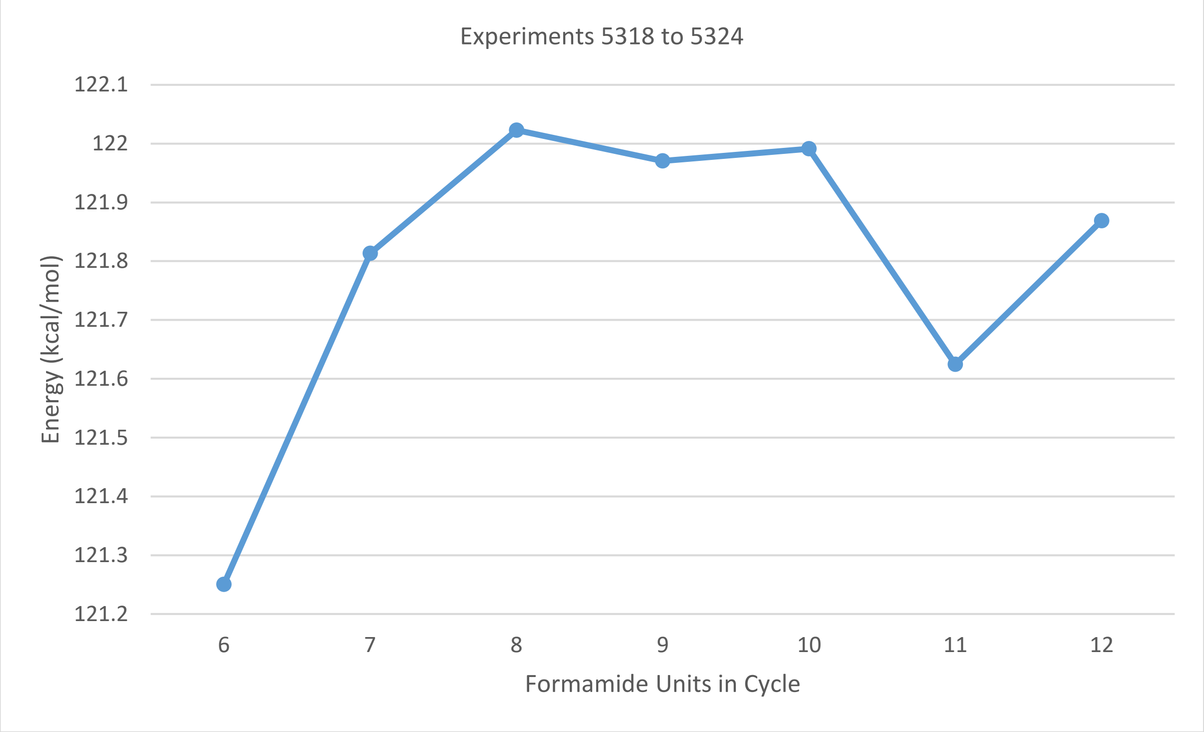


Figure 5.16. N(lp)->C-O(p)* SOPT in Hydrogen Bonded Formamide Units in Planar Cycle Optimized at LC-wPBE(w=0.4)/6-311++G**

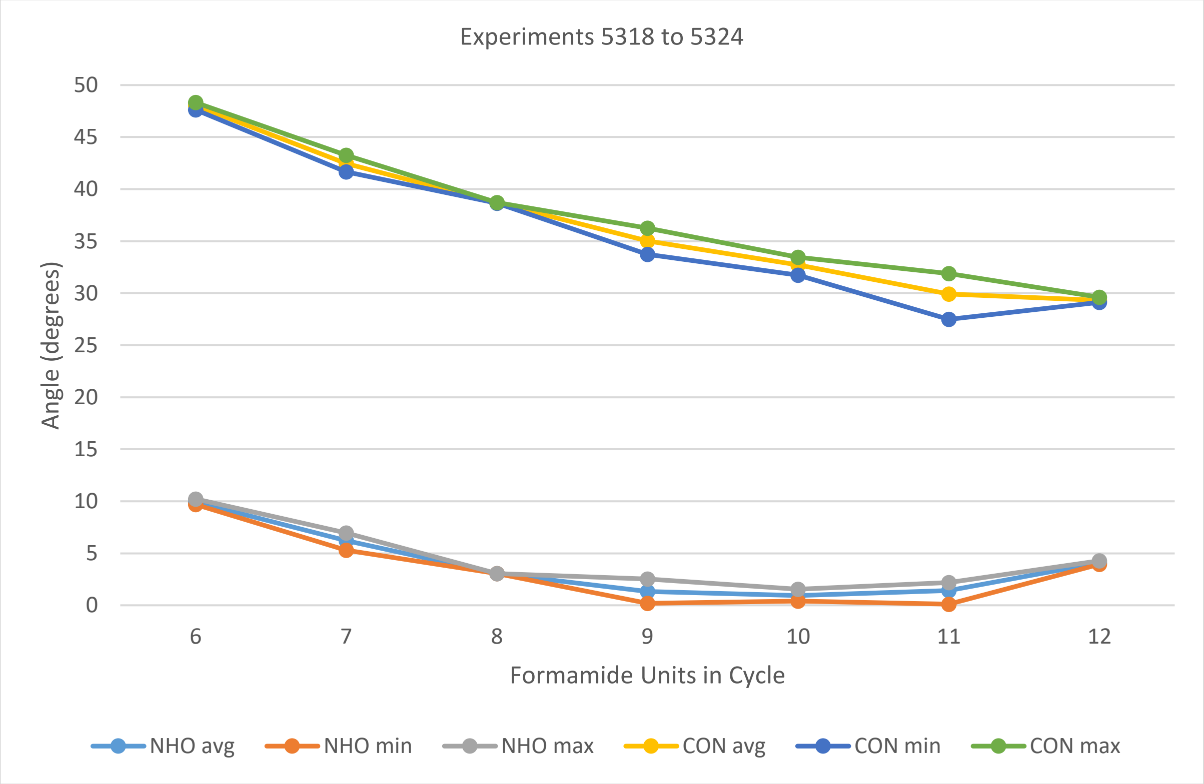


Figure 5.17. Hydrogen Bond Angles in Planar Cycles of Formamide Optimized at LC-wPBE(w=0.4)/6-311++G**

5.7.5 Three alpha helices

In exploration of what binding energy advantage exists in non-dynamic gas-phase cycles of hydrogen bonding, three Ace-Ala(15)-Nme alpha helices were connected by three formamides such that each formamide hydrogen bonds with two spines of each of two helices (Figure 5.18). Only one spine in each helix is hydrogen bonded by two formamides, and only this spine participates in a hydrogen bonding cycle (Figure 5.19). Formamide is representative of the sidegroup of asparagine or glutamine. HB and amide resonance of this three helix arrangement is compared with those of a single such helix with two capping formamides (Figure 5.20) in the manner of the three helix arrangement and the result is shown in Table 5.1.

The key to evaluating cyclic RAHB is the primary amide resonance charge transfers, which increase from ~1626 to ~1693 kcal/mol for backbone amides per helix, and decrease from ~108 to ~96 kcal/mol for each formamide. The cyclization is thus favourable by ~201 kcal/mol for the three-helix arrangement as assessed by primary backbone amide resonance charge transfer alone. In a physiological situation, the HB sites of the formamides, representing sidechain amides, in the single helix case are likely to be fully utilized, which would increase the resonance of the amides of the helix, but these isolated, gas-phase and non-dynamic experiments demonstrate that cyclization is advantageous under isolated circumstances.

The “ressb” column, N(lp)->C-O(s)*, of Table 5.1 shows purely DFT error and indicates that the findings of Chapter 3 for beta sheets extend to alpha helices although they are less marked. The “resss” column shows this DFT error for formamide, indicative of the error in asparagine and glutamine.

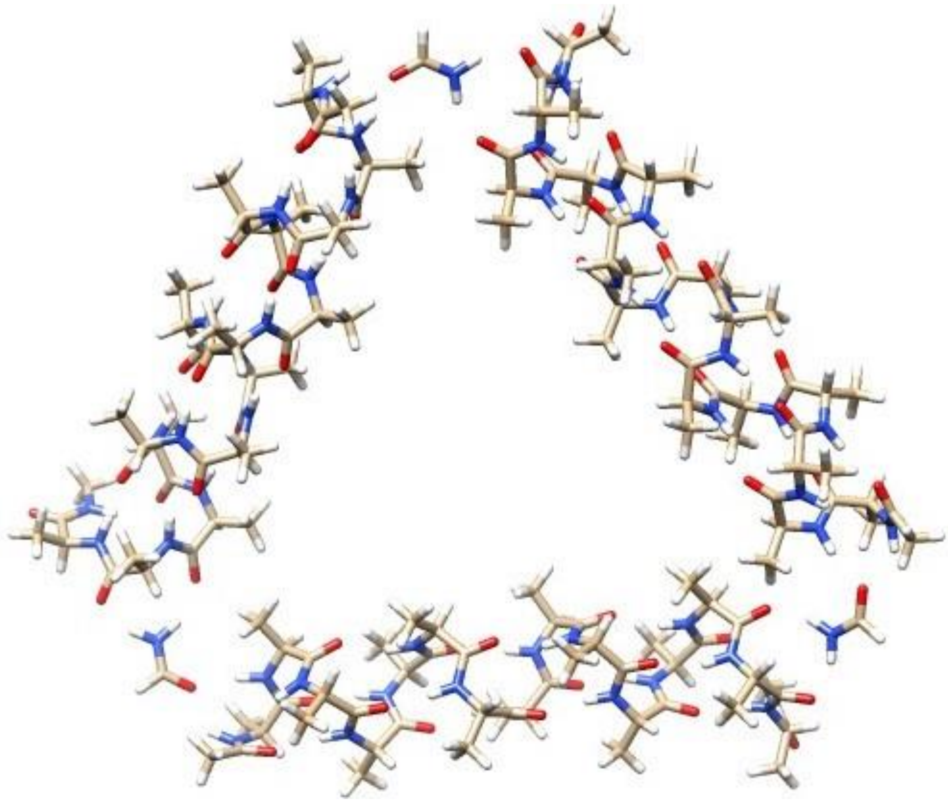


Figure 5.18. 3 ACE-Ala(15)-NME Alpha Helices with 3 Formamides Each Hydrogen Bond-Connecting Two Spines of Adjacent Helices at LC-wPBE(w=0.4)/6-311++G**

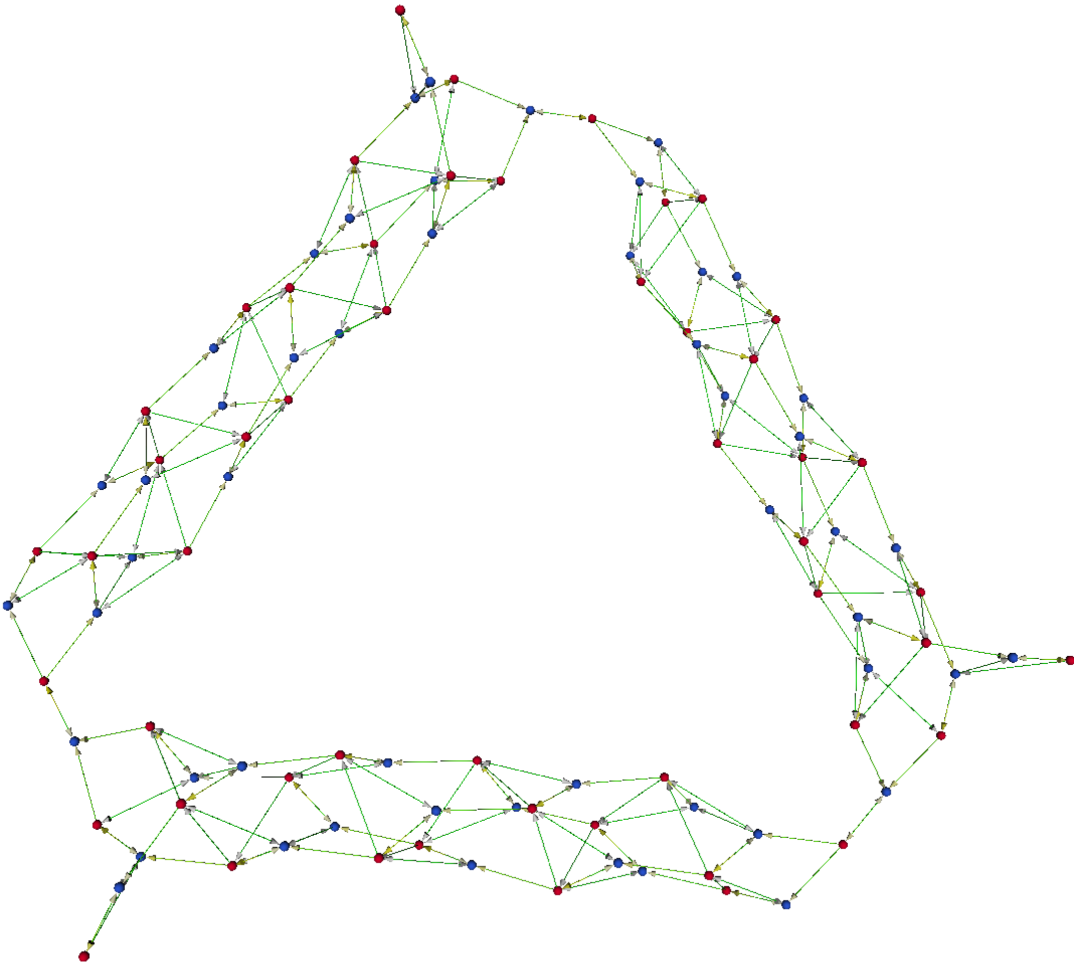


Figure 5.19. Representation of Primary Amide Charge Transfers and Inter-Amide Hydrogen Bond Charge Transfers to Show Hydrogen Bonded Chains of Amides of Figure 5.18. Red Spheres Amide O, Blue Spheres Amide N. Primary Amide CT Shown as From N to O, Inter-Amide CT as O to N

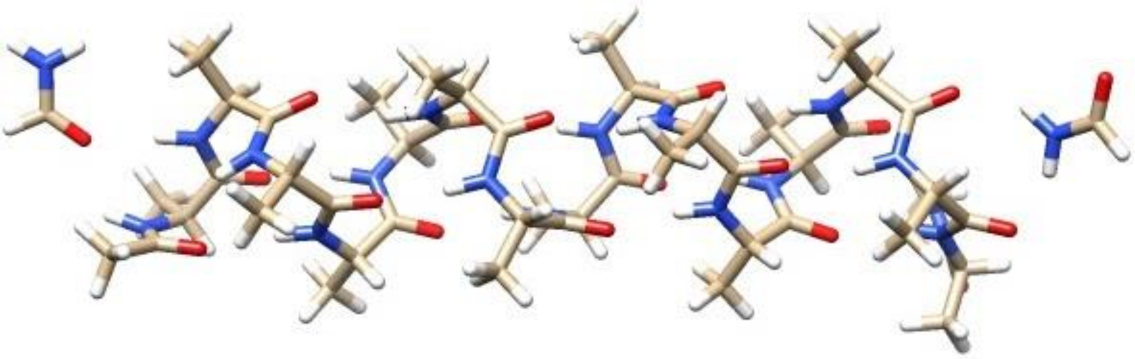


Figure 5.20. ACE-Ala(15)-NME Alpha Helix with a Formamide Hydrogen Bonded at Each Terminus at LC-wPBE(w=0.4)/6-311++G**

Table 5.1. Donor-Acceptor Interactions for 1 Alpha Helix versus 3 Alpha Helices with some Spines Connected by Formamides

helices	hbp	hbs	hbpr	hbsr	hbpl	hbsl	respb	ressb	resps	resss
1	22.6	67.87	4.88	5.57	4.69	6.27	1625.62	24.89	108.02	0.21
3	28.01	90.93	10.79	13.03	8.77	12.18	1693.11	24.92	95.93	15.53

- hbp: O(lp-p)->H-N* kcal/mol average of intra helix totals
- hbs: O(lp-s)->H-N* kcal/mol average of intra helix totals
- hbpr: O(lp-p)->H-N* kcal/mol average of helix to formamide HB CT totals
- hbsr: O(lp-s)->H-N* kcal/mol average of helix to formamide HB CT totals
- hbpl: O(lp-p)->H-N* kcal/mol average of formamide to helix HB CT totals
- hbsl: O(lp-s)->H-N* kcal/mol average of formamide to helix HB CT totals
- respb: amide resonance N(lp)->C-O(pi)* kcal/mol average of helix totals
- ressb: amide resonance N(lp)->C-O(sigma)* kcal/mol average of helix totals
- resps: amide resonance N(lp)->C-O(pi)* kcal/mol average of formamide totals
- resss: amide resonance N(lp)->C-O(sigma)* kcal/mol average of formamide totals

The 1 helix case is shown in Figure 5.20 and the 3 helix case in Figure 5.18.

5.7.6 Artificial beta sheets

Appendix 2 (Section 5.11) provides tabulation of polyvaline parallel beta sheet backbone amide N(lp)->C-O(p)* and N(lp)->C-O(s)* SOPT energy values for 4 backbone HB chain, 4 beta strand beta sheets, capped with formamides, some configurations forming HB cycles. The particular configuration for each experiment is shown in the image prefacing each experiment. The image is laid out so that the position of the backbone amides correspond to the positions of the cells in the tables for the experiment.

Appendix 3 (Section 5.12) provides the same manner of tabulation for polyvaline antiparallel beta sheet capping, and includes some examples of water capping the beta sheets though this study is focussed on cycles of protein.

In these experiments, each beta sheet including capping molecules was geometry optimized without constraint at LC-wPBE(w=0.4)/6-311G**. TeraChem was used for the geometry optimization (its method name for LC-wPBE is wpbe) and Gaussian for the production of the NBO 6 GenNBO input .47 file.

As mentioned above, non-zero N(lp)->C-O(s)* values are purely DFT error. While the total of this value is large for each experiment (Ap2:Table 5.18 for parallel sheet and Ap3:Table 5.29 for antiparallel sheet), it is the distribution of these errors that is most alarming. Inspection of the 10 N(lp)->C-O(s)* tables in Appendices 2 and 3 gives that notable instances of this error are sparse rather than evenly distributed. This is alarming because a N(lp)->C-O(p)* value in an RAHB chain is partially determined by the other such values in the chain, even where such a value is not associated at the same amide with a significant N(lp)->C-O(s)* value. It is clear that a large reduction in N(lp)->C-O(p)* value is associated with a large N(lp)->C-O(s)* value at an amide, so one large N(lp)->C-O(s)* in an RAHB chain invalidates all N(lp)->C-O(p)* values in the chain. 6-311G** can be taken to be unsatisfactory for use with beta sheets. This error should be taken as influencing geometry optimization as well as single point orbitals. In the presence of such errors, we do not proceed to discuss the tabulated values for cyclic hydrogen bonding as relevant to the study of nature. The study of RAHB cyclization in beta sheets awaits methods that do not suffer the errors described in Chapter 3 and otherwise give a good account of RAHB and computationally scale to the atom count of beta sheets.

5.7.7 Cyclic HB in the Protein Data Bank

A subset defined below of the NMR entries of the PDB was searched for cyclic HB in proteins preparatory to investigating any cooperativity in these cycles with quantum chemical methods. Non-standard residues, ligands and solvent were not considered. Data derived from NMR spectroscopy but not X-ray crystallography was used, for the protonation state of residues was required and the observed coordinates of these protons desirable. The restriction of the search to NMR spectroscopy-derived data biases the search toward smaller proteins.

Only the first model in a PDB file was considered. Also, PDB files were excluded if any error in the file was detected. The programs for spatial query of the PDB were intended to support a broad range of queries that are as straightforward as possible, and the principle that queries should not have to handle error or exception conditions was adopted, so files with these conditions were excluded from querying. Conditions evaluated were: values expected to be integer actually integer, residue sequence numbers in a chain contiguous, no atoms missing, helix start residue sequence number less than helix end residue number, helix type in the range 1..10, start and end residues of helices and beta strands being of the same chain, no secondary structures overlapping and the backbone nitrogen at the N-terminus of a chain having 3 hydrogens.

Each potentially cooperative unit was modelled as a list of polar hydrogens and a list of HB acceptor (charge transfer donor) atoms, with the connection between hydrogen and acceptor atoms within the cooperative unit left abstract until later selection for quantum chemical analysis. Determination of the extent of cooperativity was similarly deferred, and all protein HB cycles regardless of cooperativity were captured in the PDB extraction pass.

Modelling of backbone amides and sidechain amides as cooperative units was straightforward. Aspartate and glutamate sidechains have only acceptors unless a proton is bound which then allows two cooperative functional units, one being the new hydroxyl, the other being a larger unit H-O-C-O which somewhat resembles the amide H-N-C-O. Singly protonated histidine has both an acceptor and a polar hydrogen. The hydroxyl groups of serine, threonine and tyrosine were modelled as cooperative units.

The maximum HB length is taken to be 3.0 angstroms and X-H..A angle of 45 degrees as an HB in a significantly cooperative system will have length and angle appreciably less than these figures which then form an upper bound.

8378 PDB files were extracted from the RCSB Protein Data Bank on 2014-9-8 by Advanced Search with the following parameters

- Macromolecule
    - Contains Protein = Yes
    - Contains DNA = No
    - Contains RNA = No
    - Contains DNA/RNA Hybrid = No
- Experimental Method
    - Solution NMR
    - Has Experimental Data = Ignore
- Has Modified Residue(s) = No

This can be closely reproduced by further specifying a Release Date of up to 2014-9-8, returning 8377 files, the discrepancy of 1 perhaps attributable to a new file being made available during that day.

1690 files of these 8378 PDB files were filtered out as having error or exception conditions and the first model of the remaining 6,688 files was returned for querying. These were programmatically analysed for cycles of HB, and the resulting data appears as Table 5.2.

Table 5.2. Cycles of hydrogen bonding in proteins

2U2F	BA	GLY	7	A	Strand=(1,S3)	BA	VAL	75	A	Strand=(1,S4)	BA	LEU	74	A	Strand=(1,S4)	BA	GLY	8	A	Coil
2M89	SC	ASN	115	B	Helix=(RH Alpha,8)	SC	ASN	115	A	Helix=(RH Alpha,4)										
2LKD	SC	ASN	53	A	Coil	BA	ILE	57	A	Coil	BA	LYS	56	A	Coil					
2KQ2	BA	ASP	12	A	Coil	SC	ASN	124	A	Helix=(RH Alpha,7)	BA	GLY	13	A	Coil					
2KHQ	SC	GLN	53	A	Helix=(RH Alpha,5)	SC	ASN	57	A	Helix=(RH Alpha,5)										
2K2W	SC	ASN	262	A	Helix=(RH Alpha,3)	SC	THR	265	A	Strand=(3,A)										
2HD7	SC	GLN	202	A	Strand=(2,A)	SC	HIS	237	A	Strand=(3,A)										
2GIW	SC	ASN	52	A	Helix=(RH Alpha,2)	SC	THR	78	A	Coil										
2GD7	SC	ASN	53	A	Helix=(RH Alpha,2)	SC	ASN	53	B	Helix=(RH Alpha,5)										
2FKI	SC	GLN	101	A	Helix=(RH Alpha,3)	SC	ASN	105	A	Helix=(RH Alpha,3)										
1Z1D	SC	ASN	203	A	Coil	SC	ASN	210	A	Helix=(RH Alpha,1)										
1W4U	SC	GLN	34	A	Strand=(1,AA)	SC	THR	70	A	Coil	SC	THR	53	A	Strand=(2,AA)					
1QCE	SC	ASN	37	B	Helix=(RH Alpha,3)	SC	ASN	82	C	Helix=(RH Alpha,6)										
	SC	ASN	82	A	Helix=(RH Alpha,2)	SC	ASN	37	C	Helix=(RH Alpha,5)										
	SC	ASN	37	A	Helix=(RH Alpha,1)	SC	ASN	82	B	Helix=(RH Alpha,4)										
1OPZ	SC	THR	57	A	Helix=(RH Alpha,3)	SC	GLN	61	A	Helix=(RH Alpha,3)										
1M7L	SC	ASN	77	B	Coil	SC	ASN	37	A	Coil	SC	ASN	117	C	Coil					
1ESX	SC	ASP	7	A	Coil	SC	ASP	17	A	Helix=(RH Alpha,1)										
	BA	GLN	3	A	Coil	SC	GLN	11	A	Coil	SC	GLN	8	A	Coil	BA	ALA	4	A	Coil
1EDL	SC	GLN	24	A	Helix=(RH Alpha,2)	SC	GLN	53	A	Coil										
1EDK	SC	GLN	24	A	Helix=(RH Alpha,2)	SC	GLN	53	A	Coil										
1EDI	SC	GLN	24	A	Helix=(RH Alpha,2)	SC	GLN	53	A	Coil										
1DPU	SC	ASN	203	A	Coil	SC	ASN	210	A	Helix=(RH Alpha,1)										
1A03	BA	GLU	86	B	Coil	BA	GLY	90	B	Coil	BA	LYS	89	B	Coil	BA	ALA	87	B	Coil

- Each row is a cycle. Up to 4 cooperative units were found per cycle.
- A PDB model may contain multiple cycles.
- BA/SC gives whether the unit is a backbone amide or a sidechain.
- The 3 character residue code is followed by the residue sequence number and chain ID.
- The type of a helix is identified, followed by its ID.
- For a beta strand, the strand number is given followed by the sheet ID.

The outstanding features of these data are how few cycles there are in the structures considered and, where present, how small these cycles are. When cycles consisting of just two sidechain amides are excluded, only 7 cycles remain, and only 3 of these are cycles of 4 potentially cooperative units. This striking result prompted review of the program used to generate the data, and the code used to determine HB connection between cooperative units was simplified to become the quite brief Haskell list comprehension mentioned above, but the results did not change. In the absence of formal proof of correctness or lesser corroboration, any program must be held to have the possibility of error. However, given the results of this program, it must be concluded that for standard proteins of size amenable to NMR spectroscopy i.e. less than ~30 kDa in mass, potentially cooperative HB cycles purely of protein are all but completely absent, so largely so that it is suggested that the absence of significant cycles is a fundamental property of protein structure. The potentially cooperative cycles that were detected serve as some test of the program's ability to detect at least cycles of amides. Note that the cycles detected by this program are not necessarily cooperative even though are comprised of potentially cooperative units. Determination of whether a HB cycle of potentially cooperative units is actually cooperative rests with quantum chemical analysis.

Two of the cycles that involve four potentially cooperative units, 1A03 and 1ESX, of Table 5.2 were isolated from their native protein context and were subject to quantum chemical energy minimization/geometry optimization at LC-wPBE(w=0.4)/6-311G** with TeraChem, retaining near-native geometry. Gaussian single point energy calculation of the resulting coordinates at the same method and basis were used for input to NBO 6.0 analysis. The 1A03 cycle is of backbone amides, and the 1ESX cycle is of backbone and sidechain amides. The HB and amide resonance data of these cycles appears in Table 5.3 and Table 5.4 and depiction of the geometries appears in Figure 5.21 and Figure 5.22. The extract of residues 83-90.B of PDB 1A03 has an HB between the backbone amides at the junctions of residues 84:85 and 87:88, but this is not involved in the cycle. This hydrogen bond is of length 1.829 angstroms with O(lp-s)->H-N* of 3.51 kcal/mol and O(lp-p)->H-N* less than 0.1 kcal/mol, and in total likely contributes to the stability of the structure of the extract. The tabled (Table 5.4) cycle of the extract of residues 2-12.A of PDB 1ESX contains a shorter cycle created by an HB with charge transfer from atom 10 (oxygen) to atom 95 (nitrogen) not shown in Table 5.4 but appearing in Figure 5.22, but its O(lp-s)->H-N* has associated energy of 0.47 kcal/mol and its O(lp-p)->H-N* 1.15 kcal/mol and so is weaker than the larger cycle.

The least HB in these cycles has very low energy charge transfers, and it may be concluded that the cyclic charge transfer is also very low. Only one backbone amide has a N(lp)->C-O(p)* value in keeping with RAHB for the method used, that being a 109.85 kcal/mol in the 1ESX extract. Thus, according to the quantum chemical methods used, these cycles are not cooperative. Consequently, the biological significance of these cycles is not investigated.

