ABSTRACT

Title of dissertation: DENSITY FUNCTIONAL CALCULATIONS OF BACKBONE ^{15}N CHEMICAL SHIELDINGS IN PEPTIDES AND PROTEINS

Ling Cai, Doctor of Philosophy, 2009

Dissertation directed by: Professors David Fushman and Daniel S. Kosov
Department of Chemistry and Biochemistry,
Institute for Physical Science and Technology

In this dissertation, we describe computational and theoretical study of backbone ^{15}N chemical shieldings in peptides and proteins. Comprehensive density functional calculations have been performed on systems of different complexity, ranging from model dipeptides to real proteins and protein complexes.

We begin with examining the effects of solvation, hydrogen bonding, backbone conformation, and the side chain identity on ^{15}N chemical shielding in proteins by density functional calculations. N-methylacetamide (NMA) and N-formyl-alanyl-X (with X being one of the 19 naturally occurring amino acids excluding proline) were used as model systems for this purpose. The conducting polarizable continuum model was employed to include the effect of solvent in the calculations. We show that the augmentation of the polarizable continuum model with the explicit water molecules in the first solvation shell has a significant influence on isotropic ^{15}N chemical shift but not as much on the chemical shift anisotropy. The difference in the isotropic chemical shift between the standard β-sheet and standard α-helical

conformations ranges from 0.8 ppm to 6.2 ppm depending on the residue type, with the mean of 2.7 ppm. This is in good agreement with the experimental chemical shifts averaged over a database of 36 proteins containing >6100 amino acid residues. The orientation of the ^{15}N chemical shielding tensor as well as its anisotropy and asymmetry are also in the range of values experimentally observed for peptides and proteins.

Having applied density functional calculation successfully to model peptides, we develop a computationally efficient methodology to include most of the important effects in the calculation of chemical shieldings of backbone ^{15}N in a protein. We present the application to selected α-helical and β-sheet residues of protein G. The role of long-range intra-protein electrostatic interactions by comparing models with different complexity in vacuum and in charge field is analyzed. We show that the dipole moment of the α-helix can cause significant deshielding of ^{15}N ; therefore, it needs to be considered when calculating ^{15}N chemical shielding. We emphasize the importance of including interactions with the side chains that are close in space when the charged form for ionizable side chains is adopted in the calculation. We also illustrate how the ionization state of these side chains can affect the chemical shielding tensor elements. For α-helical residues, chemical shielding calculations using a 8-residue fragment model in vacuum and adopting the charged form of ionizable side chains yield a generally good agreement with experimental data. We also performed computational modeling of the chemical shift perturbations occurring upon protein-protein or protein-ligand binding. We show that the chemical shift perturbations in ubiquitin upon dimer formation can be explained

qualitatively through computation.

This dissertation hence demonstrates that quantum chemical calculations can be successfully used to obtain a fundamental understanding of the relationship between chemical shielding and the surrounding protein environment for the elusive case of ^{15}N and therefore enhance the role of ^{15}N chemical shift measurements in the analysis of protein structure and dynamics.

DENSITY FUNCTIONAL CALCULATIONS OF BACKBONE ^{15}N CHEMICAL SHIELDINGS IN PEPTIDES AND PROTEINS

by

Ling Cai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor David Fushman, Chair/Advisor
Professor Daniel S. Kosov, Advisor
Professor John Tossell
Professor Kwaku T. Dayie
Professor Jeffery T. Davis (Dean's Representative)

Dedication

In memory of my grandma, whose great strength I have always admired.

Acknowledgments

First of all, I would like to express my deepest gratitude towards both of my advisors, Dr. David Fushman and Dr. Daniel S. Kosov. By teaching the interesting classes in NMR techniques and computational chemistry that truly engaged me, they led me to the fascinating field where this dissertation topic was born. Throughout the course of this study, besides being resourceful, supportive and understanding, they were always capable of infusing me with enthusiasm and optimism, which kept my spirit high especially during the frustrating times. This and the training they provided have made me a better researcher. I must also acknowledge the committee members for their precious time and insightful comments, and Dr. Millard H. Alexander and Dr. Doug English for their helps.