Table 5.3. Hydrogen Bonding Cycle in PDB Entry 1A03 extract of residues 83-90.B for LC-wPBE(w=0.4)/6-311G**

Residue	N	N(lp)->C-O(s)*	N(lp)->C-O(p)*	O	O(lp-s)->H-N*	O(lp-p)->H-N*
BA(85,86)	14	6.86	84.96	10	1.67	3.26
BA(86,87)	21	1.9	100.99	17	2.47	0.31
BA(88,89)	35	0.81	101.39	31	2.26	3.99
BA(89,90)	42	3.8	100.28	38	7.45	2.9

- BA(X,Y) : Backbone Amide formed at junction of residue numbered X and Y
- SC(X,Y) : Sidechain Amide of residue type X number Y
- N and O : Atom Ids of Amide Nitrogen and Oxygen in extract for electronic structure calculation
- Labels including "->" : Donor-Acceptor Interaction SOPT Energy in kcal/mol
- O Interaction is with H-N of N in the line below, O of last line to N of first line

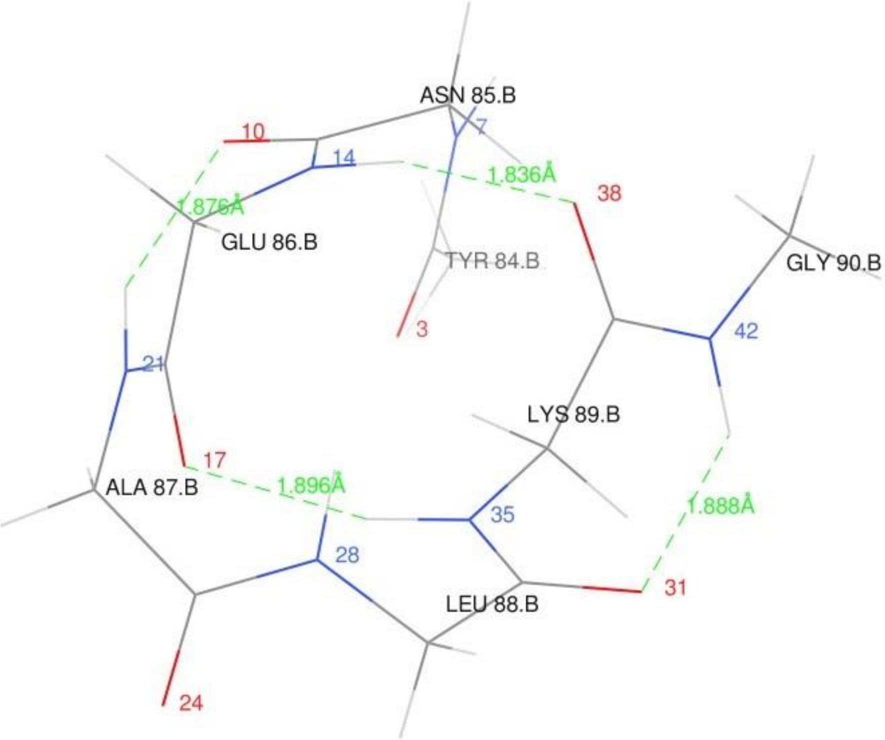


Figure 5.21. Hydrogen Bonding Cycle in PDB Entry 1A03

Table 5.4. Hydrogen Bonding Cycle in PDB Entry 1ESX extract of residues 2-12.A for LC-wPBE(w=0.4)/6-311G**

Residue	N	N(lp)->C-O(s)*	N(lp)->C-O(p)*	O	O(lp-s)->H-N*	O(lp-p)->H-N*
BA(2,3)	7	3	91.96	3	1.93	3.09
BA(3,4)	14	1.79	109.85	10	4.48	1.89
SC(GLN,8)	57	0.13	104.94	56	5.96	6.76
SC(GLN,11)	95	9.34	74.56	94	6.07	11.16

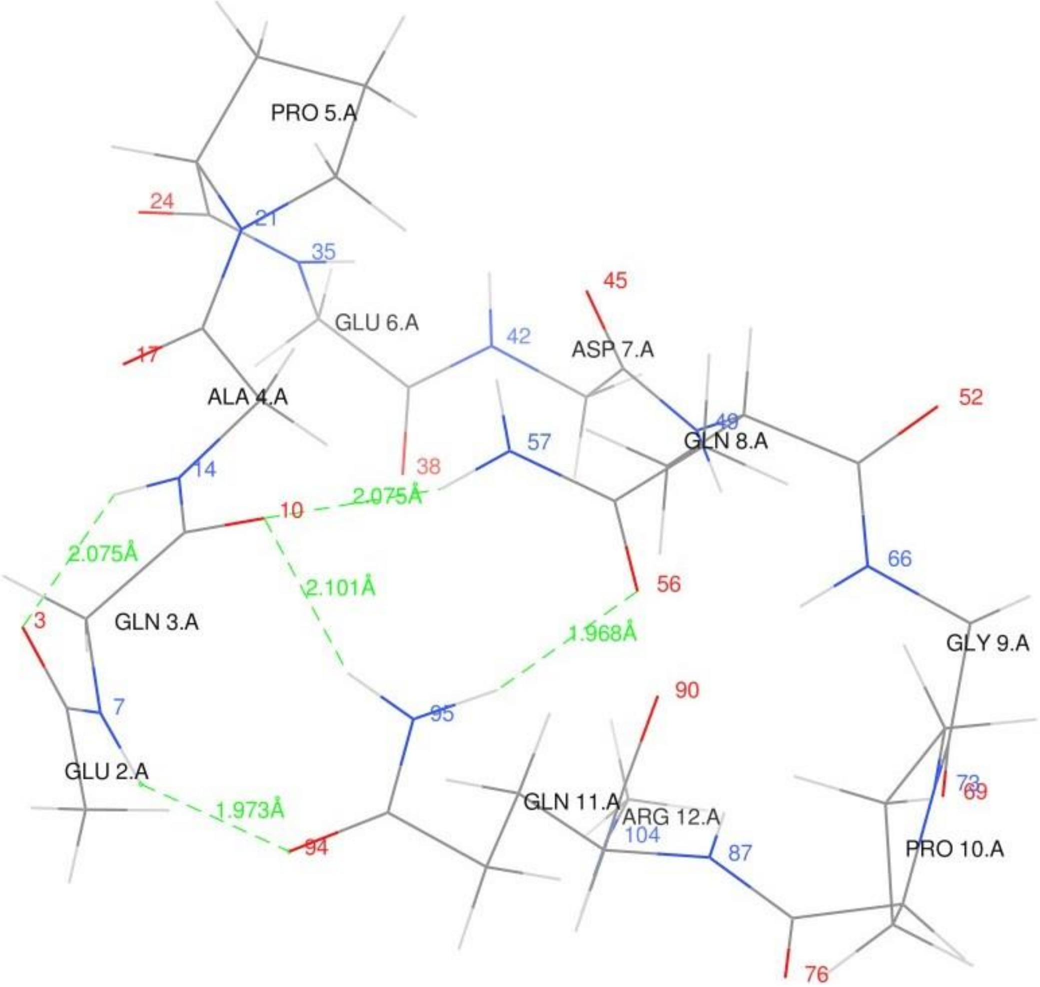


Figure 5.22. Hydrogen Bonding Cycle in PDB Entry 1ESX

5.7.8 Joined HB chains of beta sheet

Partial confirmation of the above finding was sought by querying the PDB extract without using detection of cycles. This extract was queried for asparagine or glutamine sidechain amides connecting any two beta sheet backbone amide HB chains without the requirement of cyclization, leaving detection of cycles to inspection. These queries do not require that asparagine and glutamine connect at the end of beta sheet HB chains, merely that they connect to beta sheet backbone amides. However, it is unlikely that a sidechain amide would bind to 2 beta sheet backbone amides other than at the ends of beta sheet HB chains, for such binding would weaken the beta sheet HB chains. These queries do not require that the beta sheet HB chains that are connected be in the same beta sheet. Such an asparagine or glutamine does not itself have to part of a beta sheet, or any other secondary structure. Each asparagine or glutamine sidechain amide was permitted to have multiple HB donating and accepting beta sheet backbone amides, but the results of the query were such that there were only one acceptor and donor under the defined criteria for HB which were the same as in the previous experiment.

These queries were applied to the same extract from the PDB as the previous experiment, and results were again striking. There were no instances of an asparagine sidechain amide connecting two beta sheet backbone HB chains, since there were no instances of an asparagine sidechain amide connecting two beta sheet backbone amides.

Only 18 instances of glutamine sidechain amides connecting beta sheet backbone amides exist in the models considered (Table 5.5). For a beta strand to be capped such that adjacent backbone amide RAHB chains of a beta sheet are HB connected by a sidechain amide, two configurations are possible, one in which the residue sequence numbers of the residues being HB connected do not differ, that is, the N-H and C-O are of the same residue, and the second configuration in which the residue sequence numbers of the beta strand differ by 2. 9 of these 18 instances are of the first configuration and 5 are of the second configuration. There are no instances of cycles being formed by sidechain amides and beta sheet backbone amides. The results of this experiment are consistent with the previous and more general experiment, giving some support to the view that no error exists in these programs.

Table 5.5. Glutamine sidechains connecting beta sheet backbone amides

3AIT	GLN	16	A	Strand=(1,S1)	ASP	58	A	Strand=(3,S1)	ALA	67	A	Strand=(1,S2)
2MA8	GLN	28	A	Helix=(RH Alpha,1)	ILE	75	B	Strand=(2,B)	ALA	77	B	Strand=(2,B)
2LQ7	GLN	421	A	Strand=(3,A)	THR	384	A	Strand=(4,A)	GLU	422	A	Strand=(3,A)
2LGR	GLN	61	A	Coil	VAL	70	A	Strand=(6,A)	VAL	70	A	Strand=(6,A)
2LBT	GLN	212	A	Helix=(RH Alpha,1)	ILE	323	A	Strand=(3,A)	VAL	325	A	Strand=(3,A)
2L7Q	GLN	52	A	Coil	GLY	57	A	Strand=(4,A)	GLY	57	A	Strand=(4,A)
2KQM	GLN	206	B	Strand=(1,C)	TYR	286	B	Strand=(3,D)	CYS	288	B	Strand=(3,D)
2KMX	GLN	1177	A	Strand=(4,A)	LYS	1122	A	Strand=(5,A)	TYR	1178	A	Strand=(4,A)
2KMN	GLN	131	A	Helix=(RH Alpha,3)	ALA	103	A	Strand=(4,A)	ALA	103	A	Strand=(4,A)
2KFS	GLN	133	A	Helix=(RH Alpha,7)	THR	112	A	Strand=(1,B)	THR	112	A	Strand=(1,B)
2K3K	GLN	70	A	Helix=(RH Alpha,4)	ILE	81	A	Strand=(4,A)	ILE	81	A	Strand=(4,A)
2CQO	GLN	92	A	Coil	TRP	68	A	Strand=(5,A)	TRP	68	A	Strand=(5,A)
2BBI	GLN	48	A	Strand=(2,B)	CYS	49	A	Strand=(2,B)	ARG	28	A	Strand=(1,B)
1WF9	GLN	40	A	Helix=(RH Alpha,1)	SER	22	A	Strand=(1,A)	SER	22	A	Strand=(1,A)
1QNZ	GLN	118	H	Strand=(1,D)	TYR	206	H	Strand=(2,E)	CYS	208	H	Strand=(2,E)
	GLN	6	L	Strand=(1,A)	TYR	90	L	Strand=(1,C)	CYS	92	L	Strand=(1,C)
1MKE	GLN	94	A	Strand=(1,A)	GLU	95	A	Strand=(1,A)	GLU	95	A	Strand=(1,A)
1L2M	GLN	31	A	Helix=(RH Alpha,1)	VAL	79	A	Strand=(1,B)	VAL	79	A	Strand=(1,B)

- The first outlined block identifies the residue with the connecting amide sidechain.
- The second outlined block identifies an HB accepting backbone amide.
- The third outlined block identifies an HB donating backbone amide.
- Multiple connections may occur with each PDB model.

Greater secondary structure stability is not necessarily biologically advantageous, but it seems unlikely that if greater stabilization were available that there would be no situation in which it would be advantageous, such as in proteins in which mechanical strength is crucial. Given this assumption and this PDB extract, it could be concluded that sidechain amide cyclization of beta sheet backbone HB is not stabilizing.

A hypothesis concerning stability of beta sheets was stated in Section 4.10.1, being that balance between the RAHB of the backbone amide HB chains favours the stability of beta sheets. Following this hypothesis, cooperative cyclization of a subset of the RAHB chains of a beta sheet would reduce stability.

In beta barrels such as [212] the RAHB chains are not cyclized, since the RAHB chains do not follow a diameter of the barrel. With sufficient extension of the length of the barrel by increasing the length of the beta strands, cyclization of some RAHB chains is then possible. Following the above hypothesis of the stability of beta sheets, this selective cooperative cyclization would reduce stability of the sheet. If examples of such long beta barrels exist in the PDB and were derived by X-ray crystallography [165], they would not have been detected by this survey which depends upon NMR to determine protonation state.

5.8 Conclusion

Unlike the lone pairs of the oxygen atom in water, the lone pairs of the carbonyl oxygen substantially maintain inequivalence in the presence of hydrogen bonding. One lone pair of carbonyl oxygen is p-type, has 2 lobes and is of higher energy than the s-rich lone pair. This is demonstrated by NBO analysis, and its calculations are not unitarily equivalent to a model in which the lone pairs are equivalent following sp2 hybridization [83, 205]. The higher energy of the p-type lone pair can donate more charge to an H-X* bond, depending on the geometric relation between the lobes of the p-type lone pair and the H-X* bond. This charge transfer is largest when the H-X* bond is in the plane of the oxygen p-type lone pair lobes and at a minimum when X is in the plane normal to that plane and passing through C and O.

This inequivalence of carbonyl oxygen lone pairs gives a better account of optimal C-O..N hydrogen bond angle, as an account with equivalent sp2 lone pairs could be expected to give optimal hydrogen bonds when the C-O..N angle is ~60 degrees from linear in the amide plane, whereas the shortest hydrogen bond length is found at ~75 degrees and is kept from optimality at angles yet further from linear by environmental steric interactions rather than intrinsic properties of the lone pairs. This geometry preference is not in keeping with a sp2 account of carbonyl oxygen lone pairs or a primarily dipolar account of HB.

Where classical molecular simulators model lone pairs of carbonyl/amide oxygen lone pairs, improvement in the accuracy of modelling could be had by capturing the inequivalence of the lone pairs, though this alone would not model amide resonance variation of the energy level and occupancy of these lone pairs.

A remarkable feature of amide-amide hydrogen bonding is that when the inter-amide steric exchange energy is deducted from the inter-amide donor-acceptor energy the result is close to zero, with slow increase as C-O..N becomes less linear in the plane of the amide having the C-O. The energy change associated with hydrogen bonding between amides primarily resides in variation of the resonance of the amides. The direct inter-amide NBO interactions are significant only in that they are a conduit for changes to the resonance of the amides. A particular consequence of this is that the majority of the energy associated with hydrogen bonding between amides is subject to electrostatic field via variation of amide resonance (Chapter 4), increasing the significance of electrostatic field in determining protein structure.

The lobes of an amide oxygen p-type lone pair may concurrently engage in hydrogen bonding with different H-X* bonds. While this is anti-cooperative due to the busy donor effect, depending on C-O..N angles the total amount of charge donated from an amide is nonetheless substantially increased over the single hydrogen bond case. The increase in amide resonance due to both hydrogen bonds being present reduces the impact of the busy donor effect. Provided the two H-X* bonds are in the amide plane of the charge donating oxygen, the resonance stabilization can be greater than that for the case in which the oxygen participates in a single hydrogen bond, depending on the nature of resonant moieties extending from the hydrogen bonds.

While the focus of the search for useful materials has moved from HB structures such as polyamides to covalently bonded cylinders and sheets [211], it may be that useful HB materials may still be found with hydrogen bonded structures in which the lobes of the carbonyl p-type lone pair are used for separate or bifurcated hydrogen bonds, for the increase in total charge transfer means increased resonance and hence stability of the material, provided the carbonyl is part of a resonant system in the manner of the amide group. Constraints on the search for such materials were discussed above.

Cycles of coplanar amides become less energetically favourable per hydrogen bond as the number of amides in the cycle grows, for in the case of a regular C-O..N angle this angle becomes closer to linear, and in the case of irregular C-O..N angles the cooperativity is limited by the most linear of these angles. The remaining possibility is for the amides to not be co-planar, which does not mean that hydrogen bonding is not suboptimal, for all that is required is for the H-N to be in the plane of the amide bearing the charge donating oxygen, which allows the amide bearing the H-N to be rotated about its H-N axis since the H-N orbitals are symmetric about that axis. This has the additional possible benefit of reducing steric conflict introduced by amide substituents such as the CA atoms and its substituents. We have not proposed a cyclic example of this configuration, nor suggested a regular bending of backbone chains that would accomplish this. Bearing in mind the limitation of the cooperativity to the least cooperation in the cycle, a regular bend of the backbone seems necessary for a cycle of backbone amides. In a mixed cycle of backbone and sidechain amides, the lower resonance of the sidechain amides will be limiting. Moieties with less hydrogen bonding cooperativity than amides introduce further limitation of cyclic cooperativity.

Cooperativity of cyclic HB can be diminished through the busy donor effect where an acceptor is extraneous to the cycle. A feature of RAHB secondary structures is that the backbone amide oxygens are largely protected from extraneous interactions. This protection is an additional burden on any regular backbone bend such as mentioned in the previous paragraph. Arrangements of non-planar cycles with regular backbone bends such that cycles can be mutually supportive and protecting are yet more constraining.

Our tentative finding is that cooperative HB cycles in standard proteins are radically disfavoured in nature. The use of data deriving from NMR spectroscopy biases the study toward smaller proteins, but as small cycles can be expected to be more frequently occurring than large cycles, no potentially cooperative cycles of more than 4 units being detected in the selected NMR data and the 2 4-unit examples investigated by quantum chemical means revealed to not be cooperative at the level of theory used suggest that cooperative cycles of hydrogen bonding of more than 3 units do not exist PDB structures. This observation rests on detection by program of potentially cooperative hydrogen bonded cycles in the PDB. Since the programs used for this detection have not been formally verified as solving the problem to be solved, independent confirmation of this finding is highly desirable.

Possible causes of the near-complete absence of these cycles additional to the foregoing arise from dynamics. In a cooperative HB cycle there is one HB chain, unlike the multiple spines of RAHB helices or RAHB chains of beta sheets which serve to stabilize geometry including HB lengths of the secondary structure. The single chain of a cycle can be expected to be susceptible to thermal geometry variation. Whereas RAHB secondary structures tend to be islands of stability in protein geometry fluctuation, the HB of cooperative cycles will transiently suffer stretching, and cooperativity is lost rapidly with increased HB length due to rapid reduction in charge transfer with distance. This local loss of cooperativity will propagate around the cycle, attenuating the cooperativity benefit of the cycle.

A hypothesis of the stability of beta sheets (Section 4.10.1) proposes that selective enhancement of RAHB in some backbone amide HB chains of a beta sheet reduces the stability of the sheet, which disfavours cooperative cyclization of some HB chains of a beta sheet. We are presently unable to investigate this proposal due to pronounced errors of established DFT methods as applied to beta sheets (Chapter 3).

Cyclic cooperativity of HB is fragile even in a non-dynamic calculation. The cooperativity of the cycle is limited to the least cooperative unit of the cycle, and the cooperativity at each unit is geometry sensitive. In dynamics, the entropic penalty for maintenance of geometry conducive to cooperativity for all units in the cycle will be high. Acyclic network cooperativity does not have the particular cooperativity advantage of cyclic cooperativity or its fragility. While inter-peptide HB secondary structures such as alpha helices and beta sheets are extensible without loss of the cooperativity already established, extension of a cycle involves breaking the cycle, loss of all cyclic cooperativity and change to the geometry of the cycle to allow a new element to be inserted into the cycle.

It may be that cooperative cycles result in lower density of stabilizing cooperativity than is possible in non-cyclic cooperative structures.

It may also be that evolution has selected in protein a structurally specific polymer to disfavour cooperative cycles so that energetic equality between HB cycles and non-cyclic HB chains does not lead to ambiguity in what fold a given residue sequence specifies. Where modification of the standard protein backbone, such as for beta-peptides [213], or introduction of artificial amino-acid residues results in non-specific structure, it might be investigated whether cooperative cycles of HBs are present.

5.9 Acknowledgements

Prof. John A. Carver is acknowledged for reading this manuscript and offering editing suggestions.

eResearch South Australia is acknowledged for hosting and administering machines provided under Australian Government Linkage, Infrastructure, Equipment and Facilities grants for Supercomputing in South Australia, directing funds to the acquisition of Nvidia Tesla GPU nodes and allocating 64 CPU cores and 256 GB RAM of the NeCTAR Research Cloud (a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy) to the present work.

5.10 Appendix 1

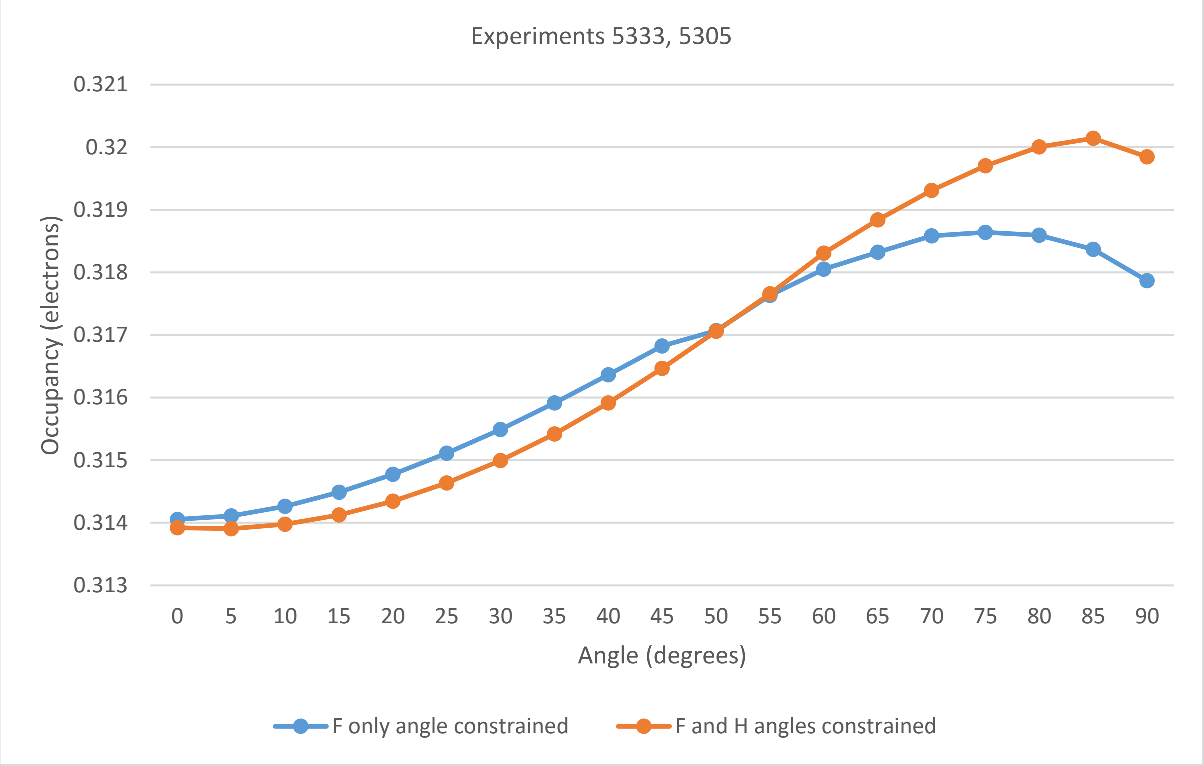


Figure 5.23. C-O(p)* NBO Occupancy in N-methylformamide with HF Hydrogen Bonded and F only or H and F Constrained to Angle from C-O at O in Amide Plane with SCS-MP2/aug-cc-pVTZ

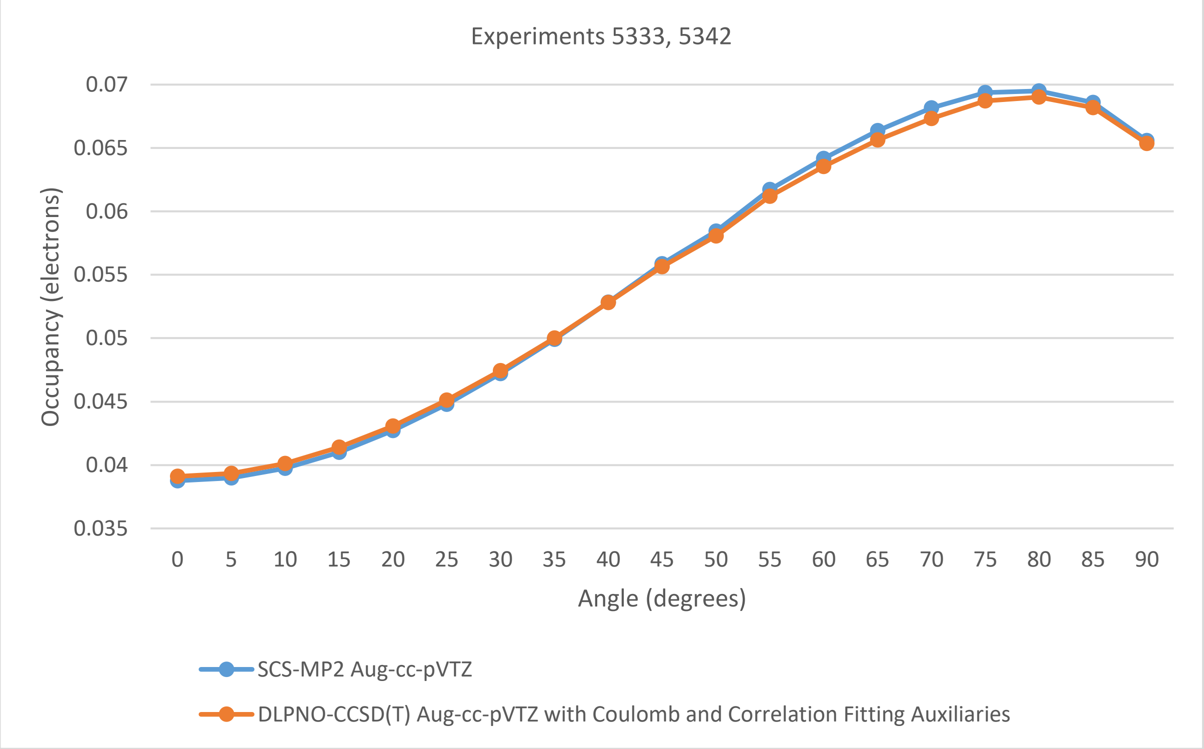


Figure 5.24. H-F* NBO Occupancy with HF Hydrogen Bonded to N-methylformamide O at C-O-F Angle in Amide Plane with Geometry Optimized at SCS-MP2/aug-cc-pVTZ

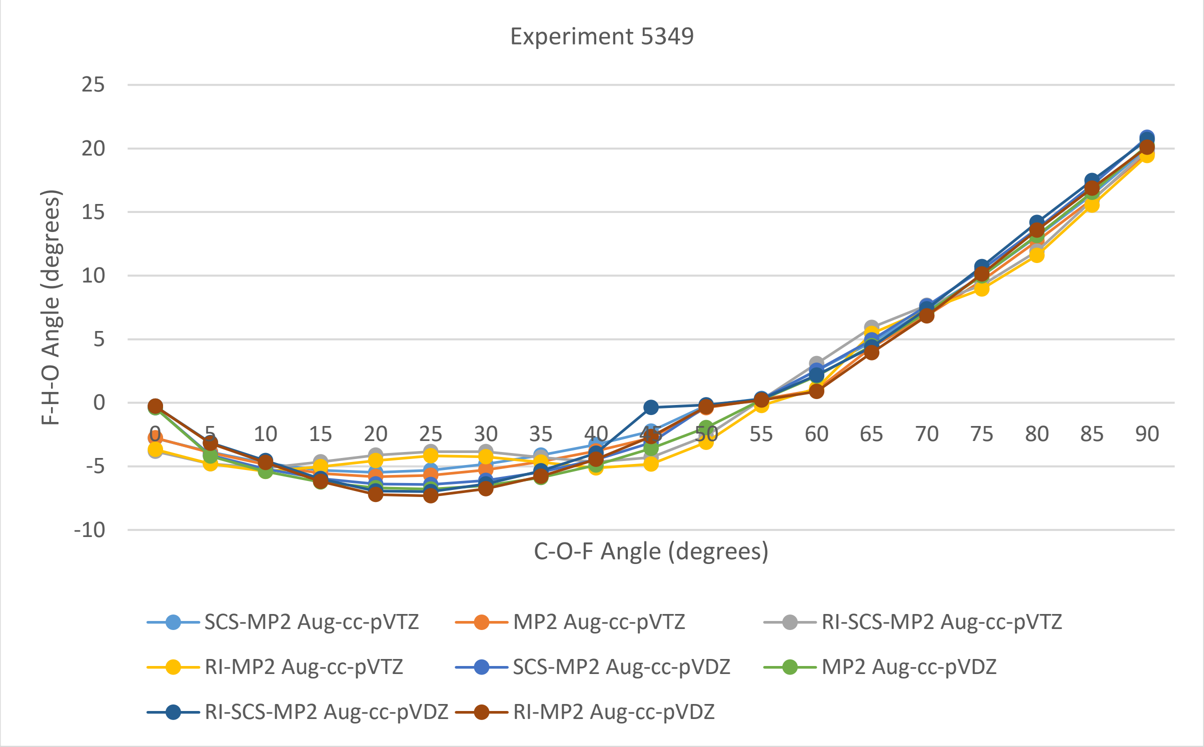


Figure 5.25. F-H-O Angle at C-O-F Angle in HF Hydrogen Bonded to N-methylformamide O in Amide Plane, F-H Pointing Outboard of O as Positive Angle and Between C and O as Negative Angle. RI Methods Used with Coulomb and Correlation Auxiliary Basis Sets

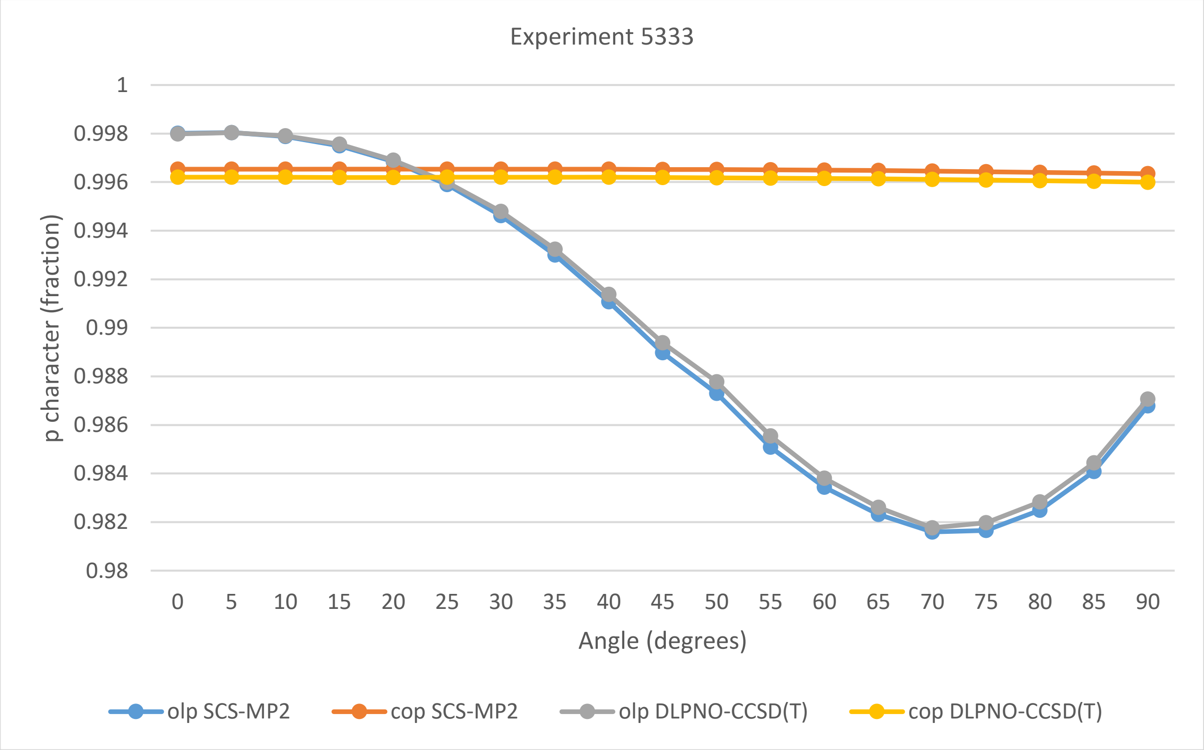


Figure 5.26. O(lp-p) and C-O(p) NBO p Character with N-methylformamide O Hydrogen Bonded to HF at C-O-F Angle in Amide Plane with F Distal to N at SCS-MP2/aug-cc-pVTZ Optimized Geometry and DLPNO-CCSD(T)/aug-cc-pVTZ with Coulomb and Correlation Auxiliaries

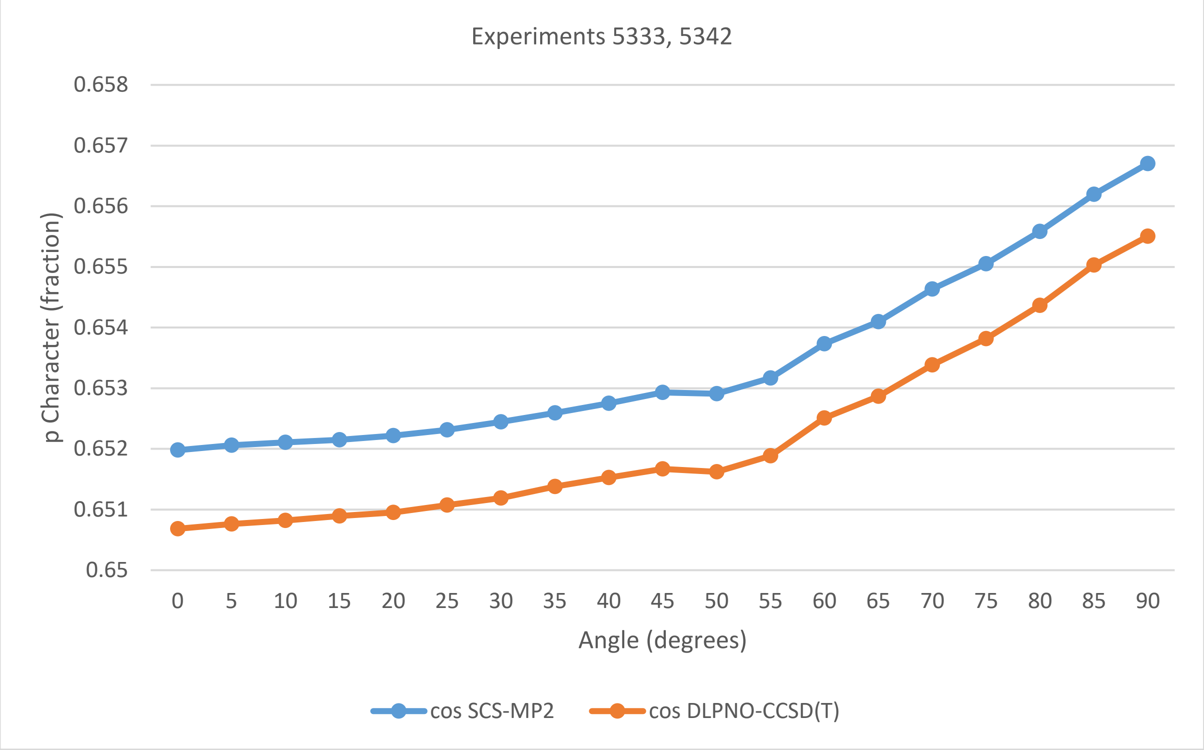


Figure 5.27. C-O(p) NBO p Character in N-methylformamide Hydrogen Bonded to H-F at C-O-F Angle in Amide Plane with F Distal to N at SCS-MP2/aug-cc-pVTZ Optimized Geometry and DLPNO-CCSD(T) with aug-cc-pVTZ and Coulomb and Correlation Auxiliaries

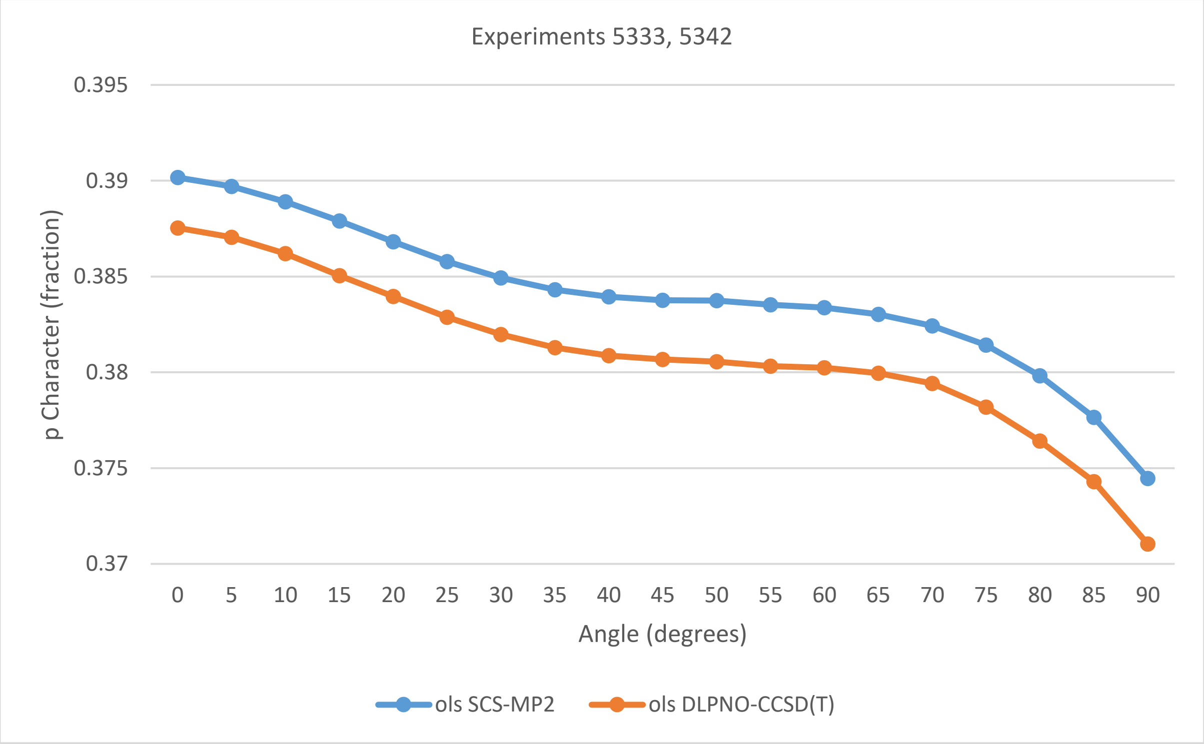


Figure 5.28. O(lp-s) NBO p Character in N-methylformamide Hydrogen Bonded to H-F at C-O-F Angle in Amide Plane with F Distal to N at SCS-MP2/aug-cc-pVTZ Optimized Geometry and DLPNO-CCSD(T)/aug-cc-pVTZ with Coulomb and Correlation Auxiliary Basis Sets

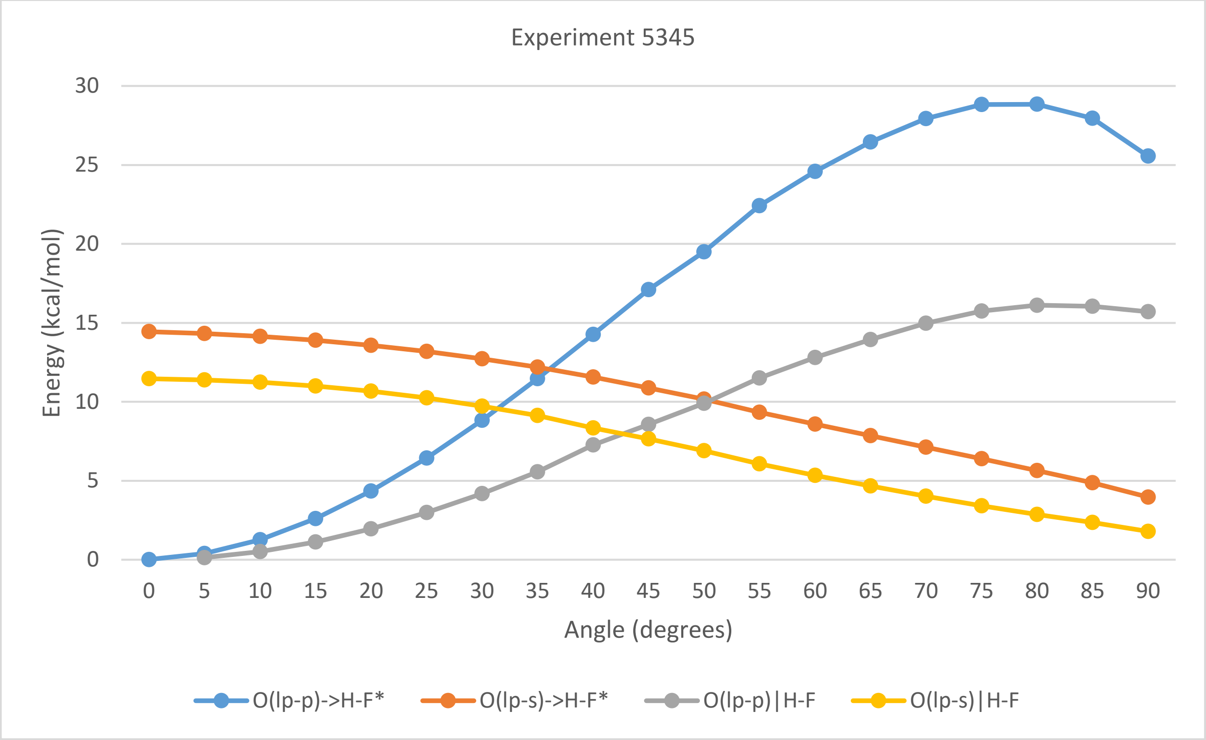


Figure 5.29. Donor-Acceptor SOPT and Steric Exchange Energies in N-methylformamide O Lone Pair Interactions with H-F and H-F* at C-O-F Angle with Geometry Optimized at SCS-MP2/aug-cc-pVTZ with Energetics at LC-wPBE(w=0.4)/aug-cc-pVTZ

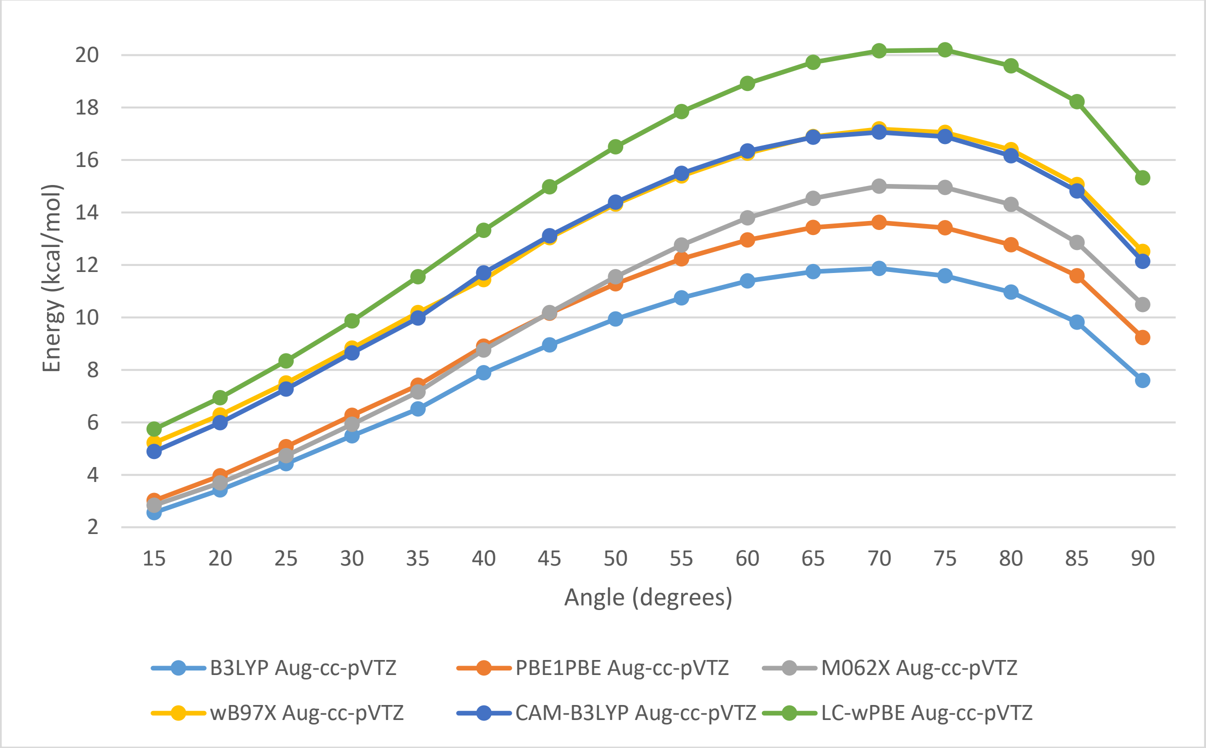


Figure 5.30. Donor-Acceptor SOPT Energy Minus Steric Exchange Energy for Interactions Between N-methylformamide O Lone Pairs with H-F and H-F* NBOs at C-O-F Angle in Amide Plane with Geometry Optimized at Same Method

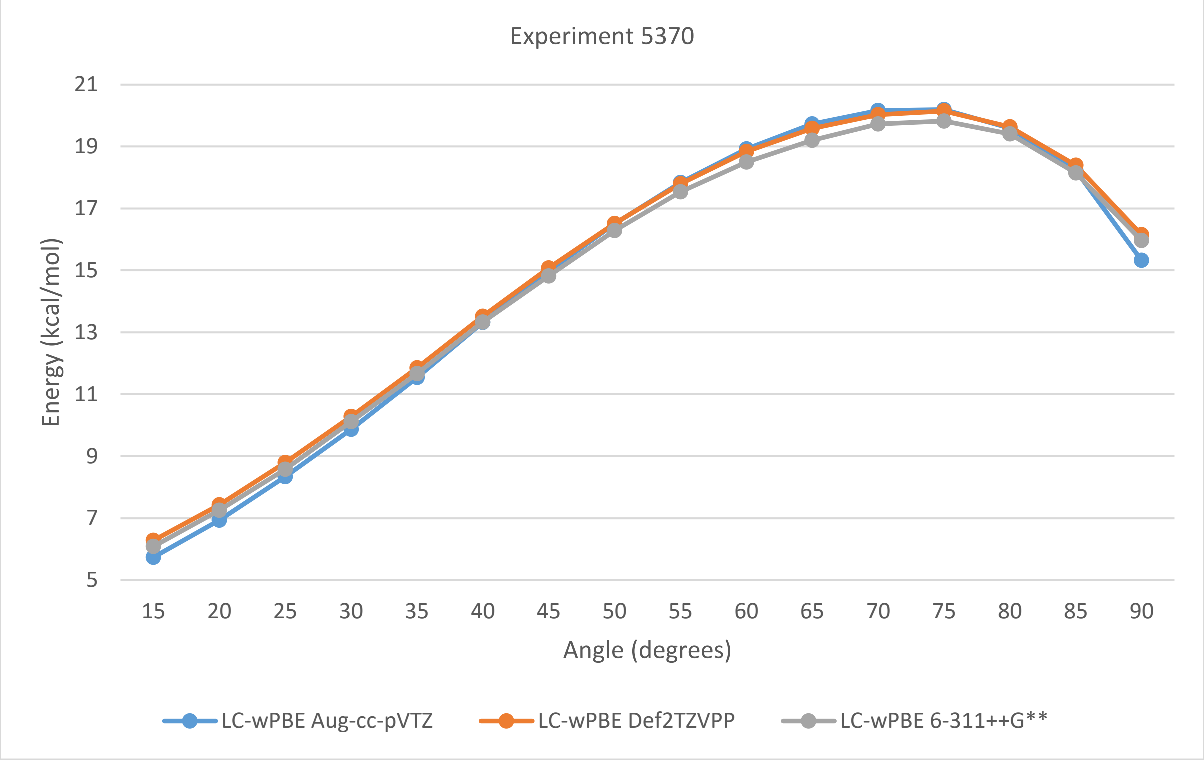


Figure 5.31. Donor-Acceptor SOPT Minus Steric Exchange Energy for N-methylformamide O Lone Pairs Interaction with H-F and H-F* NBOs at Varying C-O-F Angle in Amide Plane at 3 Basis Sets

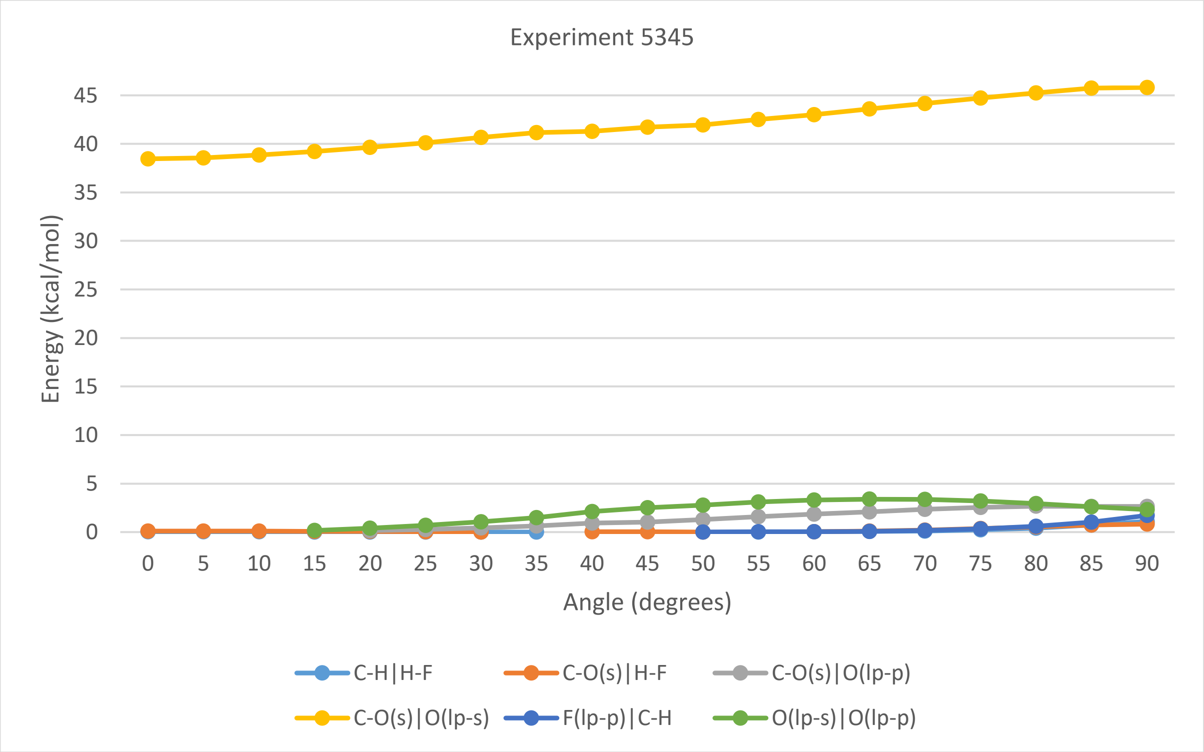


Figure 5.32. Steric Exchange Energy of Interactions Not Directly Involved in Hydrogen Bonding of HF to N-methylformamide O at C-O-F Angle in Amide Plane with Geometry Optimized as SCS-MP2/aug-cc-pVTZ and Energetics at LC-wPBE(w=0.4)/aug-cc-pVTZ

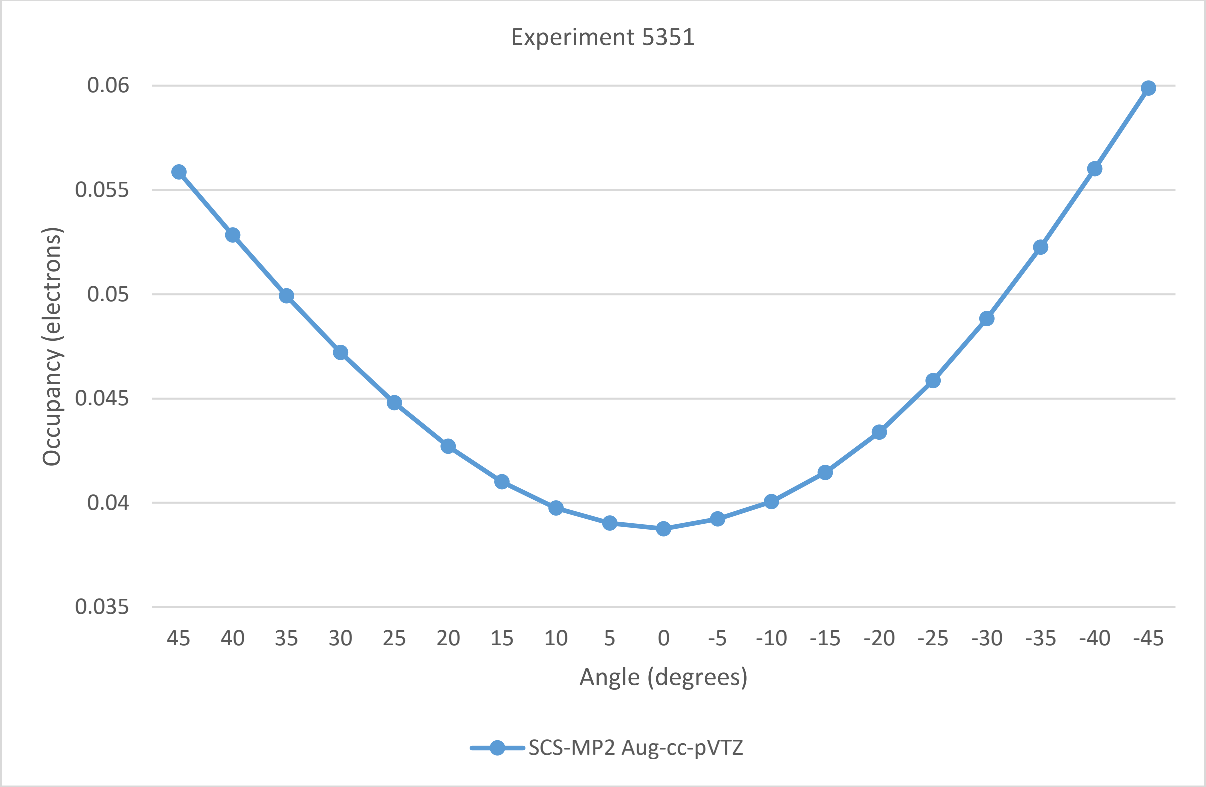


Figure 5.33. H-F* NBO Occupancy in HF Hydrogen Bonded to N-methylformamide O at C-O-F Angle in Amide Plane with Angles Distal to N being Positive

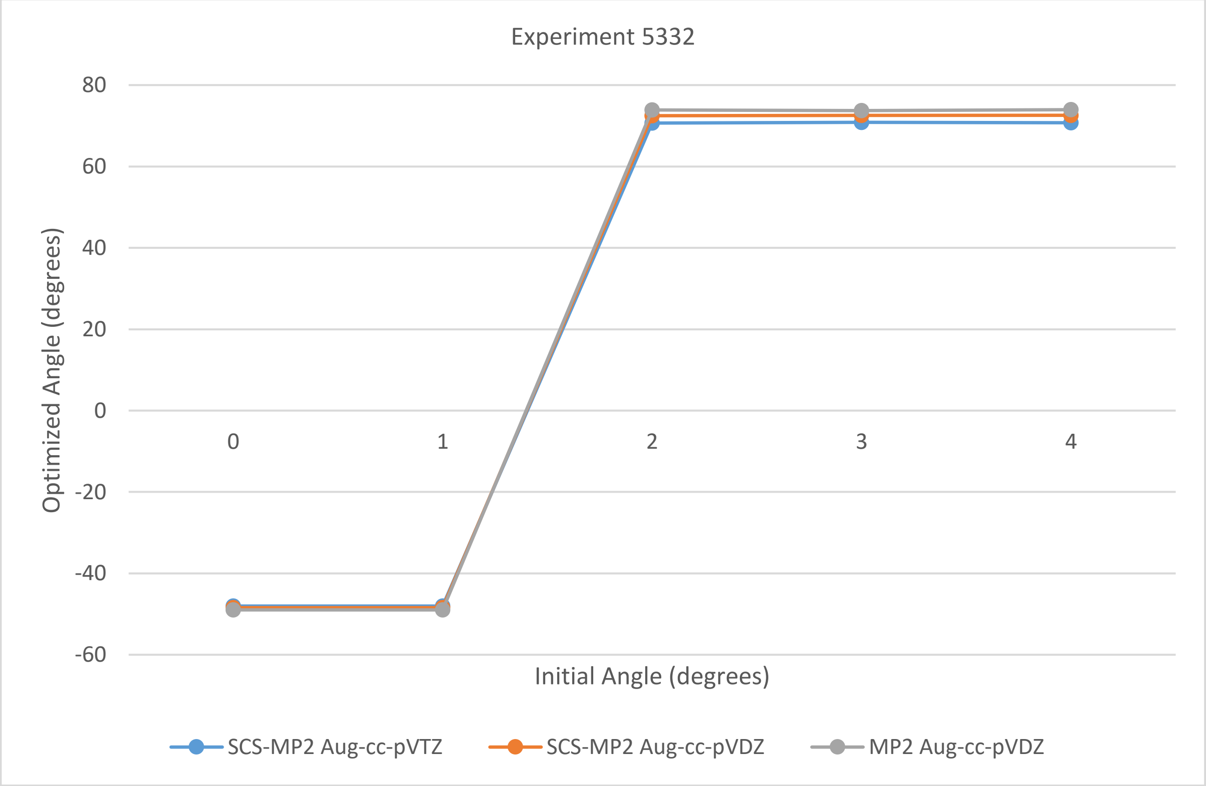


Figure 5.34. Initial versus Geometry Optimized C-O-F Angle in Amide Plane in HF Hydrogen Bonded to N-methylformamide O, Negative Angles for F Proximal to N

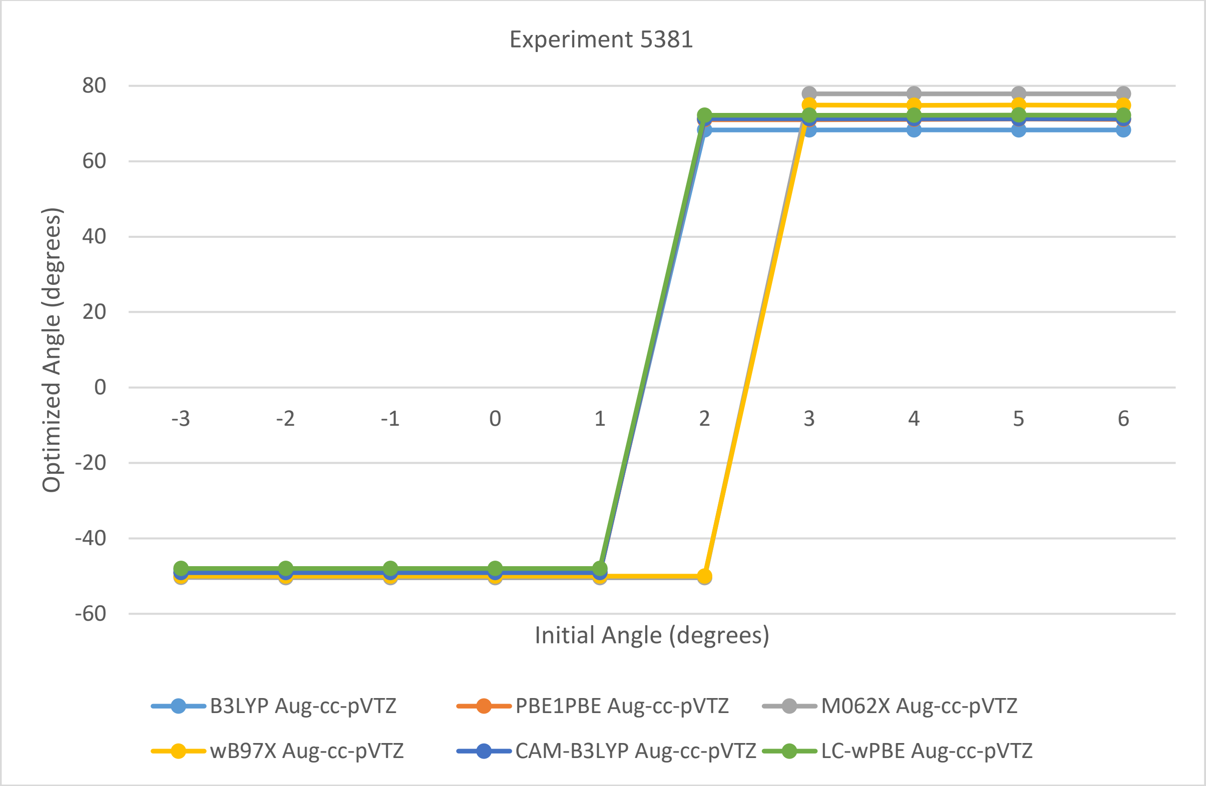


Figure 5.35. Initial versus Geometry Optimized Angle in HF Hydrogen Bonded to N-methylformamide O, Negative Angles for F Proximal to N