I would like to thank my friends from Fushman and Kosov groups, for providing me with useful discussions; especially I would like to thank Dr. Zhenyu Li, Dr. Daoning Zhang and Dr. Maxim Gelin, who generously shared with me their expertise and research experiences, as did Dr. Jacek Klos, Dr. Xianjie Liu and Dr. Shunliu Deng from other research groups here, to whom I owe my thanks as well. Thanks also go to my officemates Merle Zimmermann and Jordan Horowitz for sharing computer tips.

The Chemical Physics Program has been a great place for me. I would like to thank Dr. Michael Coplan, our director, who has supported me at every step of the way and who has made my graduate life much more smoothier. Thanks are also due to Debbie Jenkins for her effective administration. And I must thank Diane

Mancuso, who helped me during my first year at UMCP and whose genuine kindness will be missed. Of course this must not go without mentioning my Chemical Physics friends, who have made my graduate life much more fun and enjoyable!

Finally I would like to thank my family and other friends for their support and love over the years. Most of all, I would like to thank my husband Xinhui Zhou, who always believes in me and who is always there for me when I need him, for his love, faith, patience and perspective.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ASC	Apparent surface charge
CSP	Chemical shift perturbation
CFP	Charge field perturbation
CPCM	Conducting polarizable continuum model
CSA	Chemical shift anisotropy
DFT	Density-functional theory
GB1	The B1 immunoglobulin-binding domain of streptococcal protein G
GB3	The B3 immunoglobulin-binding domain of streptococcal protein G
GIAO	Gauge invariant/including atomic orbitals
HB	Hydrogen bonding
HSQC	Heteronuclear single quantum coherence
NMA	N-methylacetamide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
PDB	Protein data bank
QM	Quantum mechanical
SCRF	Self-consistent reaction field
UA0	United atom topological model

Chapter 1

Introduction

1.1 Scope of the present work

Chemical shielding reflects electronic environment of nuclei under observation and therefore contains important information about molecular structure and conformational dynamics. Detailed understanding of the sources of various contributions to chemical shielding in proteins is not only critical for our ability to predict chemical shifts and thus facilitate NMR signal assignment, but also potentially important for improvement in structure characterization of proteins (Cornilescu et al., 1999; Lipsitz and Tjandra, 2003; Shen and Bax, 2007; Shen et al., 2008, 2009) and analysis of protein dynamics (Hall and Fushman 2006).

The past several decades of research have led to the development of accurate theoretical methods and computational schemes for chemical shift calculations in peptides and proteins (reviewed in (Shen and Bax 2007)). The computational methods range from those that base on empirical shielding surface (Wishart and Nip, 1998) or sequence homology (Shen and Bax, 2007; Wishart et al., 1997) to those that use ab initio quantum mechanical (QM) calculations (de Dios et al., 1993b; Oldfield, 1995; Xu and Case, 2001, 2002). The first-principles QM calculation is very accurate and has played an important role in understanding the relative importance of various effects, such as backbone and side chain conformation, hy-

drogen bonding, and protein electrostatics, on the chemical shielding of nuclei in proteins. These calculations have also provided parameters and conceptual ideas for the development of phenomenological models (Xu and Case 2001; Xu and Case 2002). Despite their contributions to our understanding of the chemical shielding phenomena and to model-building efforts, first-principles quantum chemical calculation has not become a routine practice in predicting chemical shifts in proteins because of several difficulties. First, chemical shifts could be influenced by a multitude of factors. In general, the backbone torsion angles (Figure 1.1), the side chain orientation, the hydrogen bonding and direct interactions with nearby side chains, the long-range intra-protein electrostatics, and solvent effect can all contribute. It is challenging to include all foreseeable effects and still be computationally efficient. Second, any experimental protein structure we use for calculating chemical shifts might require some geometry optimization in order to get an accurate prediction. This is particularly true for α-carbons because their calculated chemical shielding is very sensitive to bond lengths and bond angles (de Dios et al., 1993a). Third, proteins rapidly sample the available conformational space, hence no single snapshot (X-ray) or even a finite bundle of NMR structures of a given protein can fully represent the conformational ensemble (and dynamic ensemble averaging) for which the experimental chemical shifts are measured. Moreover, when a solution structure is chosen for calculations, one has to consider not only the structural ensemble but also at what conditions (pH, temperature, etc) the structure was obtained. There can be fluctuations which may not only affect some structural parameters but also alter the ionization form of the side chains of Asp, Glu, Arg, and Lys (Vila and Scheraga

2007). Fourth, there are systematic errors associated with the level of theory used and the basis set chosen for calculating the electronic structure. In practice, an offset may exist that will need to be corrected if an absolute isotropic chemical shift prediction is the aim (Cai et al. 2008). An additional complication here is that this correction should differ between various residue types (Xu and Case, 2001).

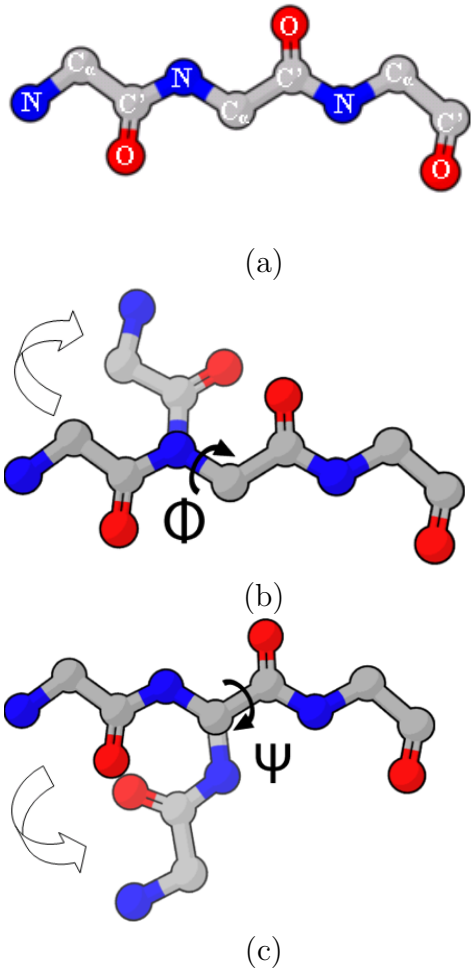


Figure 1.1: Illustration of protein backbone dihedral angles ϕ and ψ (Anonymous, June 1, 2001). Shown here is the backbone of a protein fragment in its fully extended conformation, with $\phi = 180°$ and $\psi = 180°$ (a); and the rotation of the first peptide plane along the N-C_α bond, characterized by $\phi = 0°$ (b), and along the C_α-C' bond, characterized by $\psi = 0°$ (c).

Despite these difficulties, theoretical chemical shift tensors are valuable in pep-

tide and protein structure validation and refinement (Wylie et al., 2009) and hence worth exploring. In this dissertation, we chose to focus on amide ^{15}N chemical shielding tensor for several reasons: (i) it is influenced by numerous factors and, therefore, is arguably the most difficult case for computational prediction (de Dios et al., 1993b); (ii) perturbations in amide chemical shifts are often used for mapping protein-ligand interactions (Zuiderg 2002); and (iii) amide ^{15}N chemical shift tensor contributes to ^{15}N spin relaxation rates that are widely used for analysis of protein dynamics (Fushman and Cowburn 2001; Hall and Fushman 2006). Thus, first-principles calculations may shed light onto the relationship between structural changes in proteins upon ligand binding and the accompanying changes in their chemical shifts. The quantum chemical calculations can also provide an understanding of the structural basis for site-to-site variability in ^{15}N chemical shift tensors (Fushman et al. 1998; Hall and Fushman 2006).

Previous theoretical works (de Dios et al. 1993b; Le and Oldfield 1996; Xu and Case 2002; Cai et al. 2008) have established that ^{15}N chemical shielding depends on the following factors with none seemingly dominating: ϕ, ψ, χ_1, preceding side chains identity and conformation, hydrogen bonding partners, electrostatic interactions, and solvent effect. To our knowledge, bulk solvent effects have not been considered in chemical shielding calculations for proteins. This is part of the endeavor in our present work. In addition, although short-range interaction with hydrogen-bonding partners can be included explicitly and exactly, model treatment has been required for long-range electrostatic interactions. Point charge representation can be an option here. For example, in addition to the main fragment, charges can be