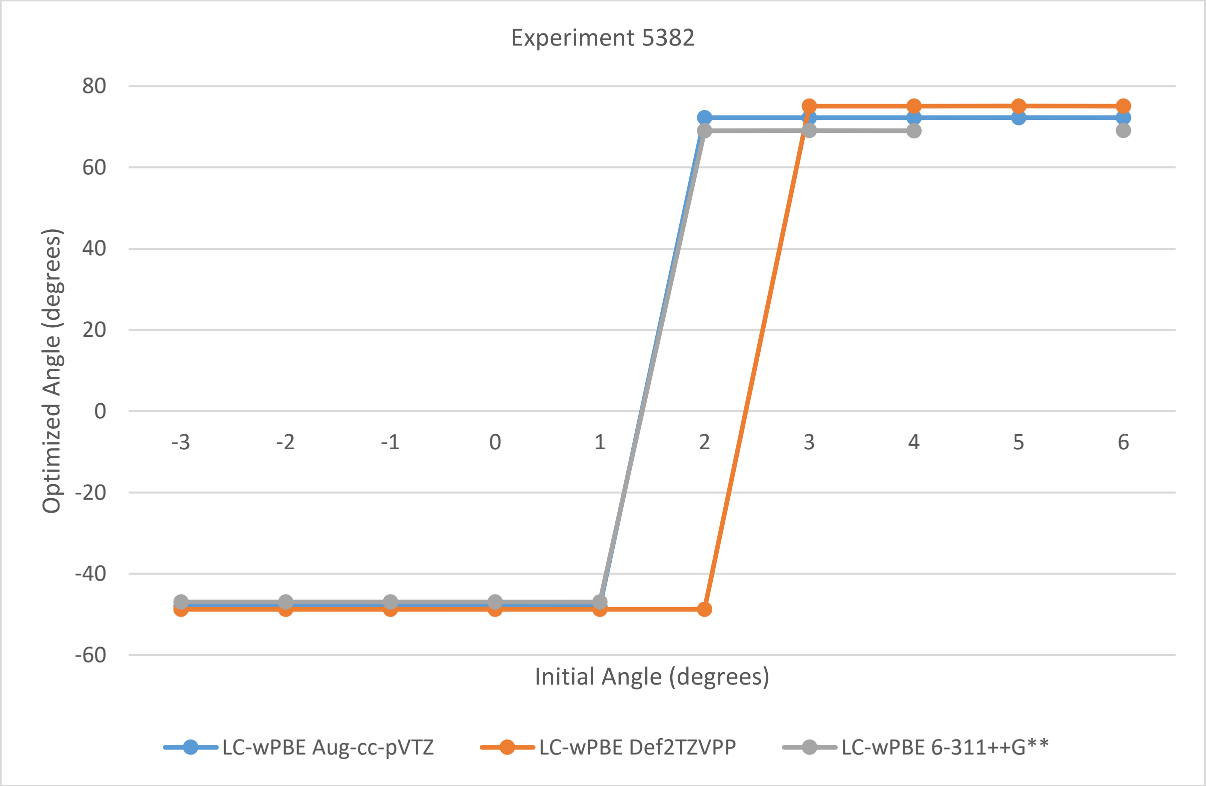


Figure 5.36. Initial versus Geometry Optimized C-O-F Angle with HF Hydrogen Bonded to N-methylformamide O, Negative Angles for F Proximal to N

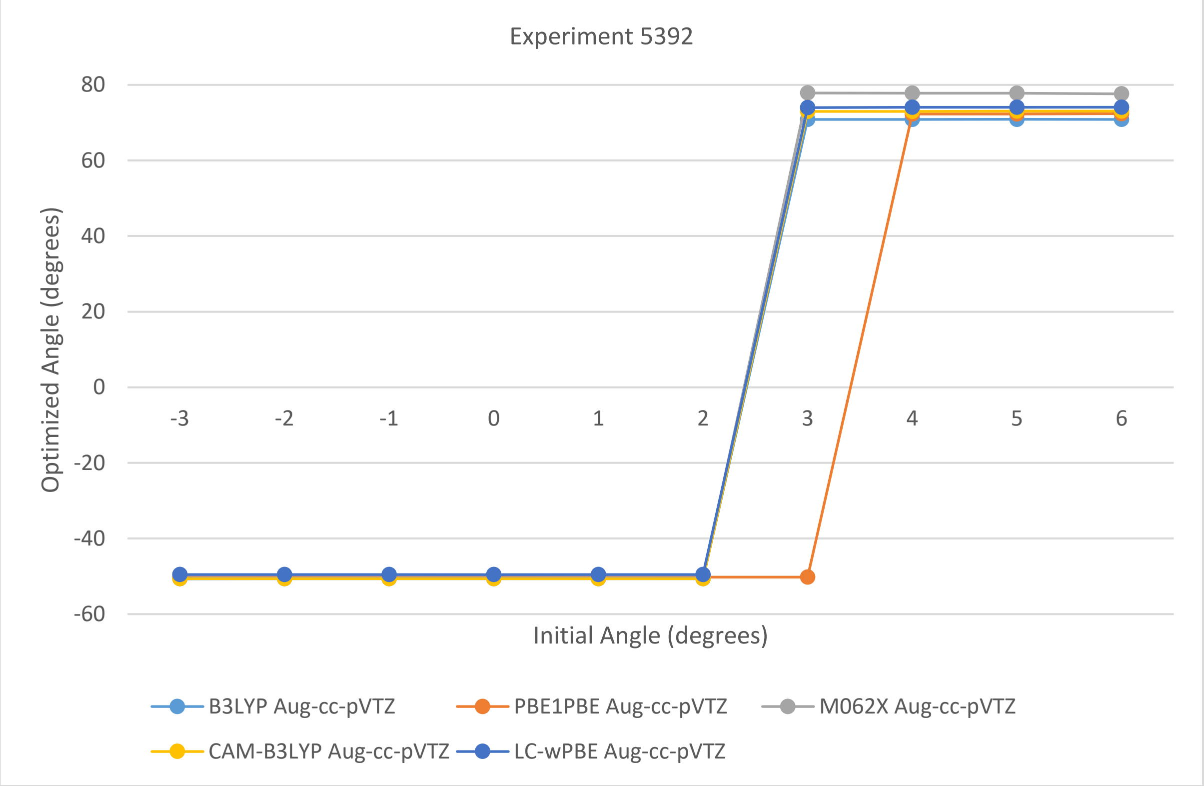


Figure 5.37. Initial versus Optimized C-O-F Angle with HF Hydrogen Bonded to N-methylformamide O in Amide Plane, Negative Angle for F Proximal to N, All Methods with D3 Dispersion Correction

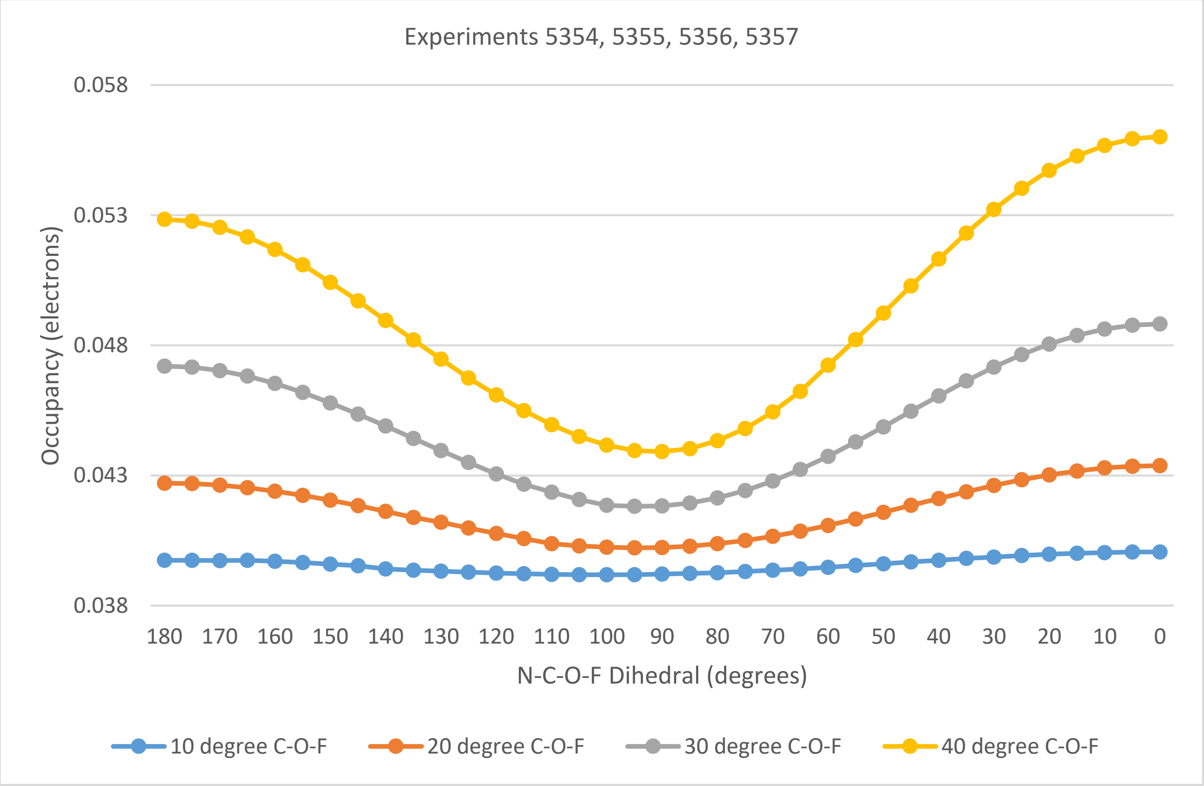


Figure 5.38. H-F* NBO Occupancy HF Hydrogen Bonded to N-methylformamide O at Constant C-O-F Angle with F Rotated About C-O Axis at SCS-MP2/aug-cc-pVTZ

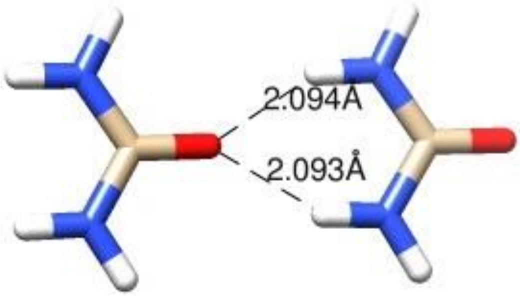


Figure 5.39. Urea Hydrogen Bonded with Urea at SCS-MP2/aug-cc-pVTZ

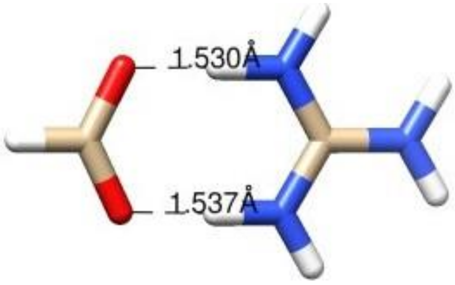


Figure 5.40. Formate Hydrogen Bonded with Guanidinium at SCS-MP2/aug-cc-pVTZ

Figure 5.41. Carbamate Hydrogen Bonded with Guanidinium at SCS-MP2/aug-cc-pVTZ

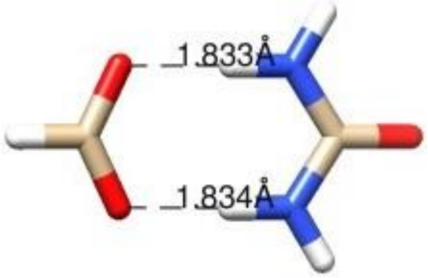


Figure 5.42. Formate Hydrogen Bonded with Urea at SCS-MP2/aug-cc-pVTZ

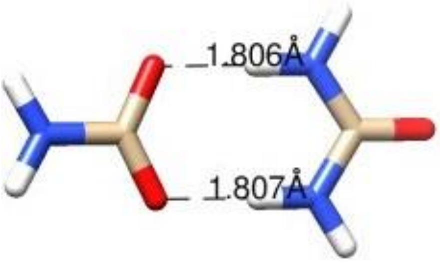


Figure 5.43. Carbamate Hydrogen Bonded with Urea at SCS-MP2/aug-cc-pVTZ

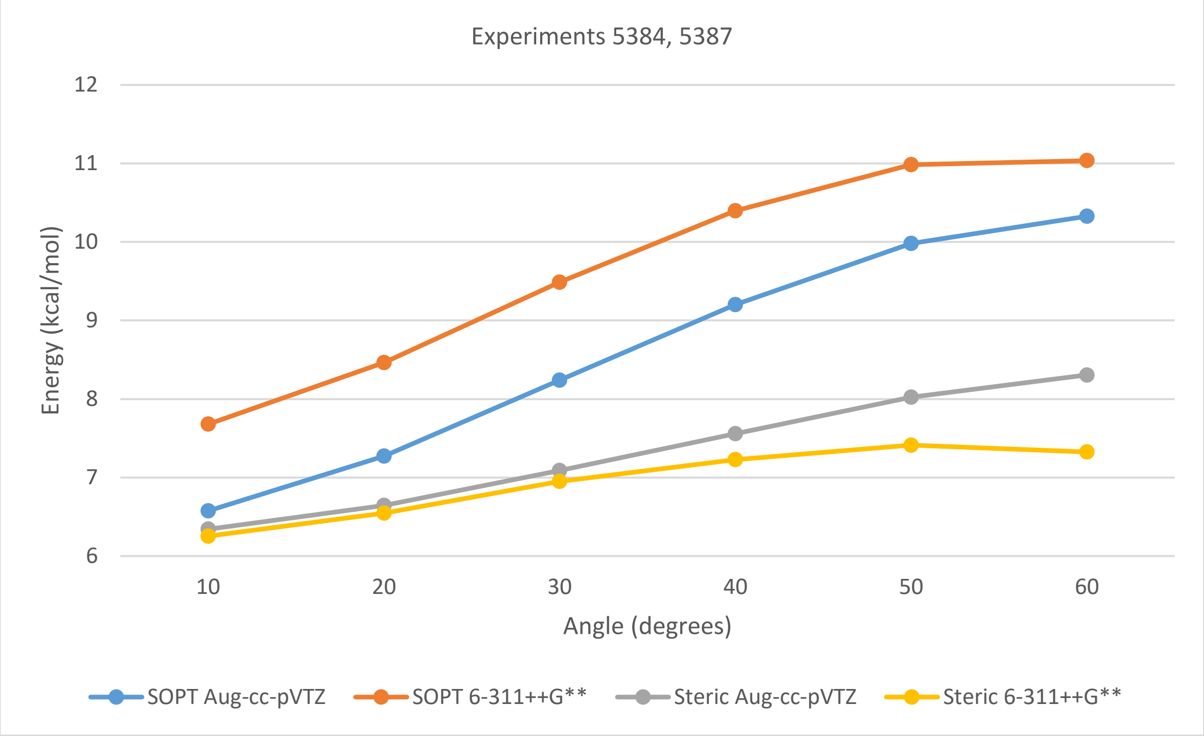


Figure 5.44. Donor-Acceptor SOPT and Steric Exchange Energies for Interactions Between Hydrogen Bonded N-methylformamides at Given C1-O1-N2 Angle with N1-C1-O1-N2 Dihedral 180 Degrees at LC-wPBE(w=0.4)

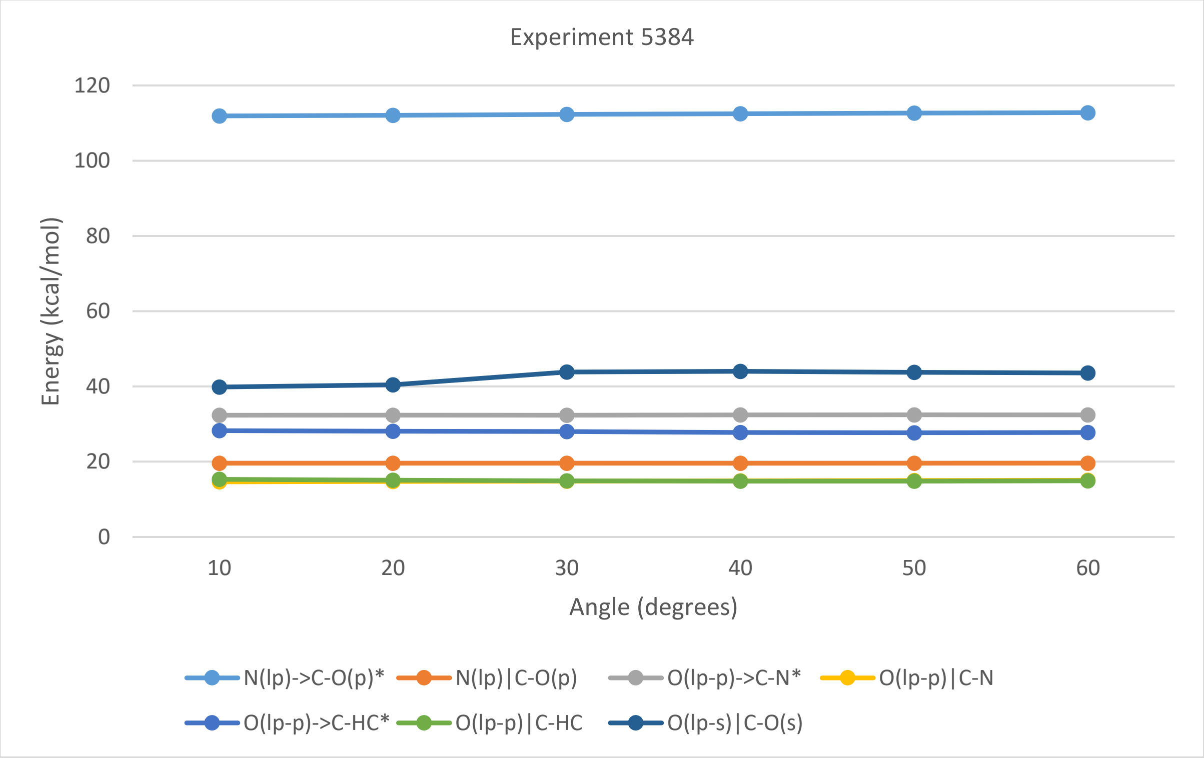


Figure 5.45. Selected Donor-Acceptor SOPT and Steric Exchange Energies for Amide 1 having O Involved in Hydrogen Bonding Between Pair of N-methylformamides at Given C-O-N Angle in Common Amide Plane at LC-wPBE(w=0.4)/aug-cc-pVTZ

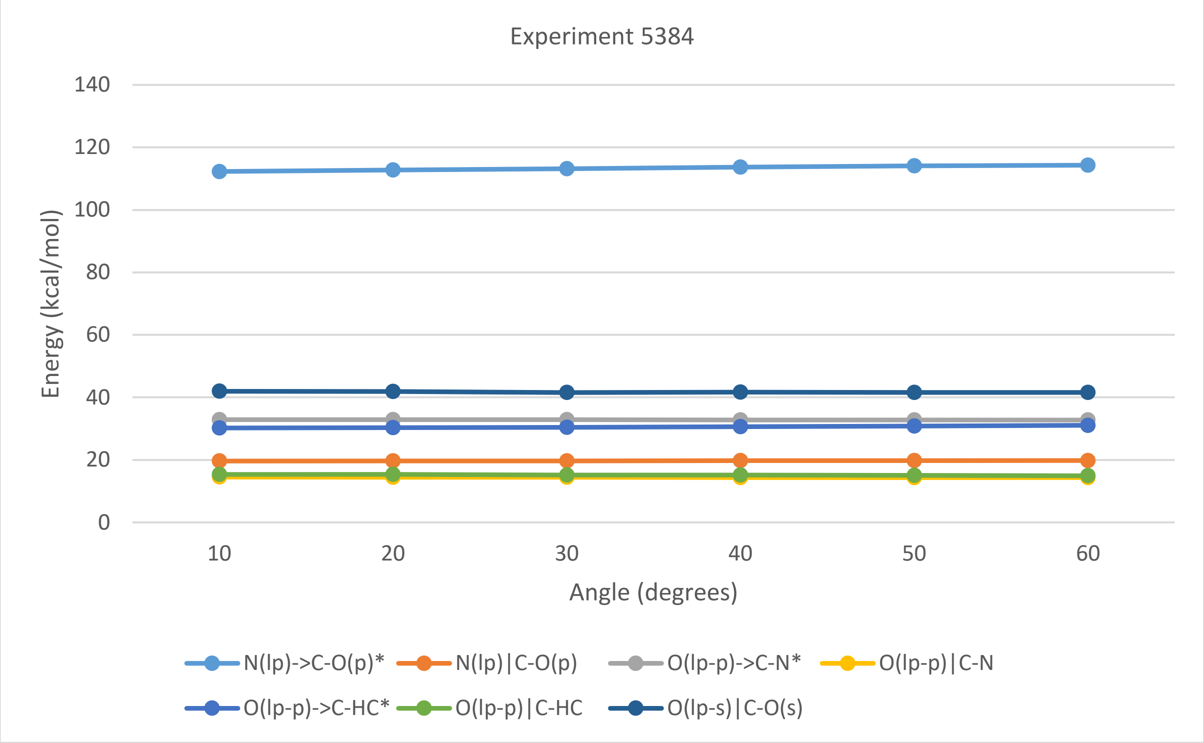


Figure 5.46. Selected Donor-Acceptor SOPT and Steric Exchange Energies in Amide 2 having H-N involved in Hydrogen Bonded Pair of N-methylformamides at Given C-O-N Angle in Common Amide Plane at LC-wPBE(w=0.4)/aug-cc-pVTZ

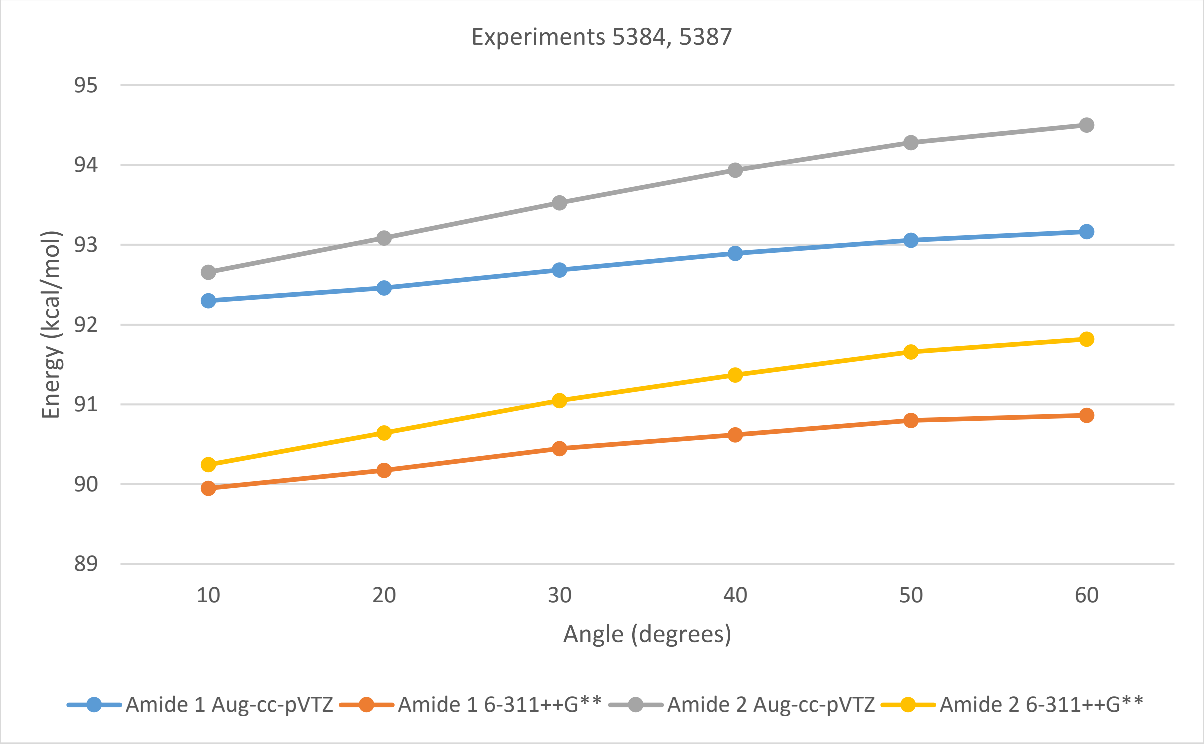


Figure 5.47. Amide Resonance Donor-Acceptor SOPT Minus Steric Exchange Energy for Hydrogen Bonded N-methylformamide 1 (C-O) and 2 (H-N) at Given C1-O1-N2 Angle and N1-C1-O1-N2 180 Degrees at LC-wPBE(w=0.4)/aug-cc-pVTZ

Table 5.6. Selected SOPT and Steric Exchange Energies for Single N-methylformamide

Method	Basis	nlcopa	nlcop	olpcna	olpcn	olpchca	olpchc	olscos
B3LYP	Aug-cc-pVTZ	67.81	16.78	25.91	13.22	22.7	13.56	42.86
PBE1PBE	Aug-cc-pVTZ	71.93	17.36	27.41	13.68	24.13	14.2	43.58
CAM-B3LYP	Aug-cc-pVTZ	87.58	18.35	30.3	14.2	27.15	14.64	42.99
LC-wPBE	Aug-cc-pVTZ	102.94	19.43	34.37	14.68	30.62	15.59	42.54
B3LYP	6-311++G**	65.43	15.84	24.87	13.48	22.41	12.15	40.92
PBE1PBE	6-311++G**	69.39	16.32	26.15	13.88	23.74	12.66	41.74
CAM-B3LYP	6-311++G**	84.76	17.34	29.12	14.4	26.65	13.28	40.12
LC-wPBE	6-311++G**	99.55	18.3	32.94	14.75	29.89	14.08	39.43

- nlcopa: N(lp)->C-O(p)* SOPT kcal/mol
- nlcop: N(lp)|C-O(p) Steric kcal/mol
- olpcna: O(lp-p)->C-N* SOPT kcal/mol
- olpcn: O(lp-p)->C-N Steric kcal/mol
- olpchca: O(lp-p)->C-HC* SOPT kcal/mol
- olpchc: O(lp-p)|C-HC Steric kcal/mol
- olscos: O(lp-s)|C-O(s) Steric kcal/mol

Table 5.7. Distribution of Hydrogen Bond C-O-N Angles Between Backbone Amides

Series	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
banc-banc	1196	3660	7555	11180	6400	1866	654	235	101	23	3		2	1			
banc-bac	66	204	322	431	302	193	98	65	31	18	11	7	3	1	1		
bac-banc	60	194	383	439	257	138	52	36	22	19	2	3	2	1	1		
bac-bac	76	228	396	458	352	185	141	64	32	14	6	5	3	3	5		1

- Banc: Backbone amide that has neither N-terminal nor C-terminal residue as random coil
- Bac: Backbone amide that has either N-terminal or C-terminal residue as random coil
- X-Y: Amide oxygen of X is hydrogen bonded to amide proton of Y

Hydrogen bond lengths are restricted to 2.2 angstroms. The hydrogen bonds reported are restricted N-C-O-N dihedral angles with absolute value less than or equal to 25 degrees or greater than or equal to 155 degrees and N-H-O angles less than or equal to 25 degrees. The numbers on columns of these tables refers to the start of 5 degree groupings of C-O-N angles. Small values intermittently exist for columns to the right of those shown.

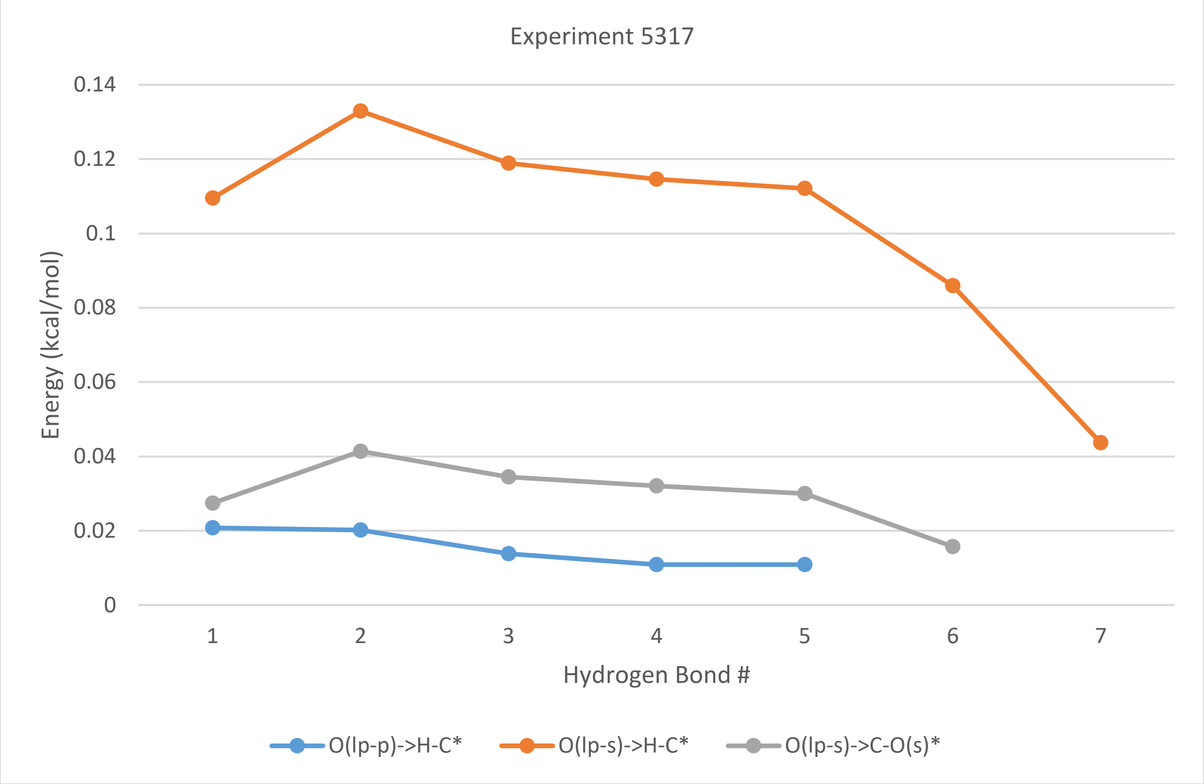


Figure 5.48. Donor-Acceptor SOPT Energy of Interactions Between Successive Units of Chain of 8 Formamides with Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

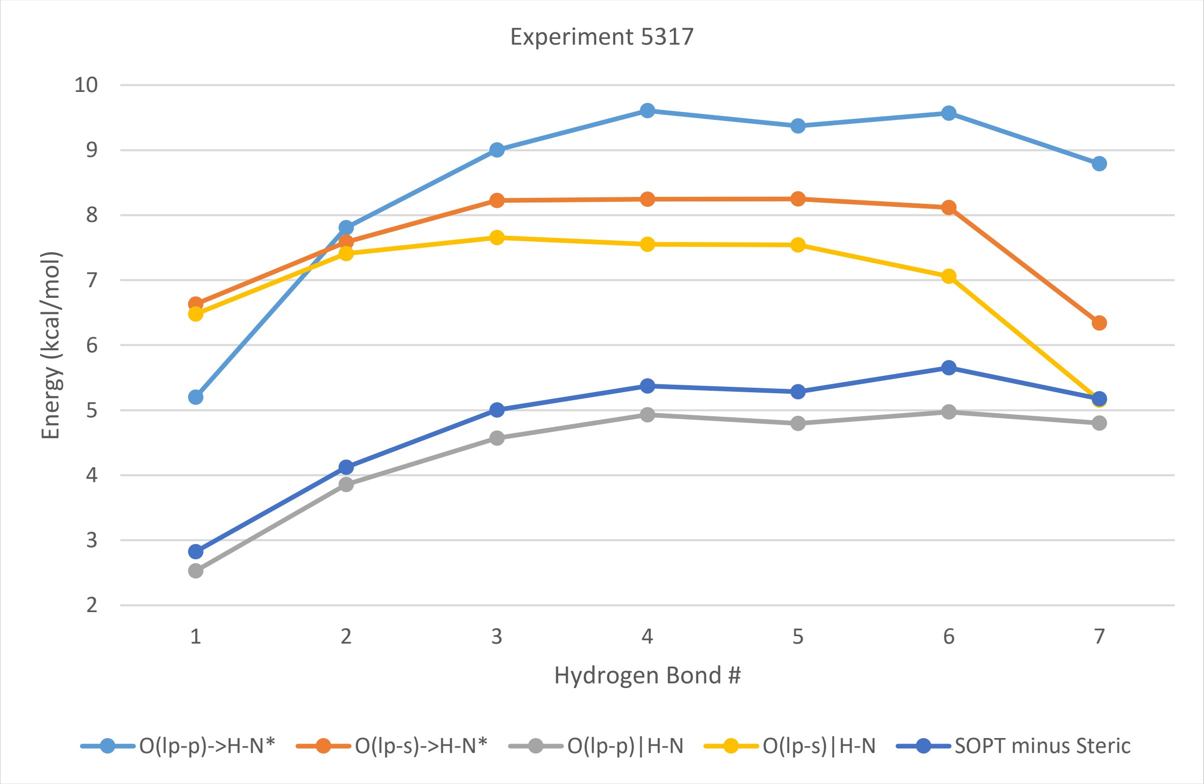


Figure 5.49. Hydrogen Bond Donor-Acceptor SOPT and Steric Exchange Energy in Chain of 8 Formamide Units with Unconstrained Geometry Optimization at LC-wPBE(w=0.4)/6-311++G**

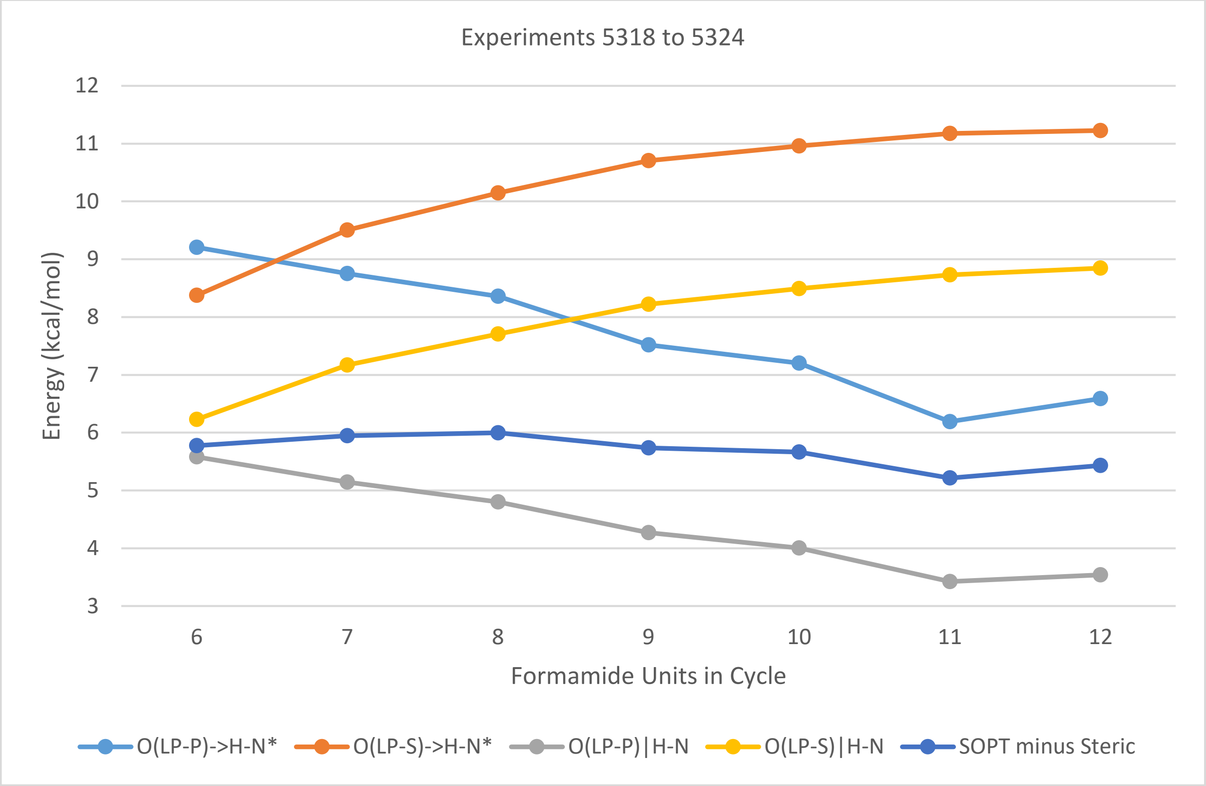


Figure 5.50. Donor-Acceptor SOPT and Steric Exchange Energy in Hydrogen Bonds of Planar Cycles of Formamide Optimized at LC-wPBE(w=0.4)/6-311++G**

5.11 Appendix 2. Polyvaline Parallel Beta Sheet

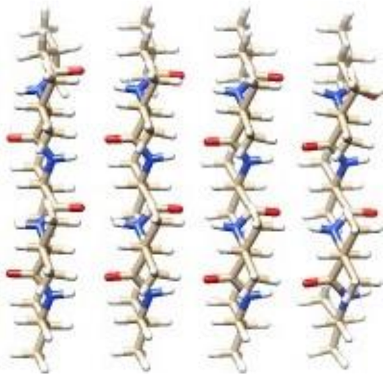


Figure 5.51. Polyvaline Parallel Beta Sheet. Experiment 997.

Table 5.8. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 997

ChainId	s1	s2	s3	s4
1	47.84	48.63	49.76	75.07
2	113.87	117.74	94.94	113.08
3	62.81	100.44	105.27	106.41
4	112.34	118.61	110.07	105.92

Table 5.9. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 997

ChainId	s1	s2	s3	s4
1	20.42	20.94	20.54	6.4
2	0	1.09	5.63	0.79
3	15.41	4.09	3.14	0.22
4	0.45	1.4	2.84	1.6

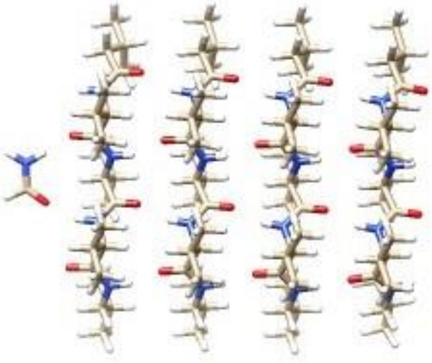


Figure 5.52. Polyvaline Parallel Beta Sheet Capped by One Formamide. Experiment 986.

Table 5.10. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 986.

ChainId	s1	s2	s3	s4
1	49.11	46.55	51.84	73.82
2	116.82	111.79	108.28	101.46
3	101.43	104.65	100.28	104.43
4	109.7	117.59	110.91	111.87

Table 5.11. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 986.

ChainId	s1	s2	s3	s4
1	19.35	22.37	19.21	6.79
2	2.23	2.3	2.58	2.28
3	5.2	3.45	4.1	0.82
4	0	1.28	2.68	0.7

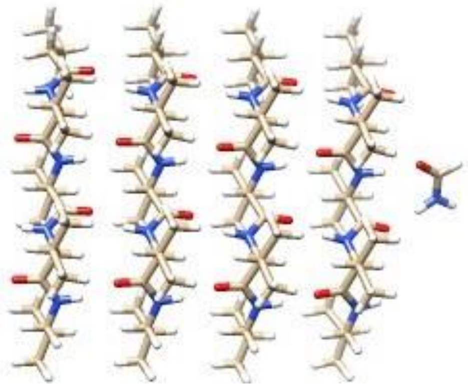


Figure 5.53. Polyvaline Parallel Beta Sheet Capped by One Formamide. Experiment 988.

Table 5.12. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 988

ChainId	s1	s2	s3	s4
1	49.56	47.1	53.57	76.81
2	114.08	111.16	109.41	73.77
3	81.19	104.38	102.44	117.03
4	114.12	118.09	110.49	112.22

Table 5.13. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 988

ChainId	s1	s2	s3	s4
1	19.55	21.89	17.87	6.13
2	0	2.25	2.35	11.21
3	8.75	3.23	3.55	0.5
4	0	1.39	2.74	0.56

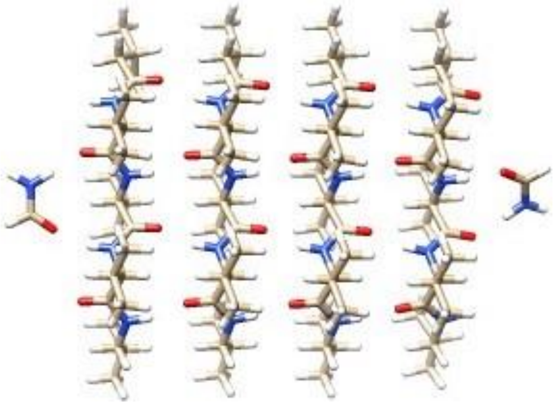


Figure 5.54. Polyvaline Parallel Beta Sheet with Two Capping Formamides. Experiment 998.

Table 5.14. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides in Experiment 998.

ChainId	s1	s2	s3	s4
1	49.25	47.05	53.56	76.93
2	116.71	112.06	109.69	74.02
3	101.68	105.05	102.35	117.23
4	109.78	117.62	110.61	112.22

Table 5.15. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides in Experiment 998

ChainId	s1	s2	s3	s4
1	19.23	21.9	17.93	6.14
2	2.26	2.26	2.32	11.14
3	5.14	3.33	3.65	0.5
4	0	1.25	2.65	0.55

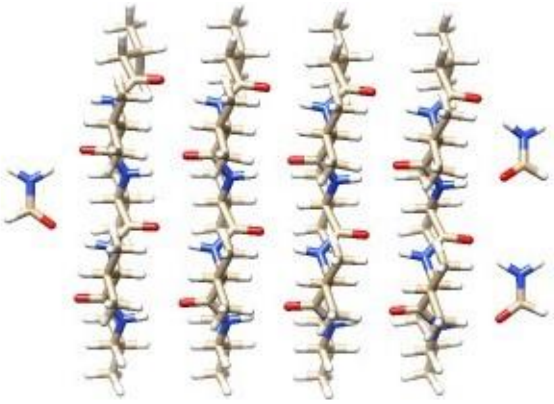


Figure 5.55. Polyvaline Parallel Beta Sheet with Three Capping Formamides. Experiment 991.

Table 5.16. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides in Experiment 991.

ChainId	s1	s2	s3	s4
1	48.91	46.64	48.36	83.19
2	118.05	111.9	105.3	91.02
3	103.41	105.12	104.94	84.94
4	110.16	119.83	120.8	106.96

Table 5.17. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides in Experiment 991.

ChainId	s1	s2	s3	s4
1	19.5	22.49	22.51	8.93
2	2.02	2.62	3.68	7.4
3	4.79	3.5	3.37	8.13
4	0	0.92	1.16	4.13

Table 5.18. Polyvaline Parallel Beta Sheet Totals

exp	pact	sact	cts	ctp
986	1520.52	95.33	105.38	23.21
998	1515.8	100.25	106.91	23.32
991	1509.52	115.16	108.22	25.04
988	1495.4	101.97	106.87	23.16
997	1482.82	104.97	105.69	23.41

- exp: experiment number
- pact: total primary backbone amide charge transfer N(lp)->C-O(pi)* kcal/mol
- sact: total secondary backbone amide charge transfer N(lp)->C-O(sigma)* kcal/mol
- cts: total intra beta sheet inter amide HB charge transfer O(lp-s)->H-N* kcal/mol
- ctp: total intra beta sheet inter amide HB charge transfer O(lp-p)->H-N* kcal/mol

5.12 Appendix 3. Polyvaline Antiparallel Beta Sheet

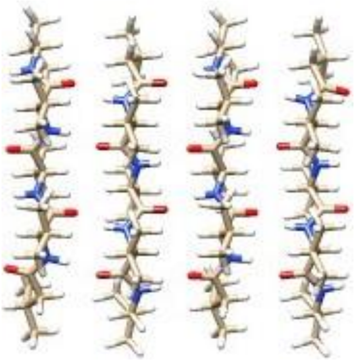


Figure 5.56. Polyvaline Antiparallel Beta Sheet. Experiment 995.

Table 5.19. N(lp)->C-O(p)* SOPT kcal/mol for Experiment 995.

ChainId	s1	s2	s3	s4
1	112.59	57.1	123.58	78.94
2	109.31	125.38	123.8	95.55
3	92.98	123.15	123.07	110.03
4	80.79	123.08	59.46	113.03

Table 5.20. N(lp)->C-O(s)* SOPT kcal/mol for Experiment 995.

ChainId	s1	s2	s3	s4
1	0	17.84	0.13	5.31
2	0.5	0.51	0.37	4.03
3	4.67	0.62	0.96	0.39
4	4.82	0.18	16.62	0

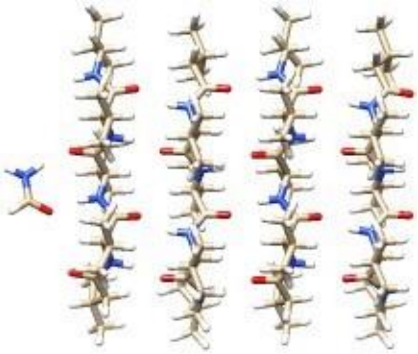
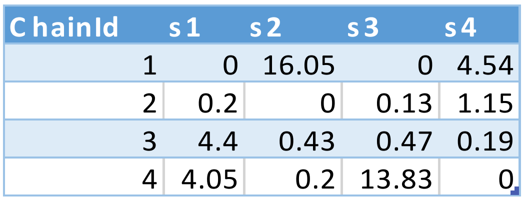


Figure 5.57. Polyvaline Antiparallel Beta Sheet with One Capping Formamide. Experiment 989.

Table 5.21. N(lp)->C-O(p)* SOPT kcal/mol for Experiment 989.

ChainId	s1	s2	s3	s4
1	112.23	60.87	123.99	82.14
2	120.32	128.32	123.54	108.38
3	97.82	124.81	125.76	110.8
4	84.52	121.66	65.35	113.24

Table 5.22. N(lp)->C-O(s)* SOPT kcal/mol for Experiment 989



ChainId	s1	s2	s3	s4
1	0	16.05	0	4.54
2	0.2	0	0.13	1.15
3	4.4	0.43	0.47	0.19
4	4.05	0.2	13.83	0

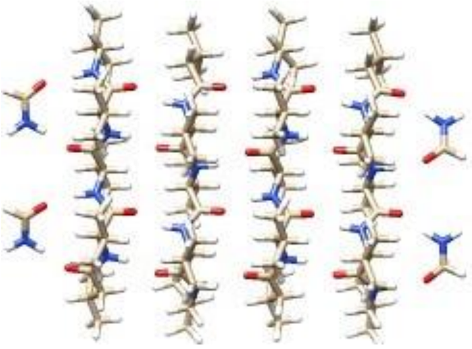
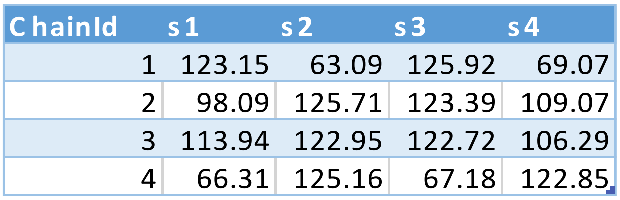


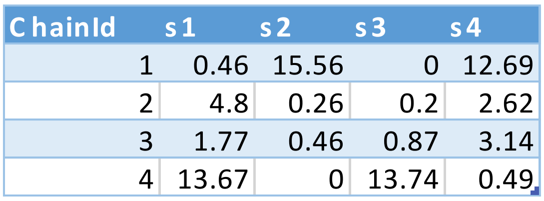
Figure 5.58. Polyvaline Antiparallel Beta Sheet with Four Capping Formamides. Experiment 990.

Table 5.23. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 990.



ChainId	s1	s2	s3	s4
1	123.15	63.09	125.92	69.07
2	98.09	125.71	123.39	109.07
3	113.94	122.95	122.72	106.29
4	66.31	125.16	67.18	122.85

Table 5.24. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 990.



ChainId	s1	s2	s3	s4
1	0.46	15.56	0	12.69
2	4.8	0.26	0.2	2.62
3	1.77	0.46	0.87	3.14
4	13.67	0	13.74	0.49

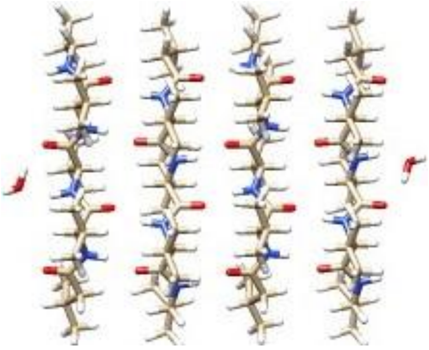


Figure 5.59. Polyvaline Antiparallel Beta Sheet with Two Capping Waters. Experiment 994.

Table 5.25. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 994.

ChainId	s1	s2	s3	s4
1	110.03	64.83	120.29	85.32
2	105.91	120.88	114.94	87.21
3	81.91	114.63	114.03	107.39
4	85.25	120.18	72.08	111.67

Table 5.26. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 994.

ChainId	s1	s2	s3	s4
1	0.41	14.2	0.53	3.42
2	3.26	1.61	2.43	7.89
3	9.33	2.56	2.91	3
4	3.47	0.59	11.41	0.17

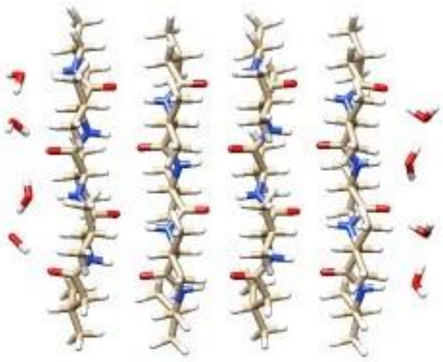


Figure 5.60. Polyvaline Antiparallel Beta Sheet with Eight Capping Waters. Experiment 996.

Table 5.27. N(lp)->C-O(p)* SOPT kcal/mol for Backbone Amides for Experiment 996.

ChainId	s1	s2	s3	s4
1	125.37	66.77	124.07	65.25
2	91.69	119.48	123.81	103.46
3	103.8	123.92	118.36	107.95
4	72.62	123.9	71.69	113.72

Table 5.28. N(lp)->C-O(s)* SOPT kcal/mol for Backbone Amides for Experiment 996.

ChainId	s1	s2	s3	s4
1	0.37	14.03	0.32	14.28
2	9.63	1.62	0	4.46
3	3.94	0	1.83	5.04
4	11.34	0.38	12.03	2.46

Table 5.29.Polyvaline Antiparallel Beta Sheet Totals

exp	pact	sact	cts	ctp
989	1703.74	45.63	100.2	46.71
990	1684.9	70.74	106.1	49.95
996	1655.88	81.74	105.5	49.62
995	1651.84	56.95	98.77	47.32
994	1616.54	67.21	104.3	44.67

- exp: experiment number
- pact: total primary backbone amide charge transfer N(lp)->C-O(pi)* kcal/mol
- sact: total secondary backbone amide charge transfer N(lp)->C-O(sigma)* kcal/mol
- cts: total intra beta sheet inter amide HB charge transfer O(lp-s)->H-N* kcal/mol
- ctp: total intra beta sheet inter amide HB charge transfer O(lp-p)->H-N* kcal/mol

6 Amino acid preference against beta sheet through allowing backbone hydration enabled by the presence of cation

6.1 Statement of authorship

Title of Paper	Amino acid preference against beta sheet through allowing backbone hydration enabled by the presence of cation	
Publication Status	☐ Published	☐ Accepted for Publication
	☐ Submitted for Publication	☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details		

Name of Principal Author (Candidate)	John Neville Sharley		
Contribution to the Paper	all		
Overall percentage (%)	100		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	2016-09-26

6.2 Author contact

john.sharley@pobox.com

6.3 Abstract

It is known that steric blocking by peptide sidechains of hydrogen bonding, HB, between water and peptide groups, PGs, in beta sheets accords with an amino acid intrinsic beta sheet preference [15]. The present observations with Quantum Molecular Dynamics, QMD, simulation with Quantum Mechanical, QM, treatment of every water molecule solvating a beta sheet that would be transient in nature suggest that this steric blocking is not applicable in a hydrophobic region unless a cation is present, so that the amino acid beta sheet preference due to this steric blocking is only effective in the presence of a cation. We observed backbone hydration in a polyalanine and to a lesser extent polyvaline alpha helix without a cation being present, but a cation could increase the strength of these HBs.

Parallel beta sheets have a greater tendency than antiparallel beta sheets of equivalent small size to retain regular structure in solvated QMD, and a 4 strand 4 inter-PG HB chain parallel beta sheet was used. Stability was reinforced by one surface being polyvaline, which buttressed the opposite surface which was used for experimentation. A single Ca^{2+} ion was used for investigation of individual binding events rather than bulk properties. No direct binding between Ca^{2+} and the PG oxygen was observed in these simulations, but perhaps it occurs at longer time scales as the transient beta sheet unfolds.

When linear scaling QMD methods that are accurate [214, 215] for peptide resonance, Resonance-Assisted Hydrogen Bonding [4, 5], RAHB, and the properties of water become available, more extensive experiments having multiple ions of multiple types could be performed at acceptable computational cost. It important that such investigations be performed on protein secondary structures rather than model amides so that sidechain limitation of backbone hydration and hence intrinsic amino acid propensity is captured.

6.4 Introduction

6.4.1 Alpha helix preferring amino acid residues in a beta sheet

Sidechain steric blocking of backbone hydration is known to correlate with amino-acid strand preference [15]. It is also known that strand preferring amino acids have minimum steric conflict with the backbone in that preferred conformation [179], but this does not provide a mechanism for disfavour of alpha helix preferring residues in beta sheets whereas steric blocking of backbone hydration does. The greatest preference for alpha helices is that of alanine [189], though the backbones of polyalanine alpha helices are hydrated, with alpha helical preference not correlating with sidechain blocking of backbone hydration in alpha helices [15]. We used a beta sheet with a polyalanine surface, with the other surface being polyvaline for stabilization of the sheet. Parallel beta sheets are more stable than antiparallel beta sheets when the sheets are small, and a 4 chain 4 inter-PG HB parallel beta sheet was used here. Valine has the largest preference for beta sheets [189], and in parallel beta sheets, valine sidechains of a polyvaline surface can interlock in a regular manner. The strands were acetyl and N-methyl capped rather than left charged, to more closely resemble unbroken chains (Figure 6.1).

Hydration and destabilization of a beta sheet backbone was studied at the level of individual HB events rather than at the level of bulk behaviour which does not necessarily illuminate basic molecular mechanism. The beta sheet studied would be transient if it occurred in nature at all, with its occurrence constituting a transient misfolding of the involved residues. Illustration of the patterns of HB contributing to the correction of this transient misfolding was sought in the following experiments.

6.4.2 Cation interactions with protein backbone oxygen

Algaer and van der Vegt [216] note that the chemical environment of the backbone amide should be considered in studying Hofmeister effects [217]. To this end, we investigated water mediated cation interactions with backbone amides in a beta sheet that would be transient in nature.

Okur *et al*. [218] found that the direct contact and solvent-separated binding for cations and amide oxygen in d-butyramide was extremely weak at biological ion concentrations, and the ordering of affinity is $Ca^{2+} > Mg^{2+} > Li^{+}$. Therefore, we used Ca^{2+} for this study to best observe cation interactions with the backbone.

6.4.3 Quantum molecular dynamics with quantum mechanical treatment of every water molecule

Okur *et al*. [218] state that “simulations that find tight associations between metal cations and the carbonyl oxygen of amides are not consistent with the spectroscopic data”. This motivates the use of fully QM rather than classical simulations.

Our QM investigations of variation in amide resonance due to electrostatic field with component parallel or antiparallel to the amide C-N bond and due to RAHB demonstrate that the variation in amide resonance and hence charge distribution in the amide group is considerable. Classical calculation has no account of variation of amide resonance. Since the backbone amides studied were already participating in secondary structure RAHB chains and further we are introducing ions and hence electrostatic field into the vicinity of the backbone group, classical methods are unsuitable. Against this indispensable advantage of QM methods, there is our own finding that established Density Functional Theory, DFT, methods are undesirable for calculation of amide resonance when the amide carbonyl is engaged in torsional hyperconjugative interactions that in particular occurs in parallel and antiparallel beta sheets (Chapter 3). However, pending the availability of methods that accurately model amide resonance and scale to QM-handled explicitly solvated beta sheets, we are obliged to use these DFT methods.

The computational expense of these simulations with available methods prohibits a comprehensive survey. Such a survey can be undertaken upon the availability of QMD methods that scale linearly in runtime and memory use with both protein atom count and solvent atom count and provide a good account of amide resonance, RAHB and the properties of water. Fragment methods [219] need to be qualified as accurately capturing RAHB in secondary structure backbone amide chains. If fragment boundaries bisect peptide bonds, particular care needs to be taken that amide resonance varies accurately with the electrostatic field component that is parallel or antiparallel to the amide C-N bond (Chapter 4) as well as with RAHB-induced variations. Surveys involving greater depth of solvent and longer simulated duration that could be undertaken with more accurate and better scaling methods are described in Section 6.9 (Future work).

6.5 Methods

The quantum chemistry package TeraChem 1.5K [91-94] was used in these large atom count QMD experiments. This version of TeraChem is cubic scaling at the atom count used, proceeding further up the cubic curve than other programs by efficient utilization of Graphical Processor Units [220]. Explicit solvent is used, with quantum mechanical treatment of all solvent molecules.

While the protein is fully solvated, the solvation shell is thin in classical simulation terms, being 7 angstroms, but is similar to the 5 angstrom minimum solvation shell recommended for best maintenance of the HOMO-LUMO gap in DFT calculations of protein electronic structure [221]. In the molecularly crowded intra-cellular environment, it will often happen that solvent between proteins is sparse, though there is a water/vacuum interface in the present simulations.

Full Hartree-Fock exchange [222] at long range is recommended to minimize a charge transfer inaccuracy arising from decreasing HOMO-LUMO gap of edge waters with oxygen not participating in HB [223]. We used LC-wPBE [88], denoted wpbe in TeraChem, for its tractability of convergence. Isborn *et al*. [223] give w=0.26 as Koopman-optimal [224] for this method, but they used 0.2 and we used 0.4. D3 empirical correction [94] is not available for this method.

No geometry constraints were applied other than protein atoms fixed for early steps in some experiments while the TIP3P solvent shell added by the Chimera [111] Solvate function took on angles, bond lengths and density suitable for QMD. The course of the QMD simulation was unguided by Ab Initio Steered Molecular Dynamics [225] or other means. Spherical boundary conditions 'mdbc spherical', constant density 'md_density 1.0' and Langevin dynamics 'thermostat langevin' were set. The Langevin temperature damping time 'lnvtime' [226] was varied during equilibration as given in the notes for each experiment. The indicator of equilibration used is approximate convergence to the target temperature, which was set at 310.15 Kelvin.

Focussing on cations has a computational benefit in that there is less need for diffuse functions [227], in contrast to anions which have more diffuse electron distribution. Diffuse functions are more challenging for convergence of quantum chemistry methods, and computational speed is diminished by their use. Also, in a QM-handled explicitly solvated simulation, electrons tend to be less diffuse than in the gas phase, reducing the need for diffuse functions.

We are presently obliged to use basis sets of a size and type we demonstrated to contribute to inaccuracy in modelling amide resonance and hence secondary structure RAHB (Chapter 3).

6.6 Results

6.6.1 Preparation

Chimera's Solvate function was used to add a 7.0 angstrom shell of TIP3BOX waters to a 4 strand 4 inter-PG HB chain parallel beta sheet with valines on one side to stabilize the sheet and alanines on the other side for the experimental surface (Figure 6.1) which had been geometry optimized at wpbe(rc_w=0.4)/6-31g** before solvation. 300 steps (frames) of QMD was applied with the coordinates of heavy atoms fixed then 200 steps with all atoms free, with all steps at a Langevin thermostat damping time parameter value lnvtime=10 fs.

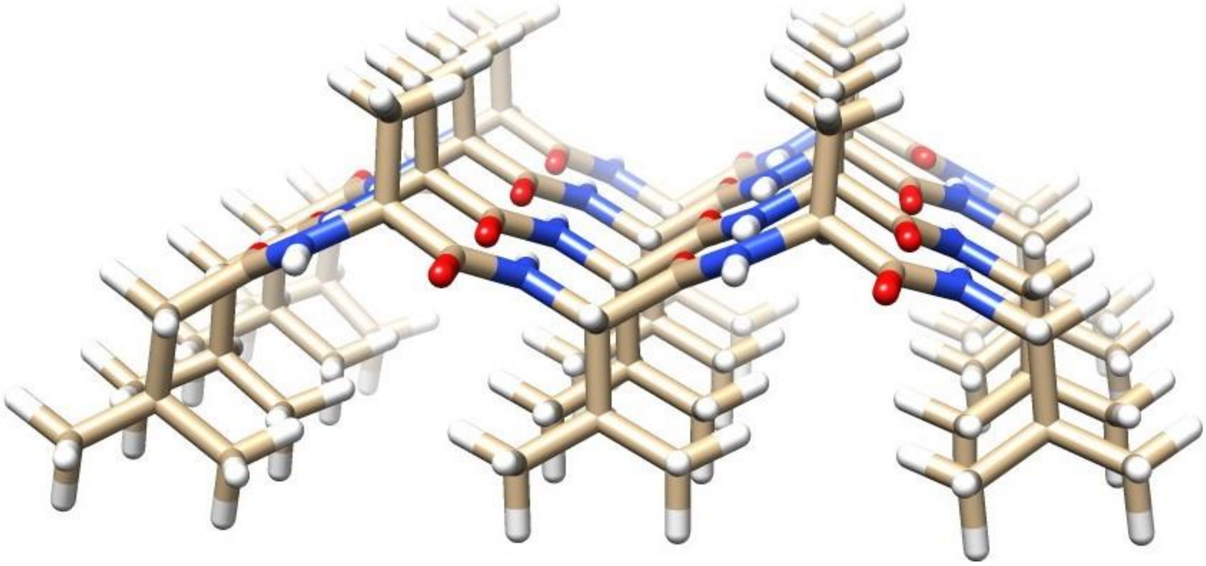


Figure 6.1. 4-Beta Strand 4-HB Chain Parallel Beta Sheet with Ala on Experimental Surface and Val on Other Surface after Unconstrained Optimization at wpbe(rc_w=0.4)/6-31g** Prior to Solvation

6.6.2 Experiment 1302

A water (oxygen serial number 436) with coordinates 3.359 angstroms from an Ala CB on one cross-strand row and 5.271 angstroms from the Ala CB on the other such row on the same strand on the same surface of the beta sheet was replaced with a Ca^{2+} atom at the oxygen coordinates and a new unconstrained MD run (1302) was started from that frame. 150 QMD steps with lnvtime=10 fs then 650 steps with lnvtime=100 fs were applied.