incorporated for the remaining atoms in the protein using some charge set. Thus, it was found that if the charge field perturbation (CFP) effects are included in the calculation of carbon chemical shielding in amino acids, the correlation between the theory and experiment is slightly improved (de Dios et al., 1994). The improvement is mainly for the sp^2 carbonyl carbon, which is more sensitive to electrostatic field effects. The sp^2-hybridized amide nitrogen shares this kind of sensitivity (Bader 2009) and showed some prediction improvement with inclusion of charge field perturbation in a dipeptide model (de Dios et al. 1993b). However, a later application to helical residues (Le and Oldfield 1996) indicated that a static charge field is inadequate to account for the long-range electrostatic field contribution to ^{15}N shielding in an α-helix. There are probably two reasons for this: (i) each type of residue has a fixed set of approximated charges, which may not be accurate enough for the purpose of calculating ^{15}N chemical shifts because the multiple-bond character of the peptide group makes the peptide nitrogen highly polarizable and hence very sensitive to the electric field that the charges generate; and (ii) the ionization state of Asp, Glu, Arg, and Lys is not well determined and can vary depending on the protein ensemble, temperature, and pH; this renders representation of the side chains of these residues as either neutral or charged somewhat arbitrary.

In this dissertation, we report our findings of the effect of solvent and electrostatic interactions through density functional calculations. In the rest of this chapter, we present the theoretical background which includes (i) quantum mechanical treatment of chemical shielding phenomena and discussion of how it is defined and computed theoretically and (ii) the method of treating solvent as a contin-

uum. In Ch. 2, solvent effect in the calculation of ^{15}N chemical shielding is modeled through two systems: NMA and Ala-X. Then the calculation of amide ^{15}N chemical shielding in a real protein, GB3, is explored in considering electrostatic effect. This calculation was performed for both classical α-helical residues (presented in Ch. 3) and β-sheet residues (presented in Ch. 4). Finally in Ch. 5, we report a preliminary computational study of ^{15}N isotropic chemical shielding sensitivity in open and closed conformation in the protein complex diubiquitin.

1.2 Theory and methods

1.2.1 Chemical shielding

Different groups of nuclei in a molecule have resonance frequencies that reflect the fact that they experience different local magnetic fields, $\mathbf{H}_{loc}$. The following shows how molecular currents induced by the external field modify the local field and hence give rise to the chemical shift (Atkins and Friedman, 1997). Let us consider a molecule containing a single magnetic nucleus B (and any number of other nuclei). The classical expression for the magnetic filed generated by a magnetic dipole is

$$\mathbf{H}_{nuc} = -\frac{\mu_0}{4\pi r^3}\{\boldsymbol{\mu} - \frac{3\mathbf{r}(\mathbf{r}\cdot\boldsymbol{\mu})}{r^2}\} \tag{1.1}$$

where $\mathbf{H}_{nuc}$ is the magnetic field generated by the nucleus B with a magnetic moment $\boldsymbol{\mu}$; $\mathbf{r}$ and r are distance vector and distance from the nucleus respectively; and μ_0 is the vacuum permeability. The corresponding vector potential is

$$\mathbf{A}_{nuc} = \left(\frac{\mu_0}{4\pi r^3}\right)\boldsymbol{\mu}\times\mathbf{r}. \tag{1.2}$$

The magnetic moment $\boldsymbol{\mu}$ of a nucleus is related to its spin angular momentum $\mathbf{I}$ by $\boldsymbol{\mu} = \gamma_N \mathbf{I}$, where γ_N is gyromagnetic ratio, therefore, the vector potential can also be written as

$$\mathbf{A}_{nuc} = \left(\frac{\mu_0 \gamma_N}{4\pi r^3}\right) \mathbf{I} \times \mathbf{r}. \tag{1.3}$$

Since the Hamiltonian for the molecule in a magnetic field is constructed in the standard way by replacing momentum $\mathbf{p}$ by $\boldsymbol{\pi} = \mathbf{p} - \frac{e}{c}\mathbf{A}_{nuc}$, the Hamiltonian is now written as

$$H = H^{(0)} + H^{(1)} + H^{(2)}, \tag{1.4}$$

with

$$H^{(0)} = \frac{\mathbf{p}^2}{2m_e} + \mathbf{V}, \tag{1.5}$$

$$H^{(1)} = (-e/2m_e c)(\mathbf{p} \cdot \mathbf{A}_{nuc} + \mathbf{A}_{nuc} \cdot \mathbf{p}), \tag{1.6}$$

$$H^{(2)} = (e^2/2m_e c^2) A_{nuc}^2. \tag{1.7}$$

We now focus on the contribution to energy from the first-order term $H^{(1)}$. The first-order perturbation energy can be written as