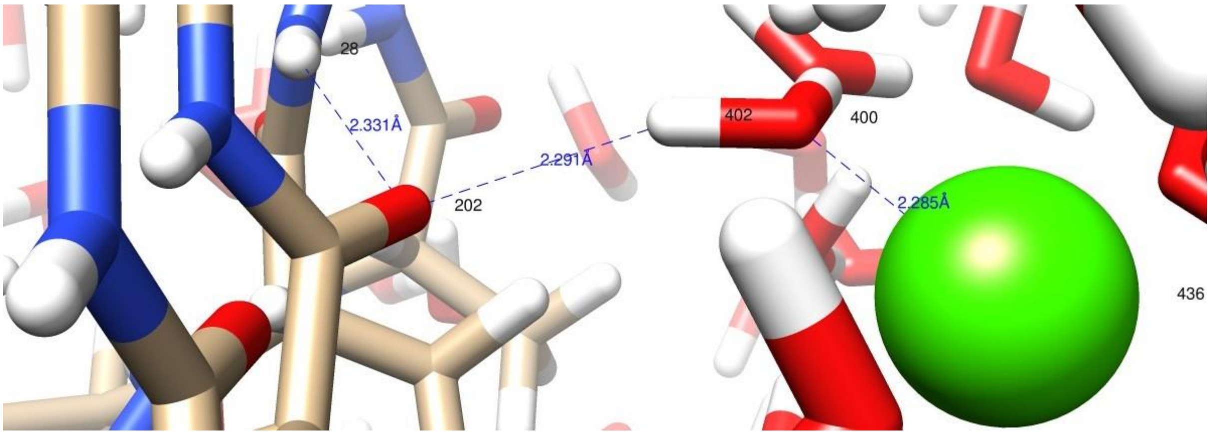


Figure 6.2. Backbone Hydration in Final Frame of Experiment 1302

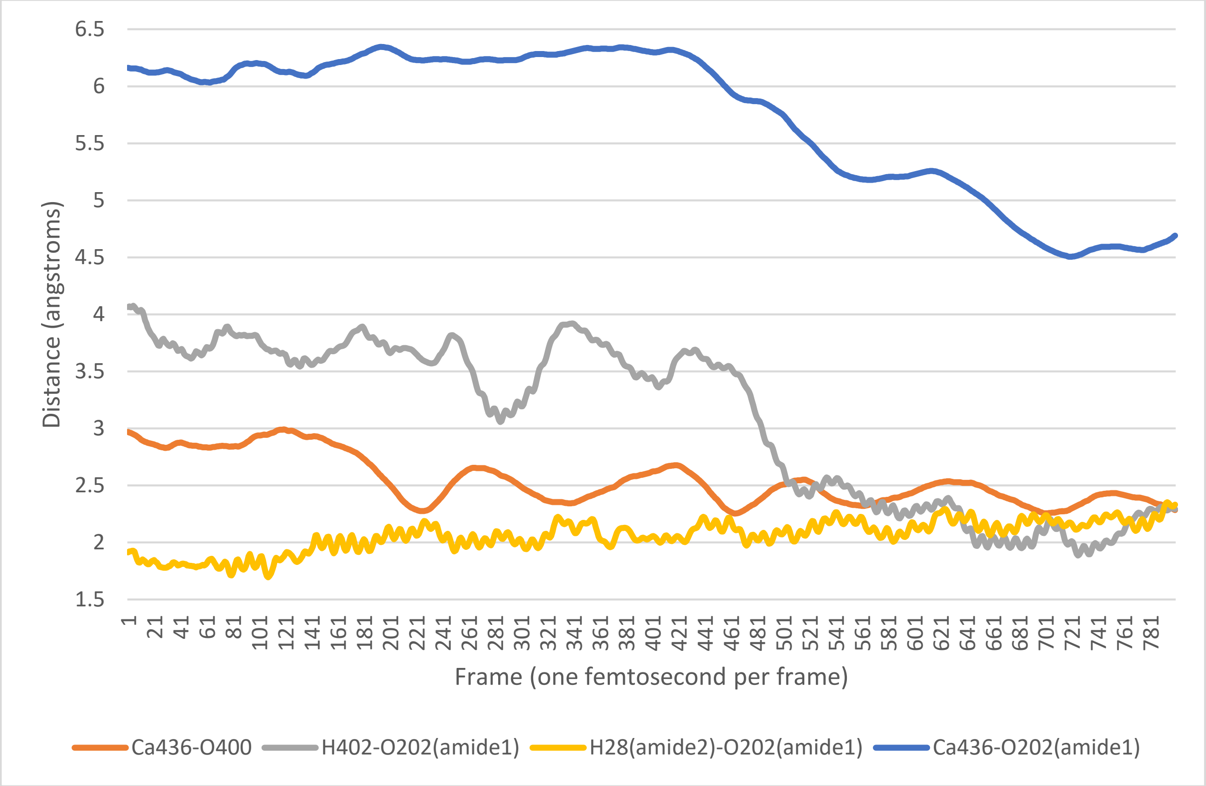


Figure 6.3. Selected Bond Lengths in Experiment 1302. Atom Ids Shown in Figure 6.2

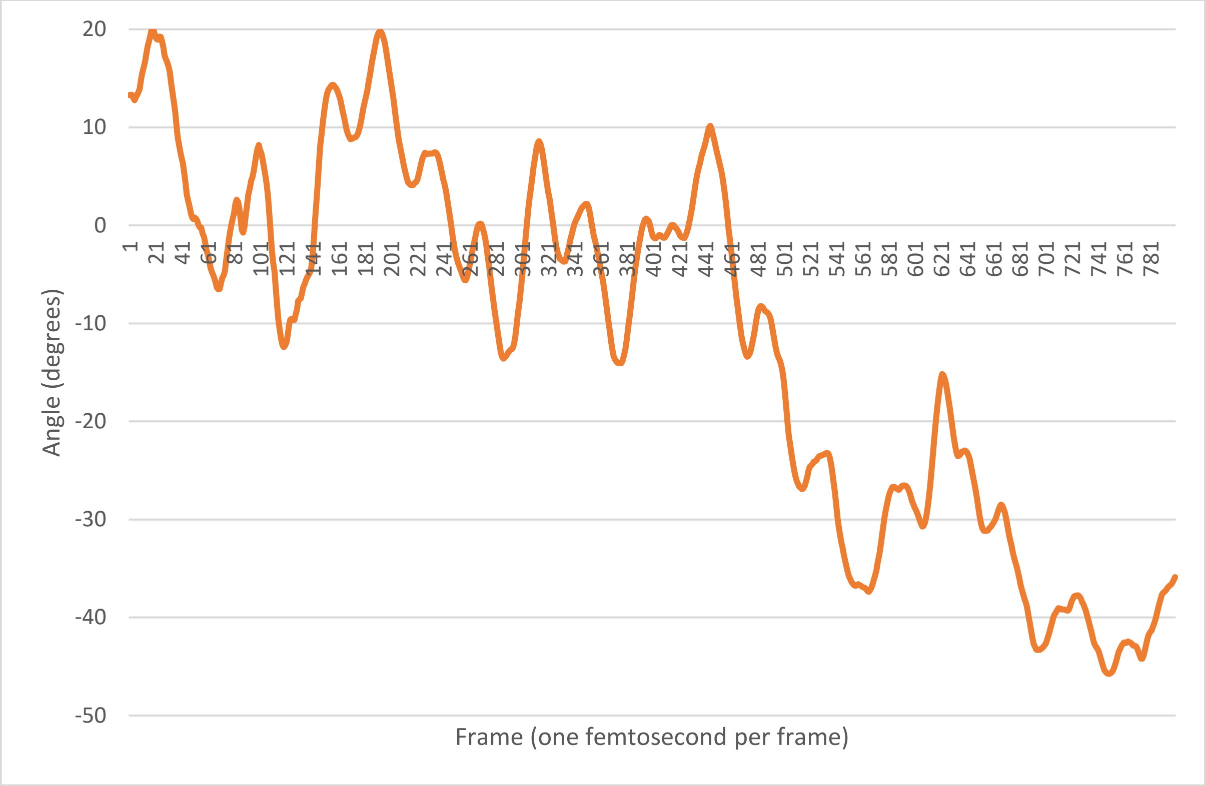


Figure 6.4. Angle of O202-H28 from O202, C200, N201 Normal Minus 90.0 in Experiment 1302

6.6.3 Experiment 1303

Starting with the final frame resulting from the common preparation (1301) and removing a water (oxygen serial number 391) with coordinates 4.915 angstroms from an Ala CB and 5.561 angstroms from another Ala CB on the same cross-strand row and placing a Ca^{2+} atom at the oxygen coordinates, a new unconstrained QMD run (1303) was started. 150 QMD steps with lnvtime=10 fs then 650 steps with lnvtime=100 fs were then applied.

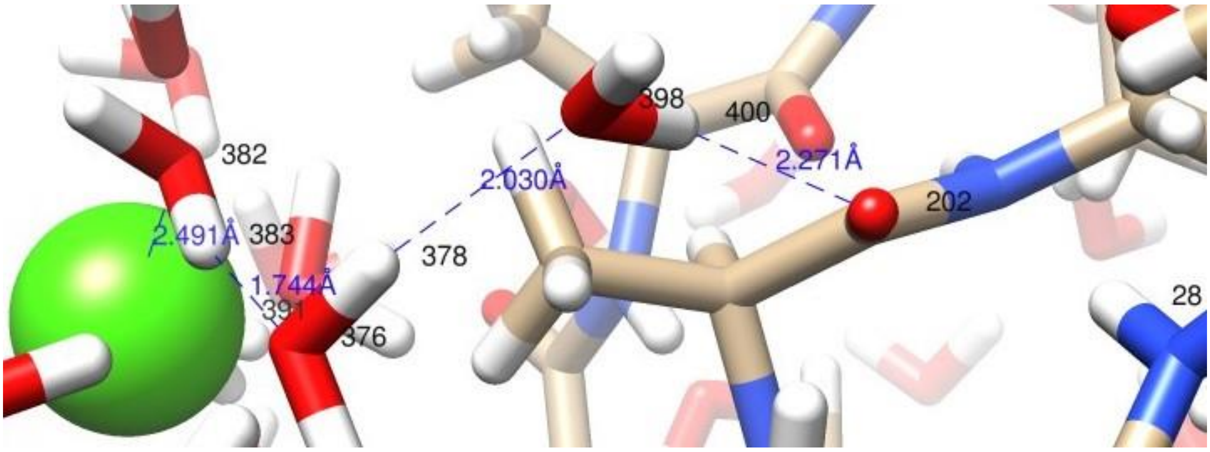


Figure 6.5. Backbone Hydration in Final Frame of Experiment 1303

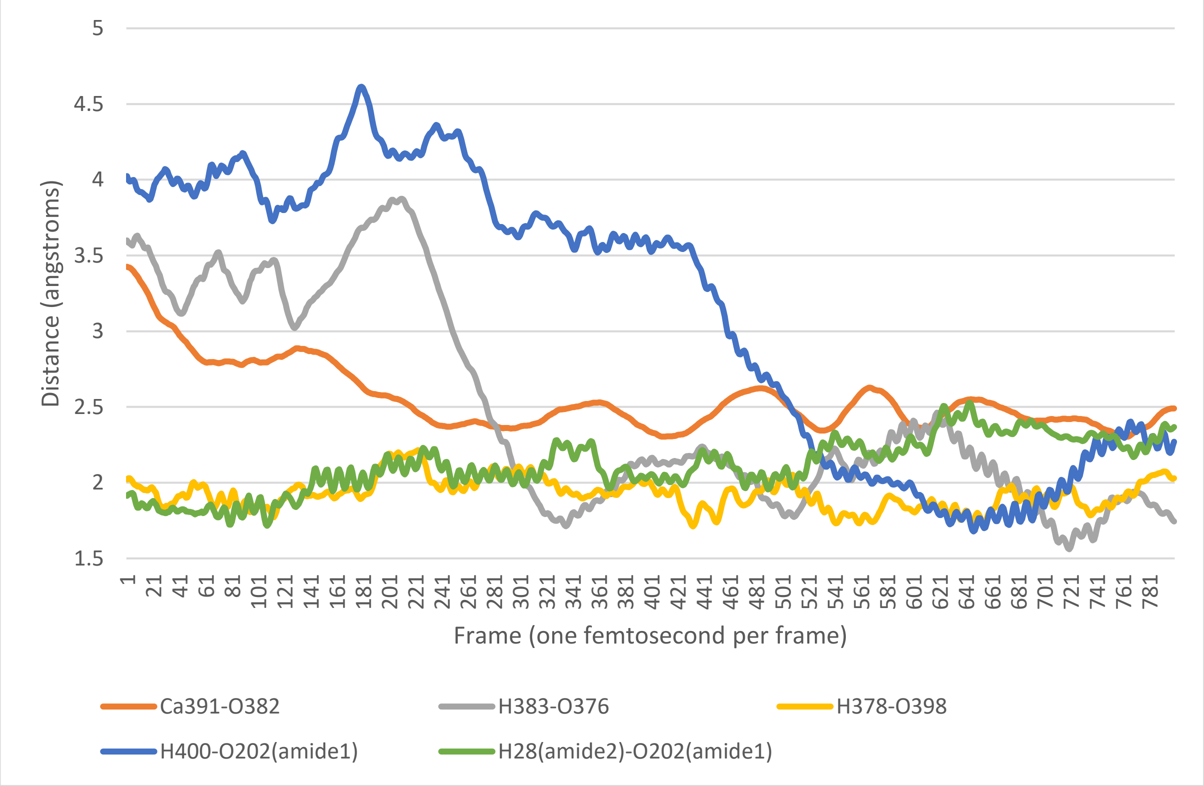


Figure 6.6. Selected Bond Lengths in Experiment 1303. Atom Ids Shown in Figure 6.5

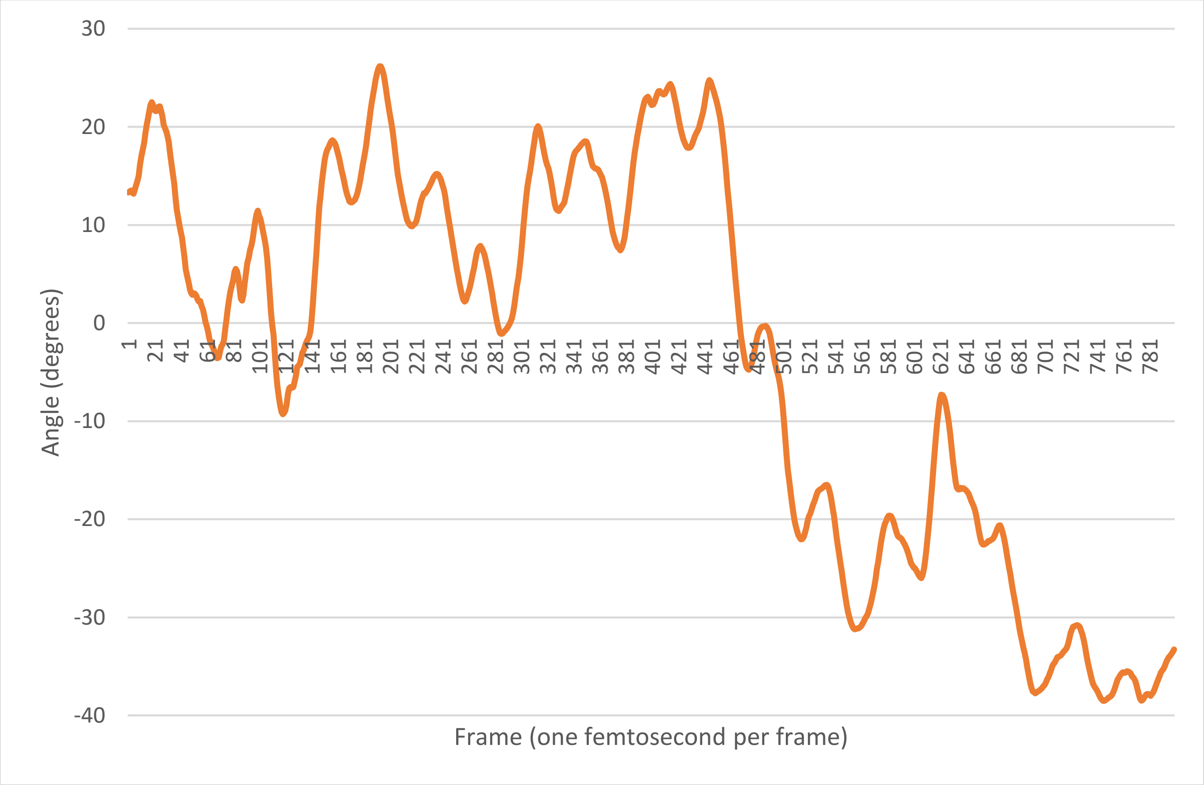


Figure 6.7. Angle of O202-H28 from O202, C200, N201 Normal Minus 90.0 in Experiment 1303

6.7 Discussion

6.7.1 HB networks of water

In both experiment 1302 and 1303, temperature is still equilibrating up to frame 440.

The experimental surface was between the two rows of alanines, since the backbone exposure outside of the alanine rows but still on that surface of the beta sheet was artificially large due to the strands not continuing beyond the final valines on the opposite surface. The two examples above, Figure 6.2 and Figure 6.5, show hydration between the rows of alanines, that is, on the experimental surface. Two examples of hydration outside the rows of alanines are shown in Ap1:Figure 6.8 and Ap2:Figure 6.11.

Figure 6.2 shows a single water connecting Ca^{2+} and an interior PG. Figure 6.5 shows a chain of three water molecules connecting Ca^{2+} and an interior PG. Ap1:Figure 6.8 shows a 3-deep network of waters connecting Ca^{2+} and an interior PG outside the experimental surface. Ap2:Figure 6.11 shows a 1-deep network of waters connecting Ca^{2+} and an interior PG outside the experimental surface. Perhaps connection of Ca^{2+} to interior PGs between the rows of alanines requires a chain rather than a network of waters due to the steric constraints given by the alanine sidechains.

The distances between atoms as they become part of a HB network connecting Ca^{2+} and PG oxygen are shown in Figure 6.3, Figure 6.6, Ap1:Figure 6.9 and Ap2:Figure 6.12.

Figure 6.4, Figure 6.7, Ap1:Figure 6.10 and Ap2:Figure 6.13 show the extent of variation of the angle between the PG normal and the inter-PG HB of its oxygen. At these very large angles from PG co-planarity, the RAHB of the inter-PG HB chain will be reduced as will the stability of the beta sheet. Whether the sheet unfolds or not will depend on the extent to which the geometry of all of the inter-PG HB chains in the sheet is ultimately de-optimized by this backbone hydration event or other such events.

In the presence of a single Ca^{2+} ion, the backbone of the beta sheet is hydrated via the polyalanine surface. This supports a view that a Ca^{2+} can rupture a hydrophobic hydration shell to enable intrinsic amino acid beta sheet preference. It cannot be concluded that water without an ion will never hydrate the backbone of the experimental surface, for the simulated time is short (0.8 picosecond, including temperature equilibration during which these HB networks do not form), but it is apparent that over the timecourse of the simulation, the backbone is more strongly hydrated in the presence of Ca^{2+}.

6.7.2 Subsequent to rupture of a transient beta sheet

If a transient beta sheet has been ruptured by cation-mediated backbone hydration, what happens next? Model amide studies are not revealing of the manner of backbone hydration of a ruptured beta sheet. Will not the HB chains and networks between the cation and the PG oxygen become stronger due to better access by water to the PG oxygen which is no longer in an inter-PG HB chain? This better access may include improved access to the PG oxygen p-type lone pair, increasing HB strength (Chapter 5). It is likely that quite a number of PG oxygens will be solvent-exposed after rupture of a beta sheet. The presence of cation makes peptide-water HB more favourable than inter-peptide HB. By what process are the cations eliminated so that the protein can fold? Perhaps in time the cations will associate with anions and drift away. Perhaps ongoing conformational change due to hydrophobia eventually dislodges these cations from the often hydrophobic residues of a beta sheet. Perhaps a number of solvent exposed PG oxygens in an otherwise hydrophobic surface is a target for molecular chaperone [13] action, whether the PG oxygens are no longer associated with cation, are indirectly or directly bound to cation.

Cations associating with PG oxygen that were not in an inter-PG HB chain would be limiting of folding progress. The prospects for a timely fold would be improved by protecting surface PG oxygen that would be buried in the correctly folded protein. Molecular chaperones might either protect such PG oxygens, or actively restructure the protein.

Perhaps a function of high Ca^{2+} concentration such as in the endoplasmic reticulum, ER, is to test the folding of proteins in that compartment. Misfolded secondary structures of proteins of that compartment become unstable in the presence of Ca^{2+} by the means described and folding chaperones then identify folding work to be done by the state of the protein after failing the test of folding.

6.7.3 Hofmeister effects

Perhaps how well other residues complement a given residue in obstructing the formation of HB networks between ions and PGs is dependent on the concentration of ions and temperature. If so, Hofmeister effects [228-230] must be studied in the presence of this sidechain complementation rather than in model amides. A structural account of this complementation such as can be given by QMD is desirable. Perhaps in aspects of Hofmeister effects, the mechanism that makes misfolded beta sheets more energetically unfavourable is dysregulated by non-physiological conditions. This might extend to sequence-dependent dehydrated sections of alpha helical backbone.

Hydrogen bonded networks connecting cations with PG oxygens are not necessarily diminished in polarization or charge transfer by the presence of anions. A water molecule might be a node of a HB network which includes both cation(s) and anion(s) such that its polarization or charge transfer is increased by interactions of its oxygen indirectly due to cation and also by interaction of one its O-H antibonding orbitals indirectly due to anion. Its free O-H antibonding orbital could then be indirectly associated with increased polarization or charge transfer from PG oxygen lone pairs.

Shi and Wang [231] suggest that cation and anion are in contact at high concentrations rather than being separately solvated. There are two quite opposing views arising from this that might be considered. The first is that this is inhibitory of HB networks connecting cations with PG oxygen. The second is that a water molecule with oxygen within charge transfer distance of the cation and a hydrogen within charge transfer distance of the anion may be yet more polarized than in the presence of cation alone, and the other hydrogen of the water molecule could start a more polarized HB chain to the backbone oxygen. Since there are likely to be multiple such waters at charge transfer distance with both cation and anion, they may collectively start a HB network to backbone oxygen which is more polarized than those from cation without anion. More polarized HB networks could be expected to rupture a hydrophobic hydration shell and access backbone oxygen where this would not occur at lower concentrations.

Another possibility, which assumes separability of ions, for the emergence of Hofmeister effects is that the electrostatic field with component parallel or antiparallel to the PG C-N bonds disturbs PG resonance and secondary structure RAHB chains (Chapter 4) at high ionic concentrations, with the electrostatic field fluctuating with the particular arrangement of ions. A momentary clustering of anions near one surface of a protein and of cations on an opposed surface may be destabilizing of the electrostatic field in the protein between these surfaces and of the PG resonances and hence secondary structure RAHB. Brief electrostatic destabilization of the protein might permit formation of HB networks between ions and PGs.

6.8 Conclusion

Based on observation of the simulated formation of water HB networks connecting Ca^{2+} and PG oxygens interior to a transient beta sheet involving residues that in a correct fold would not form this sheet, we propose that elimination of these transient misfolded sheets in the course of folding is ion-dependent. Alternatively stated, amino acid preference for beta sheet by sidechain blocking of backbone hydration is not effective unless ions are present. Backbone hydration of these transient beta sheets without ions might occur on longer time scales than simulated, but it appears this hydration is faster in the presence of Ca^{2+} and is anticipated to form closer HB to PG oxygen and to hydrate the backbone where water alone does not.

It is anticipated that alpha helix preferring residues may appear in a beta sheet if surrounding residues compensate for the deficiency of sidechain blocking of backbone hydration, referred to here as sidechain complementation. It is desirable that the sidechain structures of complementation be studied, and we expect it will soon be possible to do this with QMD. Non-hydrophobic residues will also vary the formation of HB networks to PG oxygen, and use of similar methods is indicated for such a study.

The proposed role of cations in eliminating transient secondary structures which arise due to misfolding, an early stage of which is computationally demonstrated here with Ca^{2+}, invites consideration of high Ca^{2+} concentrations in the ER as a testing environment for correct sidechain complementation and hence fold. It is suggested that the exposed PG oxygen in an otherwise hydrophobic setting, perhaps in association with a multivalent cation, could mark the hydrophobic patch for action by a folding chaperone.

Limitations due to thin solvent shell, short simulated time span and use of established DFT methods that suffer disturbances in modelling PG resonance and hence RAHB in protein secondary structures (Chapter 3) are acknowledged. QMD methods with runtime and memory use that scales linearly with the total of protein atom count and solvent atom count, are accurate in their calculation of PG resonance and are applicable through the transition metal block are highly desirable. If such methods are wavefunction, high quality basis sets are necessary for accuracy with proteins (Chapter 3). Whereas present simulated time-scales extend to ~1 picosecond, the 1 nanosecond mark might become accessible with such methods. Whereas present solvation depth is 7 angstroms over short sidechains, 20 angstroms over the longest sidechains is desirable so that surface effects of the vacuum/water interface are not experienced at the water/protein interface, and this vast increase in solvent might become possible with such methods.

6.9 Future work

Summarizing the investigations that would be enabled by the availability of the methods with the above characteristics:

- Survey backbone hydration in all RAHB secondary structure types in the presence of each biologically relevant cation, including what combinations of residues are complementary for blocking backbone hydration
- With linear scaling NBO, analyse charge transfer and polarization in HB networks of water connecting PG and ions for each QMD frame
- Survey loss of sidechain complementation in proteins of natural sequence with increasing temperature and ion concentration in mixed ionic environments, validating against data which has been physically obtained
- Observe what happens at longer simulated time scales – is there indication that folding could proceed unassisted by chaperone in the case of direct contact between cation and backbone oxygen? If such contact occurs, attempt binding of molecular chaperones to cation-disrupted secondary structures

- In the more distant development of linear scaling QMD methods that have sufficiently low constant factor on resource use, simulate entire folding of fast folding proteins to observe cation-dependent backbone hydration

The combinatorial nature of these investigations requires many simulations and a total simulation time that makes necessary methods that are linear scaling with the total of protein, solvent and ion atom count.

Ahead of the availability of methods suitable for these surveys, the Protein Data Bank might be surveyed for existence of surface backbone oxygens in a largely hydrophobic surface patch, and what features of the patch might prevent chaperone binding or action considered.

6.10 Acknowledgements

Prof. John A. Carver is acknowledged for reading this manuscript and offering editing suggestions.

eResearch South Australia is acknowledged for hosting and administering machines provided under Australian Government Linkage, Infrastructure, Equipment and Facilities grants for Supercomputing in South Australia and directing funds to the acquisition of Nvidia Tesla GPU nodes.