$$E^{(1)} = \langle 0|H^{(1)}|0\rangle = (-e/2m_e c) \int \mathbf{A}_{nuc} \cdot (\psi^* \mathbf{p} \psi + \psi \mathbf{p}^* \psi^*) d\tau = -\int \mathbf{A}_{nuc} \cdot \mathbf{j}_0 d\tau, \tag{1.8}$$

with $\mathbf{j}_0$ being the current density. When the external field is applied, $\mathbf{p}$ in the above expression needs to be replaced by $\mathbf{p} - \frac{e}{c}\mathbf{A}_{\mathrm{ex}}$, resulting in the conversion of $\mathbf{j}_0$ to $\mathbf{j}$, the current density in the presence of the applied field. Then

$$E^{(1)} = -\int \mathbf{A}_{nuc} \cdot \mathbf{j} d\tau. \tag{1.9}$$

This result shows very plainly the contribution to energy by the interaction of a magnetic nucleus dipole moment (through the vector potential) with the electron

currents which may have been induced by an external magnetic field. Inserting the explicit form of the nuclear vector potential we will get

$$E^{(1)} = -\frac{\mu_0 \gamma_N}{4\pi} \int \frac{(\mathbf{I} \times \mathbf{r}) \cdot \mathbf{j}}{r^3} d\tau = -\frac{\mu_0 \gamma_N}{4\pi} \mathbf{I} \cdot \int \frac{\mathbf{r} \times \mathbf{j}}{r^3} d\tau. \tag{1.10}$$

Since the energy of a magnetic dipole in a magnetic field $\mathbf{H}$ is $\boldsymbol{\mu} \cdot \mathbf{H}$, we can interpret this energy as the interaction of a nuclear dipole $\gamma_N \mathbf{I}$ with a local magnetic field, which is solved to be

$$\mathbf{H}_{loc} = \frac{\mu_0}{4\pi} \int \frac{\mathbf{r} \times \mathbf{j}}{r^3} d\tau. \tag{1.11}$$

An alternative way of expressing the total energy expanded as Taylor series about the zero-field value is given by Ditchfield (1974):

$$E = E^{(0)} - \sum_{\alpha} \gamma_\alpha \mathrm{H}_\alpha - \sum_{\alpha} \mu_{B\alpha} \mathrm{H}_\alpha - \frac{1}{2} \sum_{\alpha} \sum_{\beta} \chi_{\alpha\beta} \mathrm{H}_\beta \mathrm{H}_\alpha + \sum_{\alpha} \sum_{\beta} \mu_{B\alpha} \sigma_{B\alpha\beta} \mathrm{H}_\beta + \ldots, \tag{1.12}$$

where α and β indicate Cartesian coordinates x, y or z, and γ_α is a component of the permanent magnetic moment of the molecule. The permanent magnetic moment is zero for singlet state molecules with zero orbital angular momentum. The third term represents the direct interaction between the nuclear magnetic moment and the external magnetic field. The fourth term describes the diamagnetic polarization of the molecule; the total magnetic moment (in direction α) associated with the electronic currents induced by the external magnetic field is $\sum_\beta \chi_{\alpha\beta} H_\beta$, where $\chi_{\alpha\beta}$ is a component of the molecular diamagnetic susceptibility tensor $\boldsymbol{\chi}$. The secondary magnetic field (in direction α) at nucleus B due to these electronic currents is $\sum_\beta \sigma_{B\alpha\beta} H_\beta$ where $\sigma_{B\alpha\beta}$ is a component of the magnetic shielding tensor $\boldsymbol{\sigma}_B$. Thus the total magnetic field experienced by the nucleus B that determines its NMR frequency is

given by

$$\mathbf{H}_{loc} = (1 - \boldsymbol{\sigma}_B)\mathbf{H}. \tag{1.13}$$

1.2.2 Gauge invariant/including atomic orbitals

A common difficulty in the calculation of magnetic properties is that the usual wave functions do not guarantee gauge invariance. One approach to overcome this problem is to use very large basis sets (applicable for small molecules) since in the limit of complete basis sets the results should be gauge independent (Wolinski et al., 1990). For small- and medium-size systems, gauge invariant/including atomic orbitals (GIAO), first adapted by Ditchfield (1974), have been successful. In this method, the basis functions are explicitly dependent on the magnetic field by inclusion of a complex phase factor referring to the position of the basis function (usually the position of the nucleus):

$$\chi_\mu = \phi_\mu e^{-\frac{ie}{\hbar c}\mathbf{A}_\mu \cdot \mathbf{r}}. \tag{1.14}$$

with

$$\mathbf{A}_\mu = \frac{1}{2}\mathbf{H} \times (\mathbf{R}_\mu - \mathbf{R}_G). \tag{1.15}$$