6.11 Appendix 1. Backbone hydration in experiment 1302

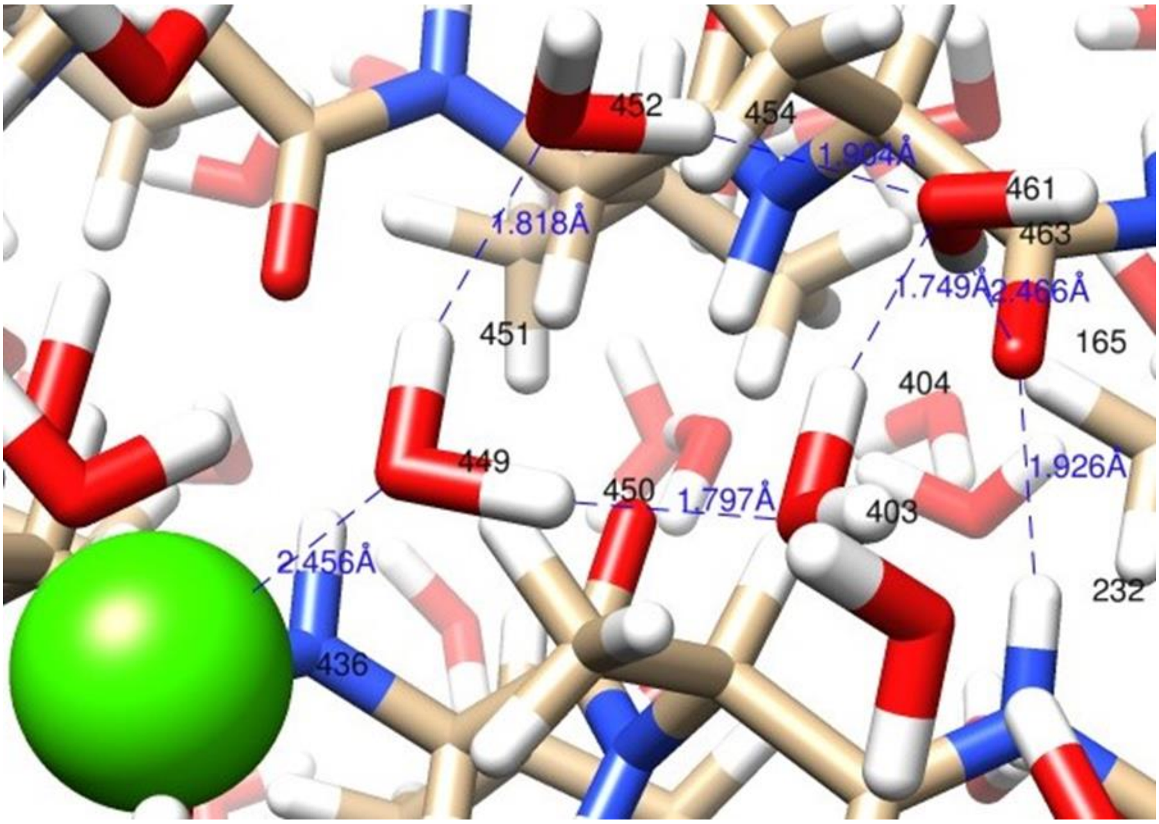


Figure 6.8. Backbone Hydration Outside Experimental Surface in Final Frame of Experiment 1302

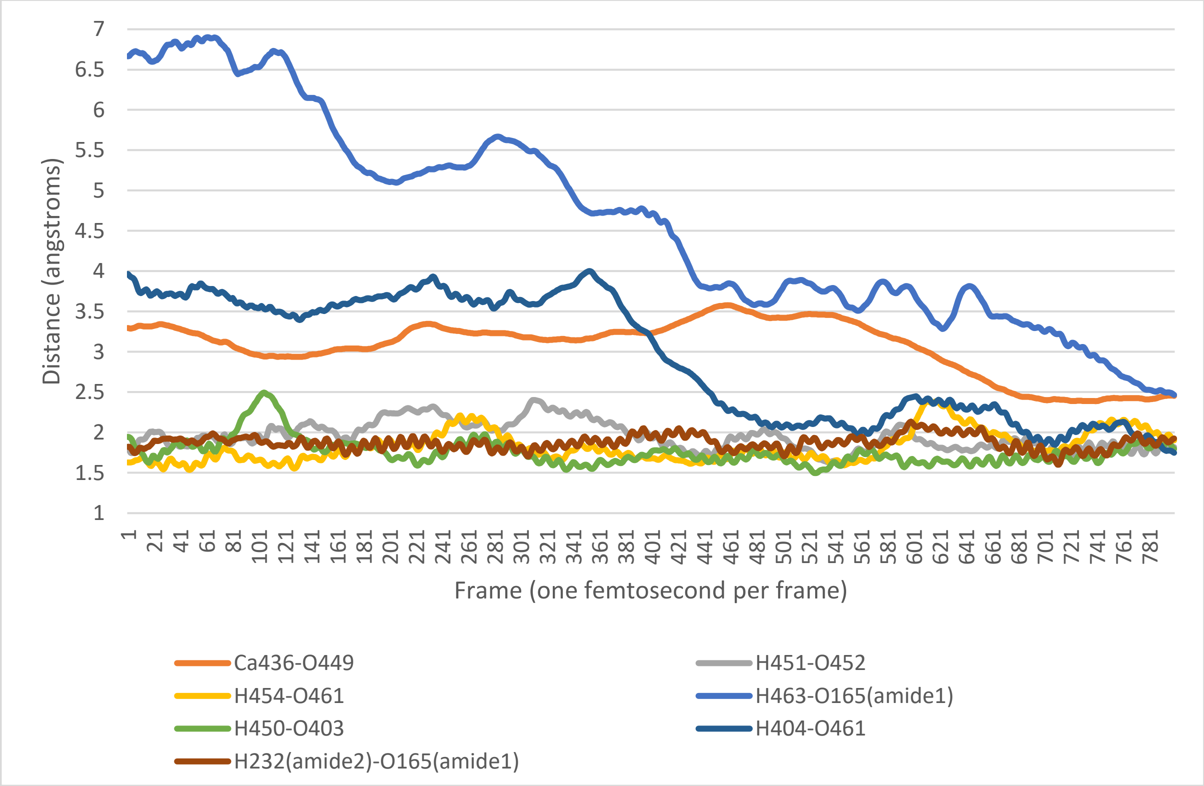


Figure 6.9. Selected bond lengths in experiment 1302. Atom Ids Shown in Figure 6.8

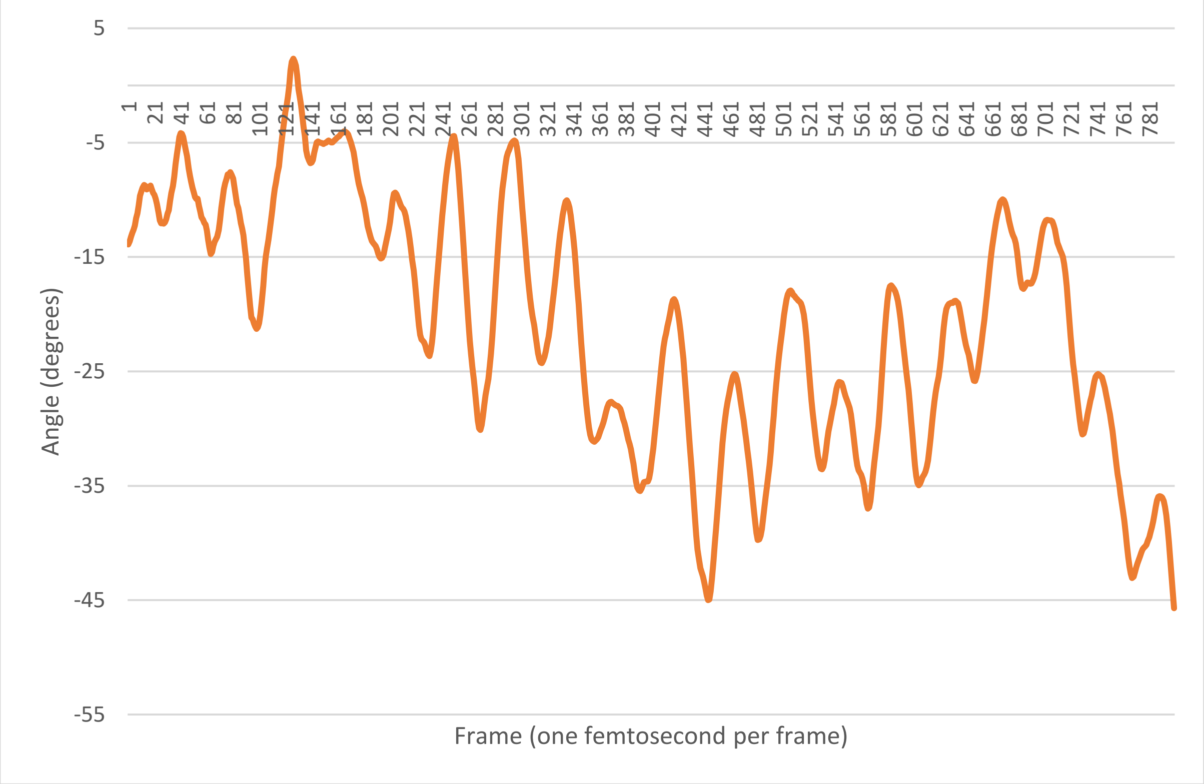


Figure 6.10. Angle of O165-H232 from O165, C163, N164 Normal Minus 90.0 in Experiment 1302

6.12 Appendix 2. Backbone hydration in experiment 1303

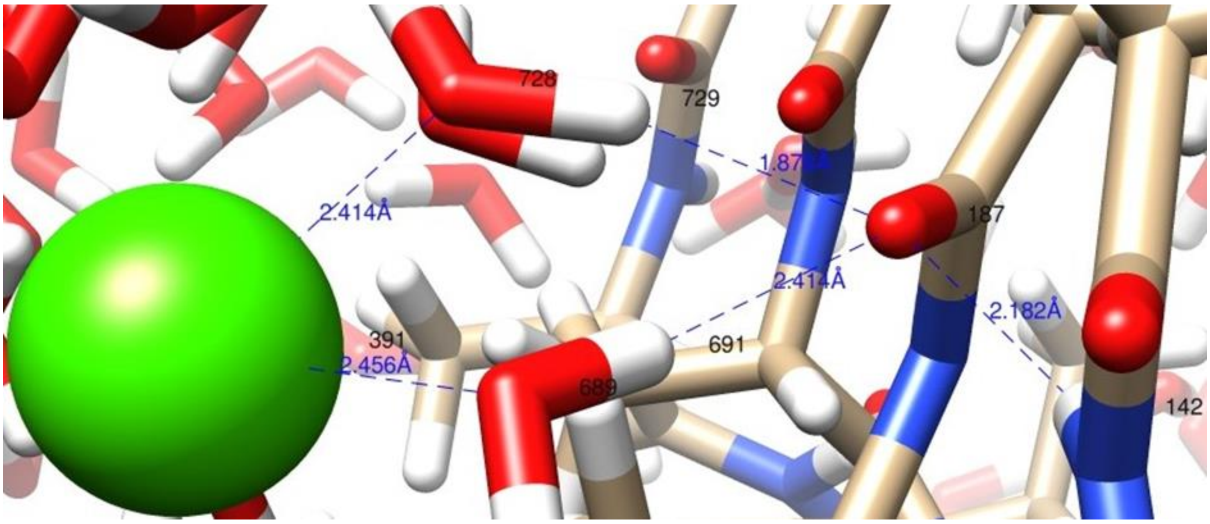


Figure 6.11. Backbone Hydration Outside Experimental Surface in Final Frame of Experiment 1303

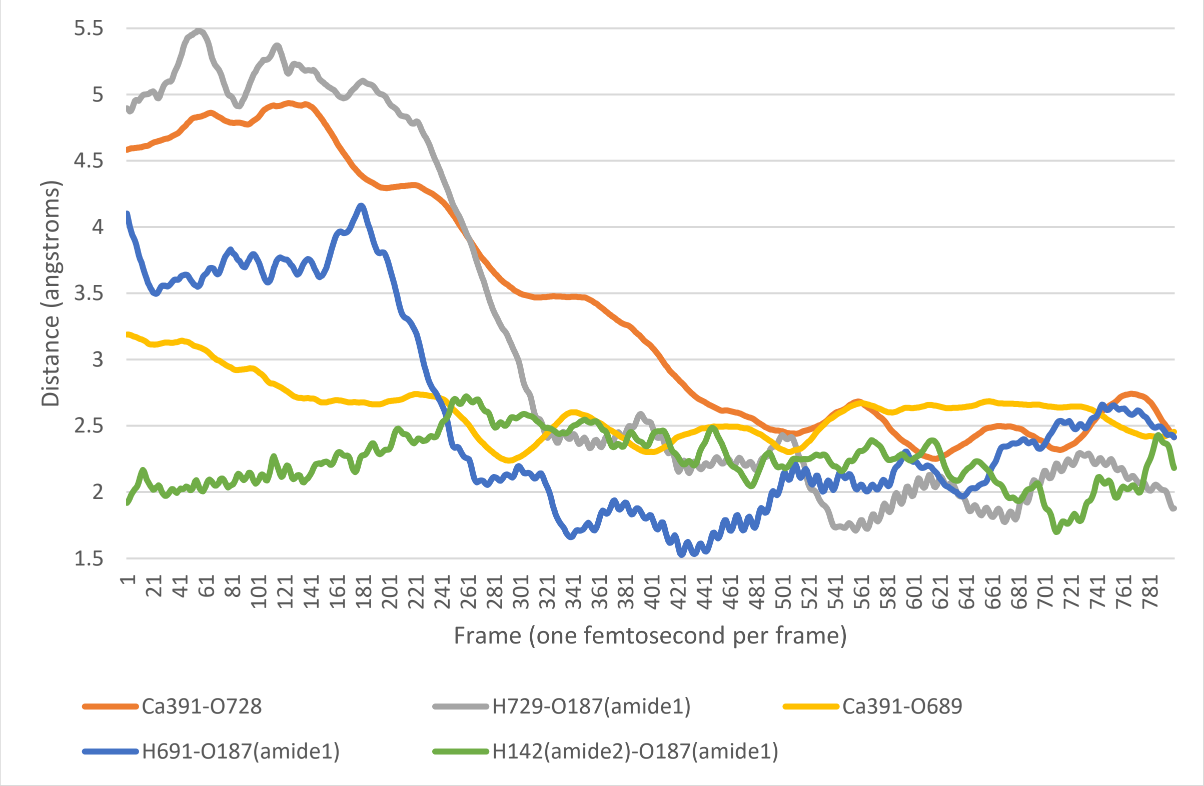


Figure 6.12. Selected Bond Lengths in Final Frame of Experiment 1303. Atom Ids Shown in Figure 6.11

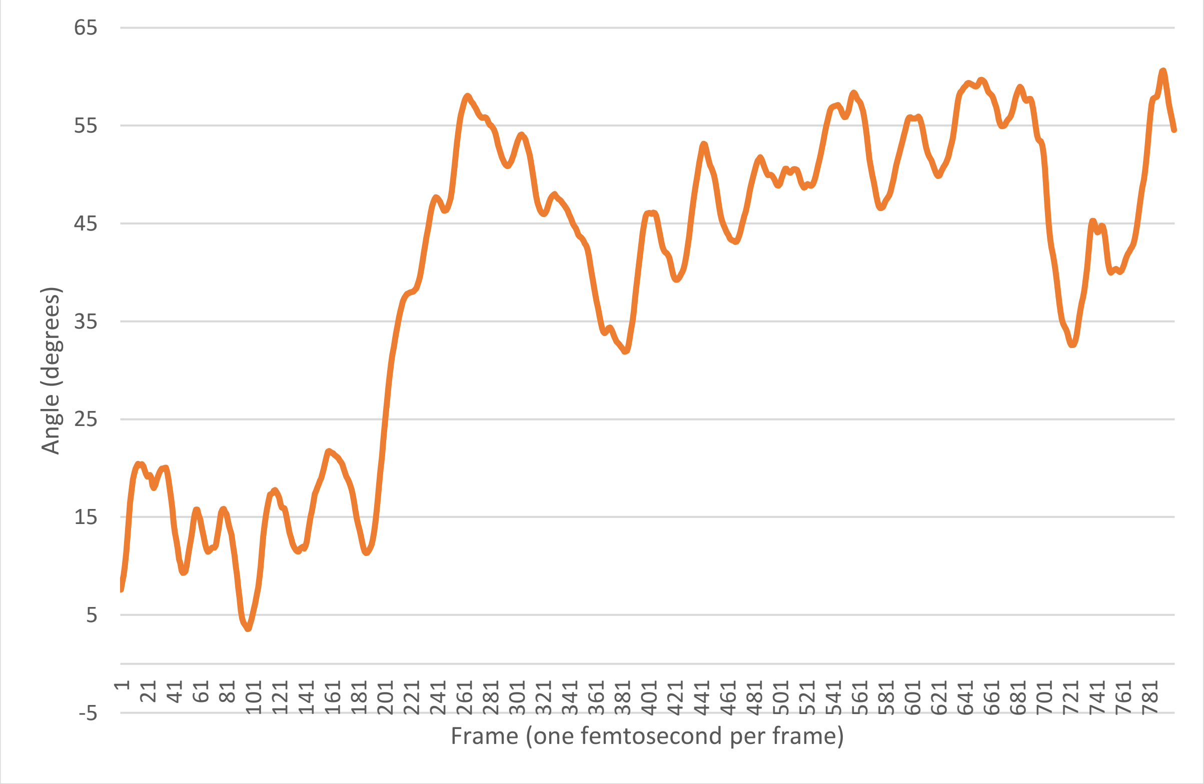


Figure 6.13. Angle of O187-H142 from O187, C185, N186 Normal Minus 90.0 in Experiment 1303

7 Conclusion

7.1 Significance and contribution

Given the ubiquity of proteins in biological processes and the relevance of fold to function [1], the protein folding problem may readily be viewed as a major unsolved problem in molecular life science and perhaps the pre-eminent such problem. Section 4.10.4 (Hypotheses/Protein folding) introduced a theory of protein folding, which is referred to here as the Resonance Theory of Protein Folding, RTPF. If understanding how proteins fold rather than predicting fold or folding kinetics is deemed to be the protein folding problem, this problem may have been solved by the present work. Of course, demonstration that this understanding is correct is needed, though the present work has provided evidence of elements of the proposed theory of protein folding. Investigation of the correctness of this theory of folding will occur over some years. This investigation is subject to the limitations of present methods mentioned below.

If RTPF is correct, much will follow. This theory would provide the basis necessary for accurate prediction of fold and folding kinetics. Protein folding having the proposed uniform mechanism facilitates protein design. Inverse design of proteins [232] proceeds from desired properties of a folded protein to an amino acid sequence for the protein, and this is simplified by the regular nature of the proposed protein folding mechanism. As an example of inverse design, design might proceed from specification of what bindings the new protein can have to other proteins and allosteric [175] constraints between these bindings. In another but related example, an entire complex of proteins might be automatically designed to satisfy specified signal transduction properties. When new species are to be designed other than by elimination of genes empirically found to be unnecessary for a desired purpose, inverse protein design based on the proposed theory of protein folding will prove indispensable. The progress toward a full molecular account of biology that such a theory represents makes plausible a goal of elimination of all disease no matter how rare by the end of the century. Having determined the mechanism responsible for specificity of structure in proteins, it may be possible to design other structurally specific polymers.

Until the proposal of RTPF, the backbone-based theory of protein folding was the most convincing theory of protein folding, the flaws of sidechain-sidechain interaction theories of protein folding having been revealed [3]. RTPF varies and extends that theory. RTPF places the mechanism of folding in the backbone itself rather than in backbone-backbone interactions, and allows the commonality of the folding process across all backbone conformations to be seen and reveals the amino acid sequence as a specification of fold which is interpreted by a universal folding mechanism. This mechanism is integral to and distributed throughout the polypeptide chain in the form of peptide group resonance, PR. These are notable conceptual advances.

Chapter 3 finds that established Density Functional Theory [81, 82], DFT, methods are subject to large errors in calculation of amide group properties when torsion is applied to the bond connecting the amide carbon to its substituent so as to resemble the geometry of a beta sheet. It is shown that these errors exist in beta sheets without artificial constraint, and further evidence for this and for errors in alpha helices is given in Chapter 5. These errors are far outside desirable limits and the calculated variation in amide resonance exceeds that due to physics in biological conditions. This renders these methods quite undesirable for application to proteins, since amide/peptide resonance is the resonance of Resonance-Assisted Hydrogen Bonding [4], RAHB, in backbone hydrogen bonding. RTPF rests upon PR, so accurate calculation of PR is essential for investigation of this theory. Extension of the amide species used in Chapter 3 to species resembling alanine and valine revealed that these errors were sensitive to local sidechain, introducing a false relationship between sidechain and protein structure. The number of papers referring to results of applying these methods to proteins/peptides is large, perhaps in the 10,000s. Developers and users of DFT methods alike need to be aware of this error, and also of the second error mentioned in Chapter 3, being of the MP2 [84] implementation dependent non-planarity of the amide group when correlation consistent basis sets are not used. The first mentioned error extends beyond proteins to conjugated systems in the presence of torsional hyperconjugation.

Chapter 4 demonstrates the variation in PR due to electrostatic field with C-N component using model amide molecules and good quality quantum chemical methods and basis sets. PR is sensitive to a biologically plausible electrostatic field with a C-N component, and quantification of this effect has not previously been reported and nor has a connection of this effect to RAHB and protein folding previously been recognised. This finding leads to a number of hypotheses in a straightforward manner. While change to barrier height of cis-trans isomerization and penalty for non-planarity [233] of X-Pro was not tested, the similarity of its electronic structure to the amide group means that it is highly likely that these quantities are significantly varied by electrostatic field with C-N component as for X-NotPro (i.e. other than proline), increased by field parallel to C->N and decreased by field antiparallel to C->N. It is possible prolyl isomerases [9] utilize this effect to lower the energy barrier to cis-trans prolyl isomerization. There is interest in prolyl isomerization for its implications for protein folding, and the observed effect is a new factor in this isomerization. Variation in PR and its associated energy due to the electrostatic field with a C-N component does not require the involvement of RAHB, and variation of PR and associated stabilization of polyproline helices [11] due to electrostatic field with component aligned with the helix axis, along with PR changes of RAHB with water, is hypothesized. Other hypotheses proposed concern the role of this effect in protein complexation and molecular chaperone action. An analogous effect is proposed for nitrogenous base pairing [12] with implications for base pair opening which is relevant to replication and transcription of nucleic acids [234]. Hypotheses concerning the stability of beta sheets and amyloid fibrils [10] are offered, though the cause of variation in RAHB leading to destabilization is not limited to this effect.

Chapter 5 tentatively finds that cooperative cycles of hydrogen bonding are radically disfavoured in proteins. The finding is tentative because it rests on detection of cycles by a program without confirmation, except if a possible example is flagged for investigation by quantum chemical methods, and independent confirmation is highly desirable. No indication of such a cycle was found in the literature. Whether such cycles exist is an important matter, for if they do, a range of phenomena as described in Section 5.4.2 can be anticipated, as can frustration of the extension of secondary structures during folding. The intrinsic limitation of cyclic cooperativity to the least cooperativity in the cycle, the sensitivity of cooperativity to geometry of hydrogen bonding between amides and that these cycles cannot extend without losing all cyclic cooperativity favour the tentative finding. Whether the absence of these cycles in proteins arises because of the properties of proteins or the intrinsic weaknesses of cooperative cycles needs to be determined, the result of which would have implications for the design of other polymers of programmed specific structure. If it happens the tentative finding is false, investigation of the range of phenomena then anticipated is important in understanding protein function. In any event, opening the debate as to whether such cycles exist in proteins is important. That the Natural Bond Orbital [6], NBO, binding energy of amide-amide hydrogen bonding is close to zero and that the hydrogen binding energy is primarily associated with the resonance of the amides is an important finding in its own right, one reason for which being that the amide-amide binding energy will be substantially subject to variation by electrostatic field with an amide C-N component. The observation that the optimal hydrogen bonding angle from linear between N-methylformamide oxygen and hydrogen fluoride, HF, is 15 degrees greater than the 60 degrees expected in an equivalent sp2 view of amide/carbonyl lone pairs is a contribution. The observation that two HF, one hydrogen bonded to each lobe of an amide oxygen p-type lone pair increases amide resonance 70% over that associated with a single such HF may be useful in design of RAHB polymers of improved stability since resonance is cooperative in extended RAHB structures.

Chapter 6 investigates the finding that amino acid preference for inclusion in a beta sheet correlates with sidechain steric blocking of hydration of the backbone in that preferred conformation [15], and does so by Quantum Molecular Dynamics, QMD, simulation of a beta sheet with full quantum mechanical treatment of each solvent molecule. Backbone hydration via a polyalanine surface is observed, but only in the presence of the cation used for these experiments, Ca^{2+}. Dependence of amino acid preference for beta sheet on the presence of a cation is novel, and leads to the suggestion that aspects of Hofmeister salt effects [228, 229, 235] arise from dysregulation of the amino acid secondary structure preferences which depend on the extent of backbone hydration. QMD observation of the first stages of the unfolding of a beta sheet due to the presence of alpha helix preferring residues is novel. If the suggestion that peptide oxygen in an otherwise hydrophobic surface patch, perhaps bound to a multivalent cation, is a signal for molecular chaperone [13] action proves correct, it would be quite useful in understanding chaperone function.

7.2 Future work

The veracity of the hypotheses appearing in this thesis and mentioned above will be investigated as the availability of suitable methods permits. Presently available methods will allow gas-phase investigation variation by electrostatic field with C-N component of cis-trans isomerization barrier height of X-Pro peptide, stability of polyproline helices and of RAHB in nitrogenous base pairs. These investigations may be conducted soon. QMD investigation of the other hypotheses requires a linear scaling method which accurately calculates PR, electrostatic field variation of PR, RAHB and the properties of water. The field of research of such methods is vigorous, and with the development of methods that have a constant factor or prefactor sufficiently low, observation of the complete folding of fast folding proteins can commence. It is presumed that such methods will be available before physical experimental methods for observation of protein structural dynamics and each PR throughout the folding of a single molecule are available. Such QMD or physical methods are necessary for a full investigation of RTPF.

Means of exploiting the regular and uniform properties of the universal folding mechanism given by RTPF for prediction and inverse design of fold will be considered. This might be pursued ahead of full confirmation of RTPF, and support for RTPF would be given by accurate protein fold predictions.

References

1. Kendrew, J.C., *The structure of globular proteins.* Comparative Biochemistry and Physiology, 1962. **4**(2–4): p. 249-252.
2. Anson, M.L. and A.E. Mirsky, *Protein Coagulation and its Reversal : Globin.* The Journal of General Physiology, 1931. **14**(5): p. 605-609.
3. Rose, G.D., et al., *A backbone-based theory of protein folding.* Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16623-33.
4. Gilli, G., et al., *Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the .beta.-diketone fragment.* Journal of the American Chemical Society, 1989. **111**(3): p. 1023-1028.
5. Bertolasi, V., et al., *Intermolecular N–H···O hydrogen bonds assisted by resonance. Heteroconjugated systems as hydrogen-bond-strengthening functional groups.* Acta Crystallographica Section B, 1995. **51**(6): p. 1004-1015.
6. Weinhold, F.A. and C.R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*. 2005: Cambridge University Press.
7. Pauling, L. and R.B. Corey, *The Pleated Sheet, A New Layer Configuration of Polypeptide Chains.* Proceedings of the National Academy of Sciences, 1951. **37**(5): p. 251-256.
8. Pauling, L., R.B. Corey, and H.R. Branson, *The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain.* Proceedings of the National Academy of Sciences, 1951. **37**(4): p. 205-211.
9. Fischer, G. and F.X. Schmid, *The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell.* Biochemistry, 1990. **29**(9): p. 2205-2212.
10. Sunde, M., et al., *Common core structure of amyloid fibrils by synchrotron X-ray diffraction1.* Journal of Molecular Biology, 1997. **273**(3): p. 729-739.
11. Adzhubei, A.A., M.J.E. Sternberg, and A.A. Makarov, *Polyproline-II Helix in Proteins: Structure and Function.* Journal of Molecular Biology, 2013. **425**(12): p. 2100-2132.
12. Watson, J.D. and F.H.C. Crick, *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.* Nature, 1953. **171**(4356): p. 737-738.
13. Ellis, J., *Proteins as molecular chaperones.* Nature, 1987. **328**(6129): p. 378-379.
14. Moran, D., et al., *Popular Theoretical Methods Predict Benzene and Arenes To Be Nonplanar.* Journal of the American Chemical Society, 2006. **128**(29): p. 9342-9343.
15. Bai, Y. and S.W. Englander, *Hydrogen bond strength and β-sheet propensities: The role of a side chain blocking effect.* Proteins: Structure, Function, and Bioinformatics, 1994. **18**(3): p. 262-266.
16. Hausmann, R., *To Grasp the Essence of Life: A History of Molecular Biology*. 1 ed. 2002: Springer Netherlands.
17. Vauquelin, L.N. and P.J. Robiquet, *The discovery of a new plant principle in Asparagus sativus.* Annales de Chimie 1806. **57**: p. 88–93.
18. McCoy, R.H., C.E. Meyer, and W.C. Rose, *Feeding Experiments with Mixtures of Highly Purified Amino Acids: VIII. Isolation and Identification of a New Essential Amino Acid.* Journal of Biological Chemistry, 1935. **112**(1): p. 283-302.
19. Moore, T.S. and T.F. Winmill, J. Chem. Soc., 1912. **101**.
20. Pauling, L., *The Structure and Entropy of Ice and of Other Crystals with Some Randomness of Atomic Arrangement.* Journal of the American Chemical Society, 1935. **57**(12): p. 2680-2684.
21. Isaacs, E.D., et al., *Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement.* Physical Review Letters, 1999. **82**(3): p. 600-603.
22. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains.* Science, 1973. **181**(4096): p. 223-230.
23. Anfinsen, C.B., et al., *The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain.* Proceedings of the National Academy of Sciences of the United States of America, 1961. **47**(9): p. 1309-1314.

24. Chellapa, G.D. and G.D. Rose, *On interpretation of protein X-ray structures: Planarity of the peptide unit.* Proteins: Structure, Function, and Bioinformatics, 2015. **83**(9): p. 1687-1692.
25. Edison, A.S., *Linus Pauling and the planar peptide bond.* Nat Struct Mol Biol, 2001. **8**(3): p. 201-202.
26. Wedemeyer, W.J., E. Welker, and H.A. Scheraga, *Proline Cis−Trans Isomerization and Protein Folding.* Biochemistry, 2002. **41**(50): p. 14637-14644.
27. Levinthal, C., *How to Fold Graciously*, in *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House*. 1969: Monticello Illinois. p. 22–24.
28. Baldwin, R., *The nature of protein folding pathways: The classical versus the new view.* Journal of Biomolecular NMR, 1995. **5**(2): p. 103-109.
29. Englander, S.W. and L. Mayne, *The nature of protein folding pathways.* Proceedings of the National Academy of Sciences, 2014. **111**(45): p. 15873-15880.
30. Bryngelson, J.D. and P.G. Wolynes, *Spin glasses and the statistical mechanics of protein folding.* Proceedings of the National Academy of Sciences, 1987. **84**(21): p. 7524-7528.
31. Anderson, P., *Spin Glass VI: Spin Glass as Cornucopia*, in *Physics Today*. 1989, American Institute of Physics. p. 9-11.
32. Cook, S.A., *The complexity of theorem-proving procedures*, in *Proceedings of the third annual ACM symposium on Theory of computing*. 1971, ACM: Shaker Heights, Ohio, USA. p. 151-158.
33. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The Energy Landscapes and Motions of Proteins.* Science, 1991. **254**(5038): p. 1598-1603.
34. Bryngelson, J.D., et al., *Funnels, pathways, and the energy landscape of protein folding: A synthesis.* Proteins: Structure, Function, and Bioinformatics, 1995. **21**(3): p. 167-195.
35. Go, N., *Theoretical Studies of Protein Folding.* Annual Review of Biophysics and Bioengineering, 1983. **12**(1): p. 183-210.
36. Anderson, P.W., *Spin Glass V: Real Power Brought to Bear.* Physics Today, 1989. **42**(7): p. 9-11.
37. Auton, M. and D.W. Bolen, *Additive Transfer Free Energies of the Peptide Backbone Unit That Are Independent of the Model Compound and the Choice of Concentration Scale.* Biochemistry, 2004. **43**(5): p. 1329-1342.
38. Lin, T.-Y. and S.N. Timasheff, *Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein.* Biochemistry, 1994. **33**(42): p. 12695-12701.
39. Rose, G.D., L.M. Glerasch, and J.A. Smith, *Turns in Peptides and Proteins*, in *Advances in Protein Chemistry*. 1985, Academic Press. p. 1-109.
40. Drenth, J., et al., *Structure of Papain.* Nature, 1968. **218**(5145): p. 929-932.
41. Karplus, M. and D.L. Weaver, *Protein-folding dynamics.* Nature, 1976. **260**(5550): p. 404-406.
42. Frieden, C., *The Kinetics of Side Chain Stabilization during Protein Folding.* Biochemistry, 2003. **42**(43): p. 12439-12446.
43. Fleming, P.J., H. Gong, and G.D. Rose, *Secondary structure determines protein topology.* Protein Science, 2006. **15**(8): p. 1829-1834.
44. Krishna, M.M.G., Y. Lin, and S. Walter Englander, *Protein Misfolding: Optional Barriers, Misfolded Intermediates, and Pathway Heterogeneity.* Journal of Molecular Biology, 2004. **343**(4): p. 1095-1109.
45. Ptitsyn, O.B., *Molten Globule and Protein Folding*, in *Advances in Protein Chemistry*. 1995, Academic Press. p. 83-229.
46. Kuwajima, K., *The molten globule state of alpha-lactalbumin.* The FASEB Journal, 1996. **10**(1): p. 102-9.
47. Baldwin, R.L. and G.D. Rose, *Molten globules, entropy-driven conformational change and protein folding.* Current Opinion in Structural Biology, 2013. **23**(1): p. 4-10.
48. Mezei, M., et al., *Polyproline II helix is the preferred conformation for unfolded polyalanine in water.* Proteins: Structure, Function, and Bioinformatics, 2004. **55**(3): p. 502-507.
49. Gong, H., et al., *Local Secondary Structure Content Predicts Folding Rates for Simple, Two-state Proteins.* Journal of Molecular Biology, 2003. **327**(5): p. 1149-1154.
50. Myers, J.K. and C.N. Pace, *Hydrogen bonding stabilizes globular proteins.* Biophysical Journal, 1996. **71**(4): p. 2033-2039.

51. Pace, C.N., et al., *Contribution of hydrogen bonds to protein stability.* Protein Science, 2014. **23**(5): p. 652-661.
52. Kauzmann, W., *Some Factors in the Interpretation of Protein Denaturation*, in *Advances in Protein Chemistry*. 1959, Academic Press. p. 1-63.
53. Baskakov, I. and D.W. Bolen, *Forcing Thermodynamically Unfolded Proteins to Fold.* Journal of Biological Chemistry, 1998. **273**(9): p. 4831-4834.
54. Auton, M. and D.W. Bolen, *Predicting the energetics of osmolyte-induced protein folding/unfolding.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(42): p. 15065-15068.
55. Weinhold, F., C.R. Landis, and E.D. Glendening, *What is NBO analysis and how is it useful?* International Reviews in Physical Chemistry, 2016. **35**(3): p. 399-440.
56. Fock, V.A., Z. Phys., 1930. **61**: p. 126.
57. Glendening, E.D., C.R. Landis, and F. Weinhold, *NBO 6.0: Natural bond orbital analysis program.* Journal of Computational Chemistry, 2013. **34**(16): p. 1429-1437.
58. Hudak, P., et al., *A history of Haskell: being lazy with class*, in *Proceedings of the third ACM SIGPLAN conference on History of programming languages*. 2007, ACM: San Diego, California. p. 12-1-12-55.
59. *Extensible Markup Language (XML)*. Available from: http://www.w3.org/XML/.
60. Consortium, W.W.W.
61. Backus, J., *The history of FORTRAN I, II, and III.* SIGPLAN Not., 1978. **13**(8): p. 165-180.
62. W3C, *Document Object Model*. 2009.
63. W3C, *XSL Transformations (XSLT) Version 3.0*. 2014.
64. Rosenblum, L.J., *Scientific Visualization: Advances and Challenges*. 1994: Academic.
65. W3C, *XQuery 3.0: An XML Query Language*. 2014.
66. Hudak, P., *Conception, evolution, and application of functional programming languages.* ACM Comput. Surv., 1989. **21**(3): p. 359-411.
67. Web3DConsortium, *X3D*. 2013.
68. Strothotte, T., *Computational Visualization: Graphics, Abstraction, and Interactivity*. 1998: Springer.
69. *Saxon-PE*. 2015, Saxonica.
70. W3C, *XML Path Language (XPath) 3.1*. 2015.
71. Behr, J., U. Bockholt, and D. Fellner, *Instantreality — A Framework for Industrial Augmented and Virtual Reality Applications*, in *Virtual Reality & Augmented Reality in Industry: The 2nd Sino-German Workshop*, D. Ma, et al., Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 91-99.
72. Neese, F., *The ORCA program system.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(1): p. 73-78.
73. Kossmann, S. and F. Neese, *Efficient Structure Optimization with Second-Order Many-Body Perturbation Theory: The RIJCOSX-MP2 Method.* Journal of Chemical Theory and Computation, 2010. **6**(8): p. 2325-2338.
74. Riplinger, C. and F. Neese, *An efficient and near linear scaling pair natural orbital based local coupled cluster method.* The Journal of Chemical Physics, 2013. **138**(3): p. 034106.
75. Frisch, M.J., et al., *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.
76. Stroustrup, B., *Evolving a language in and for the real world: C++ 1991-2006*, in *Proceedings of the third ACM SIGPLAN conference on History of programming languages*. 2007, ACM: San Diego, California. p. 4-1-4-59.
77. Leskovec, J. and R. Sosi. *SNAP: A general purpose network analysis and graph mining library in C++*. 2014; Available from: http://snap.stanford.edu/snap.
78. Bertolasi, V., et al., *Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters in eight intramolecularly hydrogen bonded 1,3-diaryl-1,3-propanedione enols.* Journal of the American Chemical Society, 1991. **113**(13): p. 4917-4925.
79. Bertolasi, V., et al., *pi-Bond cooperativity and anticooperativity effects in resonance-assisted hydrogen bonds (RAHBs).* Acta Crystallogr B, 2006. **62**(Pt 5): p. 850-63.

80. Weinhold, F. and C.R. Landis, *Discovering Chemistry with Natural Bond Orbitals*. 2012: Wiley.

81. Burke, K., *Perspective on density functional theory.* The Journal of Chemical Physics, 2012. **136**(15): p. 150901.

82. Becke, A.D., *Perspective: Fifty years of density-functional theory in chemical physics.* The Journal of Chemical Physics, 2014. **140**(18): p. 18A301.

83. Clauss, A.D., et al., *Rabbit ears concepts of water lone pairs: a reply to comments of Hiberty, Danovich, and Shaik.* Chemistry Education Research and Practice, 2015.

84. Moeller, C. and M.S. Plesset, *Note on an approximation treatment for many-electron systems.* Phys. Rev., 1934. **46**: p. 0618-22.

85. Dunning, T.H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen.* The Journal of Chemical Physics, 1989. **90**(2): p. 1007-1023.

86. Grimme, S., *Semiempirical hybrid density functional with perturbative second-order correlation.* The Journal of Chemical Physics, 2006. **124**(3): p. 034108.

87. Zhao, Y., B.J. Lynch, and D.G. Truhlar, *Doubly Hybrid Meta DFT: New Multi-Coefficient Correlation and Density Functional Methods for Thermochemistry and Thermochemical Kinetics.* The Journal of Physical Chemistry A, 2004. **108**(21): p. 4786-4791.

88. Vydrov, O.A. and G.E. Scuseria, *Assessment of a long-range corrected hybrid functional.* The Journal of Chemical Physics, 2006. **125**(23): p. 234109.

89. Boys, S.F. and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors.* Molecular Physics, 1970. **19**(4): p. 553-566.

90. Gupta, A. and R.J. Boyd, *Density difference representation of electron correlation.* The Journal of Chemical Physics, 1978. **68**(4): p. 1951-1957.

91. Ufimtsev, I.S. and T.J. Martinez, *Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics.* Journal of Chemical Theory and Computation, 2009. **5**(10): p. 2619-2628.

92. Kästner, J., et al., *DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations†.* The Journal of Physical Chemistry A, 2009. **113**(43): p. 11856-11865.

93. Grimme, S., et al., *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu.* The Journal of Chemical Physics, 2010. **132**(15): p. 154104.

94. Grimme, S., S. Ehrlich, and L. Goerigk, *Effect of the damping function in dispersion corrected density functional theory.* Journal of Computational Chemistry, 2011. **32**(7): p. 1456-1465.

95. Glendening, E.D., et al., *NBO 6.0*. 2013, Theoretical Chemistry Institute, University of Wisconsin, Madison.

96. *Excel*. 2013, Microsoft.

97. *Jmol: an open-source Java viewer for chemical structures in 3D*.

98. *Gaussian 09 User's Reference*. Available from: http://www.gaussian.com/g_tech/g_ur/g09help.htm.

99. Van Voorhis, T. and G.E. Scuseria, *A novel form for the exchange-correlation energy functional.* The Journal of Chemical Physics, 1998. **109**(2): p. 400-410.

100. Peverati, R. and D.G. Truhlar, *Exchange–Correlation Functional with Good Accuracy for Both Structural and Energetic Properties while Depending Only on the Density and Its Gradient.* Journal of Chemical Theory and Computation, 2012. **8**(7): p. 2310-2319.

101. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* The Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.

102. Peverati, R. and D.G. Truhlar, *Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics.* Physical Chemistry Chemical Physics, 2012. **14**(47): p. 16187-16191.

103. Chai, J.-D. and M. Head-Gordon, *Systematic optimization of long-range corrected hybrid density functionals.* The Journal of Chemical Physics, 2008. **128**(8): p. 084106.

104. Schwabe, T. and S. Grimme, *Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects.* Physical Chemistry Chemical Physics, 2006. **8**(38): p. 4398-4401.

105. Grimme, S., *Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies.* The Journal of Chemical Physics, 2003. **118**(20): p. 9095-9102.
106. Neese, F., et al., *Assessment of Orbital-Optimized, Spin-Component Scaled Second-Order Many-Body Perturbation Theory for Thermochemistry and Kinetics.* Journal of Chemical Theory and Computation, 2009. **5**(11): p. 3060-3073.
107. Meissner, L., *Size-consistency corrections for configuration interaction calculations.* Chemical Physics Letters, 1988. **146**(3–4): p. 204-210.
108. Iikura, H., et al., *A long-range correction scheme for generalized-gradient-approximation exchange functionals.* The Journal of Chemical Physics, 2001. **115**(8): p. 3540-3544.
109. Becke, A.D., *Real-space post-Hartree–Fock correlation models.* The Journal of Chemical Physics, 2005. **122**(6): p. 064101.
110. Scuseria, G.E., C.L. Janssen, and H.F. Schaefer, *An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations.* The Journal of Chemical Physics, 1988. **89**(12): p. 7382-7387.
111. Pettersen, E.F., et al., *UCSF Chimera—A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
112. Peverati, R. and D.G. Truhlar, *Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2014. **372**(2011).
113. Peverati, R. and D.G. Truhlar, *An improved and broadly accurate local approximation to the exchange-correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics.* Physical Chemistry Chemical Physics, 2012. **14**(38): p. 13171-13174.
114. Zhao, Y. and D.G. Truhlar, *Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States.* The Journal of Physical Chemistry A, 2006. **110**(49): p. 13126-13130.
115. Mardirossian, N. and M. Head-Gordon, *How Accurate Are the Minnesota Density Functionals for Noncovalent Interactions, Isomerization Energies, Thermochemistry, and Barrier Heights Involving Molecules Composed of Main-Group Elements?* Journal of Chemical Theory and Computation, 2016.
116. Perdew, J.P. and K. Schmidt, *Jacob's ladder of density functional approximations for the exchange-correlation energy.* AIP Conference Proceedings, 2001. **577**(1): p. 1-20.
117. Perdew, J.P., et al., *Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits.* The Journal of Chemical Physics, 2005. **123**(6): p. 062201.
118. Grimme, S., *Semiempirical GGA-type density functional constructed with a long-range dispersion correction.* Journal of Computational Chemistry, 2006. **27**(15): p. 1787-1799.
119. Rohrdanz, M.A., K.M. Martins, and J.M. Herbert, *A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states.* The Journal of Chemical Physics, 2009. **130**(5): p. 054112.
120. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple.* Physical Review Letters, 1996. **77**(18): p. 3865-3868.
121. Vosko, S.H., L. Wilk, and M. Nusair, *Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis.* Canadian Journal of Physics, 1980. **58**(8): p. 1200-1211.
122. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**(2): p. 785-789.
123. Gill, P.M.W., *A new gradient-corrected exchange functional.* Molecular Physics, 1996. **89**(2): p. 433-445.

124. Slater, J.C., *The Self-Consistent Field for Molecules and Solids, Quantum Theory of Molecules and Solids* Vol. 4. 1974, New York: McGraw-Hill.
125. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects.* Physical Review, 1965. **140**(4A): p. A1133-A1138.
126. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review, 1964. **136**(3B): p. B864-B871.
127. Becke, A.D. and M.R. Roussel, *Exchange holes in inhomogeneous systems: A coordinate-space model.* Physical Review A, 1989. **39**(8): p. 3761-3767.
128. Yu, H.S., X. He, and D.G. Truhlar, *MN15-L: A New Local Exchange-Correlation Functional for Kohn–Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids.* Journal of Chemical Theory and Computation, 2016. **12**(3): p. 1280-1293.
129. Ditchfield, R., W.J. Hehre, and J.A. Pople, *Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules.* The Journal of Chemical Physics, 1971. **54**(2): p. 724-728.
130. Martin, J.M.L., P.R. Taylor, and T.J. Lee, *The harmonic frequencies of benzene. A case for atomic natural orbital basis sets.* Chemical Physics Letters, 1997. **275**(3–4): p. 414-422.
131. Martin, J.M.L., T.J. Lee, and P.R. Taylor, *A purely ab initio spectroscopic quality quartic force field for acetylene.* The Journal of Chemical Physics, 1998. **108**(2): p. 676-691.
132. Boese, A.D. and N.C. Handy, *New exchange-correlation density functionals: The role of the kinetic-energy density.* The Journal of Chemical Physics, 2002. **116**(22): p. 9559-9569.
133. Peverati, R. and D.G. Truhlar, *M11-L: A Local Density Functional That Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics.* The Journal of Physical Chemistry Letters, 2012. **3**(1): p. 117-124.
134. Zhao, Y. and D. Truhlar, *The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals.* Theoretical Chemistry Accounts, 2008. **120**(1-3): p. 215-241.
135. Hegarty, D. and M.A. Robb, *Application of Unitary Group-Methods to Configuration-Interaction Calculations.* Molecular Physics, 1979. **38**(6): p. 1795-1812.
136. Dunning, T.H. and P.J. Hay, *Modern Theoretical Chemistry*. Vol. 3. 1977, New York: Plenum.
137. Weigend, F. and R. Ahlrichs, *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy.* Physical Chemistry Chemical Physics, 2005. **7**(18): p. 3297-3305.
138. Yu, H.S., et al., *MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions.* Chemical Science, 2016. **7**(8): p. 5032-5051.
139. Tubman, N.M., et al., *A deterministic alternative to the full configuration interaction quantum Monte Carlo method.* The Journal of Chemical Physics, 2016. **145**(4): p. 044112.
140. Holmes, A.A., N.M. Tubman, and C.J. Umrigar, *Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling.* Journal of Chemical Theory and Computation, 2016. **12**(8): p. 3674-3680.
141. Robson, B. and R.H. Pain, *Analysis of the code relating sequence to conformation in proteins: Possible implications for the mechanism of formation of helical regions.* Journal of Molecular Biology, 1971. **58**(1): p. 237-257.
142. Shaik, S., S.P. de Visser, and D. Kumar, *External Electric Field Will Control the Selectivity of Enzymatic-Like Bond Activations.* Journal of the American Chemical Society, 2004. **126**(37): p. 11746-11749.
143. Meir, R., et al., *Oriented Electric Fields Accelerate Diels–Alder Reactions and Control the endo/exo Selectivity.* ChemPhysChem, 2010. **11**(1): p. 301-310.
144. Gryn'ova, G., et al., *Switching radical stability by pH-induced orbital conversion.* Nat Chem, 2013. **5**(6): p. 474-481.
145. Gryn’ova, G. and M.L. Coote, *Origin and Scope of Long-Range Stabilizing Interactions and Associated SOMO–HOMO Conversion in Distonic Radical Anions.* Journal of the American Chemical Society, 2013. **135**(41): p. 15392-15403.

146. Aragonès, A.C., et al., *Electrostatic catalysis of a Diels–Alder reaction.* Nature, 2016. **531**(7592): p. 88-91.
147. Shaik, S., et al., *Charge-shift bonding and its manifestations in chemistry.* Nat Chem, 2009. **1**(6): p. 443-449.
148. Sini, G., et al., *Covalent, ionic and resonating single bonds.* Journal of Molecular Structure: THEOCHEM, 1991. **229**: p. 163-188.
149. Warshel, A., et al., *Electrostatic Basis for Enzyme Catalysis.* Chemical Reviews, 2006. **106**(8): p. 3210-3235.
150. Hirao, H., et al., *Effect of External Electric Fields on the C−H Bond Activation Reactivity of Nonheme Iron−Oxo Reagents.* Journal of the American Chemical Society, 2008. **130**(11): p. 3319-3327.
151. Lai, W., et al., *External Electric Field Can Control the Catalytic Cycle of Cytochrome P450cam: A QM/MM Study.* The Journal of Physical Chemistry Letters, 2010. **1**(14): p. 2082-2087.
152. Fried, S.D., S. Bagchi, and S.G. Boxer, *Extreme electric fields power catalysis in the active site of ketosteroid isomerase.* Science, 2014. **346**(6216): p. 1510.
153. Saitta, A.M. and F. Saija, *Miller experiments in atomistic computer simulations.* Proceedings of the National Academy of Sciences, 2014. **111**(38): p. 13768-13773.
154. Karafiloglou, P., *Control of delocalization and structural changes by means of an electric field.* Journal of Computational Chemistry, 2006. **27**(15): p. 1883-1891.
155. Boxer, S.G., *Stark Realities†.* The Journal of Physical Chemistry B, 2009. **113**(10): p. 2972-2983.
156. *Peptide resonance states: Effect of external field*. 2006; Available from: http://people.cs.uchicago.edu/~ridg/cspbnfolektur/omegaflex.pdf.
157. Bublitz, G.U. and S.G. Boxer, *Stark Spectroscopy: Applications in Chemistry, Biology, and Materials Science.* Annual Review of Physical Chemistry, 1997. **48**(1): p. 213-242.
158. Geissinger, P., B.E. Kohler, and J.C. Woehl, *Electric field and structure in the myoglobin heme pocket.* The Journal of Physical Chemistry, 1995. **99**(45): p. 16527-16529.
159. Callis, P.R. and B.K. Burgess, *Tryptophan Fluorescence Shifts in Proteins from Hybrid Simulations: An Electrostatic Approach.* The Journal of Physical Chemistry B, 1997. **101**(46): p. 9429-9432.
160. Manas, E.S., et al., *The Influence of Protein Environment on the Low Temperature Electronic Spectroscopy of Zn-Substituted Cytochrome c.* The Journal of Physical Chemistry B, 2000. **104**(29): p. 6932-6941.
161. Drobizhev, M., et al., *Long- and Short-Range Electrostatic Fields in GFP Mutants: Implications for Spectral Tuning.* Scientific Reports, 2015. **5**: p. 13223.
162. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**(6): p. 3098-3100.
163. Adamo, C. and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model.* The Journal of Chemical Physics, 1999. **110**(13): p. 6158-6170.
164. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
165. Bragg, L., *The Rutherford Memorial Lecture, 1960. The Development of X-ray Analysis.* Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1961. **262**(1309): p. 145-158.
166. Palmer, A.G. and D.J. Patel, *Kurt Wuthrich and NMR of Biological Macromolecules.* Structure. **10**(12): p. 1603-1604.
167. Serrano, L. and A.R. Fersht, *Capping and [alpha]-helix stability.* Nature, 1989. **342**(6247): p. 296-299.
168. Ren, P., C. Wu, and J.W. Ponder, *Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules.* Journal of chemical theory and computation, 2011. **7**(10): p. 3143-3161.
169. Shi, Y., et al., *The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins.* Journal of chemical theory and computation, 2013. **9**(9): p. 4046-4063.
170. Schnieders, M.J. and J.W. Ponder, *Polarizable Atomic Multipole Solutes in a Generalized Kirkwood Continuum.* Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2083-2097.
171. Fried, S.D., et al., *Calculations of the Electric Fields in Liquid Solutions.* The journal of physical chemistry. B, 2013. **117**(50): p. 16236-16248.

172. Hol, W.G.J., *Effects of the α-helix dipole upon the functioning and structure of proteins and peptides.* Advances in Biophysics, 1985. **19**: p. 133-165.
173. Honeywell. *Dielectric Constant Table*. Available from: https://www.honeywellprocess.com/library/marketing/tech-specs/Dielectric%20Constant%20Table.pdf.
174. Richman, D.E., A. Majumdar, and B. Garcia-Moreno E, *Conformational reorganization coupled to the ionization of internal Lys residues in proteins.* Biochemistry, 2015.
175. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited.* Protein Science : A Publication of the Protein Society, 2008. **17**(8): p. 1295-1307.
176. Bhate, Manasi P., et al., *Signal Transduction in Histidine Kinases: Insights from New Structures.* Structure, 2015. **23**(6): p. 981-994.
177. Marshall, C.B., et al., *Calmodulin and STIM proteins: Two major calcium sensors in the cytoplasm and endoplasmic reticulum.* Biochemical and Biophysical Research Communications, 2015. **460**(1): p. 5-21.
178. Kurnik, M., et al., *Folding without charges.* Proc Natl Acad Sci U S A, 2012. **109**(15): p. 5705-10.
179. Street, A.G. and S.L. Mayo, *Intrinsic β-sheet propensities result from van der Waals interactions between side chains and the local backbone.* Proceedings of the National Academy of Sciences, 1999. **96**(16): p. 9074-9076.
180. Ganesan, S.J. and S. Matysiak, *Role of Backbone Dipole Interactions in the Formation of Secondary and Supersecondary Structures of Proteins.* J Chem Theory Comput, 2014. **10**(6): p. 2569-2576.
181. Ganesan, S.J. and S. Matysiak, *Interplay between the hydrophobic effect and dipole interactions in peptide aggregation at interfaces.* Physical Chemistry Chemical Physics, 2016. **18**(4): p. 2449-2458.
182. Moult, J., et al., *A large-scale experiment to assess protein structure prediction methods.* Proteins: Structure, Function, and Bioinformatics, 1995. **23**(3): p. ii-iv.
183. Alvizo, O., B.D. Allen, and S.L. Mayo, *Computational protein design promises to revolutionize protein engineering.* Biotechniques, 2007. **42**(1): p. 31, 33, 35 passim.
184. Pinski, P., et al., *Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals.* The Journal of Chemical Physics, 2015. **143**(3): p. 034108.
185. Riplinger, C., et al., *Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory.* The Journal of Chemical Physics, 2016. **144**(2): p. 024109.
186. Palenik, M.C. and B.I. Dunlap, *Density perturbation theory.* The Journal of Chemical Physics, 2015. **143**(4): p. 044115.
187. Nelson, R., et al., *Structure of the cross-[beta] spine of amyloid-like fibrils.* Nature, 2005. **435**(7043): p. 773-778.
188. Kumar, M.D.S., et al., *ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions.* Nucleic Acids Research, 2006. **34**(suppl 1): p. D204-D206.
189. Malkov, S.N., et al., *A reexamination of the propensities of amino acids towards a particular secondary structure: classification of amino acids based on their chemical structure.* J Mol Model, 2008. **14**(8): p. 769-75.
190. *Practical Aspects of Declarative Languages - 18th International Symposium, PADL 2016, St. Petersburg, FL, USA, January 18-19, 2016. Proceedings*. 2016. Springer.
191. Chellapa, G.D. and G.D. Rose, *Reducing the dimensionality of the protein-folding search problem.* Protein Science : A Publication of the Protein Society, 2012. **21**(8): p. 1231-1240.
192. Street, T.O., et al., *Physical-chemical determinants of turn conformations in globular proteins.* Protein Science : A Publication of the Protein Society, 2007. **16**(8): p. 1720-1727.
193. Niklasson, A.M.N., et al. *Graph-based linear scaling electronic structure theory*. ArXiv e-prints, 2016. arXiv:1603.00937.
194. Weinhold, F. and R.A. Klein, *Anti-Electrostatic Hydrogen Bonds.* Angewandte Chemie International Edition, 2014. **53**(42): p. 11214-11217.

195. Weinhold, F., *Nature of H-bonding in clusters, liquids, and enzymes: an ab initio, natural bond orbital perspective.* Journal of Molecular Structure: THEOCHEM, 1997. **398–399**(0): p. 181-197.

196. Knowles, J.R., *Enzyme-Catalyzed Phosphoryl Transfer Reactions.* Annual Review of Biochemistry, 1980. **49**(1): p. 877-919.

197. Mentel, Ł.M. and E.J. Baerends, *Can the Counterpoise Correction for Basis Set Superposition Effect Be Justified?* Journal of Chemical Theory and Computation, 2014. **10**(1): p. 252-267.

198. Jansen, H.B. and P. Ros, *Non-empirical molecular orbital calculations on the protonation of carbon monoxide.* Chemical Physics Letters, 1969. **3**(3): p. 140-143.

199. Liu, B. and A.D. McLean, *Accurate calculation of the attractive interaction of two ground state helium atoms.* The Journal of Chemical Physics, 1973. **59**(8): p. 4557-4558.

200. Cousot, P., *Formal verification by abstract interpretation*, in *Proceedings of the 4th international conference on NASA Formal Methods*. 2012, Springer-Verlag: Norfolk, VA. p. 3-7.

201. *Why Haskell just works*. 2011; Available from: https://wiki.haskell.org/Why_Haskell_just_works.

202. Backus, J., *Can programming be liberated from the von Neumann style?: a functional style and its algebra of programs.* Commun. ACM, 1978. **21**(8): p. 613-641.

203. *List comprehension*. 2007; Available from: https://wiki.haskell.org/List_comprehension.

204. Baker, E.N. and R.E. Hubbard, *Hydrogen bonding in globular proteins.* Progress in Biophysics and Molecular Biology, 1984. **44**(2): p. 97-179.

205. Clauss, A.D., et al., *Rabbit-ears hybrids, VSEPR sterics, and other orbital anachronisms.* Chemistry Education Research and Practice, 2014. **15**(4): p. 417-434.

206. Palmer, R.J., *Polyamides, Plastics*, in *Encyclopedia of Polymer Science and Technology*. 2002, John Wiley & Sons, Inc.

207. Seymour, R.B. and G.B. Kauffman, *Polyurethanes: A class of modern versatile materials.* Journal of Chemical Education, 1992. **69**(11): p. 909.

208. Simic, V., L. Bouteiller, and M. Jalabert, *Highly Cooperative Formation of Bis-Urea Based Supramolecular Polymers.* Journal of the American Chemical Society, 2003. **125**(43): p. 13148-13154.

209. De Jong, K.A., *Evolutionary computation : a unified approach*. 2006: MIT Press.

210. Cordier, P., et al., *Self-healing and thermoreversible rubber from supramolecular assembly.* Nature, 2008. **451**(7181): p. 977-980.

211. Yang, G., et al., *Graphene-like two-dimensional layered nanomaterials: applications in biosensors and nanomedicine.* Nanoscale, 2015. **7**(34): p. 14217-14231.

212. Huang, P.-S., et al., *De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy.* Nat Chem Biol, 2016. **12**(1): p. 29-34.

213. Seebach, D., et al., *β-Peptides: Synthesis by Arndt-Eistert homologation with concomitant peptide coupling. Structure determination by NMR and CD spectroscopy and by X-ray crystallography. Helical secondary structure of a β-hexapeptide in solution and its stability towards pepsin.* Helvetica Chimica Acta, 1996. **79**(4): p. 913-941.

214. Sharley, J.N. *Variation of protein backbone amide resonance by electrostatic field*. ArXiv e-prints, 2015. arXiv:1512.05488.

215. Sharley, J.N. *Do cooperative cycles of hydrogen bonding exist in proteins?* ArXiv e-prints, 2016. arXiv:1601.01792.

216. Algaer, E.A. and N.F.A. van der Vegt, *Hofmeister Ion Interactions with Model Amide Compounds.* The Journal of Physical Chemistry B, 2011. **115**(46): p. 13781-13787.

217. Hofmeister, F., Arch. Exp. Pathol. Pharmakol., 1888. **24**: p. 247-260.

218. Okur, H.I., J. Kherb, and P.S. Cremer, *Cations Bind Only Weakly to Amides in Aqueous Solutions.* Journal of the American Chemical Society, 2013. **135**(13): p. 5062-5067.

219. Kitaura, K., et al., *Fragment molecular orbital method: an approximate computational method for large molecules.* Chemical Physics Letters, 1999. **313**(3–4): p. 701-706.

220. Dongarra, J., et al., *Accelerating Numerical Dense Linear Algebra Calculations with GPUs*, in *Numerical Computations with GPUs*, V. Kindratenko, Editor. 2014, Springer International Publishing: Cham. p. 3-28.

221. Greg, L., et al., *Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules.* Journal of Physics: Condensed Matter, 2013. **25**(15): p. 152101.

222. Froese Fischer, C., *General Hartree-Fock program.* Computer Physics Communications, 1987. **43**(3): p. 355-365.

223. Isborn, C.M., et al., *The Charge Transfer Problem in Density Functional Theory Calculations of Aqueously Solvated Molecules.* The Journal of Physical Chemistry B, 2013. **117**(40): p. 12189-12201.

224. Baer, R., E. Livshits, and U. Salzner, *Tuned Range-Separated Hybrids in Density Functional Theory.* Annual Review of Physical Chemistry, 2010. **61**(1): p. 85-109.

225. Ong, M.T., et al., *First Principles Dynamics and Minimum Energy Pathways for Mechanochemical Ring Opening of Cyclobutene.* Journal of the American Chemical Society, 2009. **131**(18): p. 6377-6379.

226. *TeraChem User's Guide Version 1.5K*. PetaChem LLC.

227. Bauzá, A., et al., *Is the Use of Diffuse Functions Essential for the Properly Description of Noncovalent Interactions Involving Anions?* The Journal of Physical Chemistry A, 2013. **117**(12): p. 2651-2655.

228. Crevenna, Alvaro H., et al., *Effects of Hofmeister Ions on the α-Helical Structure of Proteins.* Biophysical Journal, 2012. **102**(4): p. 907-915.

229. Lo Nostro, P. and B.W. Ninham, *Hofmeister Phenomena: An Update on Ion Specificity in Biology.* Chemical Reviews, 2012. **112**(4): p. 2286-2322.

230. Light, T.P., et al., *Hofmeister Ion-Induced Changes in Water Structure Correlate with Changes in Solvation of an Aggregated Protein Complex.* Langmuir, 2016. **32**(5): p. 1360-1369.

231. Shi, J. and J. Wang, *Interaction between Metal Cation and Unnatural Peptide Backbone Mediated by Polarized Water Molecules: Study of Infrared Spectroscopy and Computations.* The Journal of Physical Chemistry B, 2014. **118**(43): p. 12336-12347.

232. Yue, K. and K.A. Dill, *Inverse protein folding problem: designing polymer sequences.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(9): p. 4163-4167.

233. Corey, R.B. and L. Pauling, *Fundamental Dimensions of Polypeptide Chains.* Proceedings of the Royal Society of London. Series B, Biological Sciences, 1953. **141**(902): p. 10-20.

234. Papachristodoulou, D., et al., *Biochemistry and Molecular Biology*. 5 ed. 2014: OUP.

235. Light, T.P., et al., *Hofmeister Ion-Induced Changes in Water Structure Correlate with Changes in Solvation of an Aggregated Protein Complex.* Langmuir, 2016.