Here, ϕ_μ is the atomic orbital with the nuclear position $\mathbf{R}_\mu$, $\mathbf{R}_G$ is the origin of the Coulomb gauge used, and so $\mathbf{A}_\mu$ is the value of the vector potential at the nucleus position $\mathbf{R}_\mu$. The effect is that the matrix elements involving GIAOs only contain a difference in vector potentials, thereby removing the reference to an absolute gauge origin. For the overlap and potential energy it is straightforward to see that matrix elements become independent of the gauge origin because the term $\mathbf{A}_\mu - \mathbf{A}_v$ does

not depend on $\mathbf{R}_G$:

$$\langle \chi_\mu | \chi_v \rangle = \langle \phi_\mu | e^{\frac{ie}{\hbar c}(\mathbf{A}_\mu - \mathbf{A}_v)\mathbf{r}} | \phi_v \rangle \tag{1.16}$$

$$\langle \chi_\mu | \mathbf{V} | \chi_v \rangle = \langle \phi_\mu | e^{\frac{ie}{\hbar c}(\mathbf{A}_\mu - \mathbf{A}_v)\mathbf{r}} \mathbf{V} | \phi_v \rangle \tag{1.17}$$

$$\mathbf{A}_\mu - \mathbf{A}_v = \frac{1}{2}\mathbf{H} \times (\mathbf{R}_\mu - \mathbf{R}_v) \tag{1.18}$$

The kinetic energy is slightly more complicated, but it can be shown that the following relation holds so it does not dependent on $\mathbf{R}_G$ either:

$$\begin{aligned} \langle \chi_\mu | \boldsymbol{\pi}^2 | \chi_v \rangle &= \langle \chi_\mu | (p - \frac{e}{c}\frac{1}{2}\mathbf{H} \times (\mathbf{r} - \mathbf{R}_G))^2 | \chi_v \rangle \\ &= \langle \phi_\mu | e^{\frac{ie}{\hbar c}(\mathbf{A}_\mu - \mathbf{A}_\mu)\mathbf{r}} (p - \frac{e}{2c}\mathbf{H} \times (\mathbf{r} - \mathbf{R}_v))^2 | \phi_v \rangle \end{aligned} \tag{1.19}$$

The use of GIAOs as basis functions makes all matrix elements, and therefore all properties, independent of the gauge origin (Jensen, 2007).

1.2.3 Self-consistent reaction field (SCRF) model

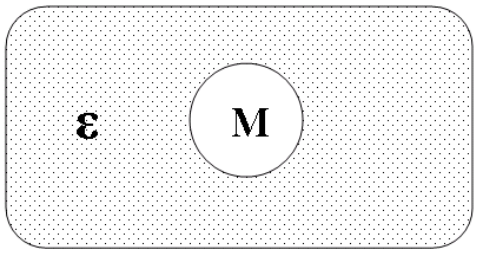


Figure 1.2: Reaction field model. Molecule M is placed in a cavity in the solvent, which is treated as a polarizable continuum with a dielectric constant ϵ

SCRF model belongs to the family of continuum solvation models that treat the solvent environment as a "reaction field" without considering the individual

solvent molecules explicitly. It considers the solvent as a uniform polarizable medium with a dielectric constant of ϵ, with the solute M placed in a suitably shaped hole, also called cavity, in the medium (Figure 1.2) (Cossi et al., 2003). The charge distribution of the solute polarizes the solvent producing a reaction potential. Very often, the apparent surface charge (ASC) approach is used so that the reaction potential V_σ, everywhere in the space, can be described in terms of an apparent charge distribution $\sigma(\mathbf{r}_s)$ spread on the cavity surface.

$$V_\sigma(\mathbf{r}) = \oint \frac{\sigma(\mathbf{r}_s)}{|\mathbf{r} - \mathbf{r}_s|} d\mathbf{r}_s. \tag{1.20}$$

It can be shown from Gauss's law at an electric field discontinuity that

$$4\pi\epsilon\sigma(\mathbf{r}_s) = (\epsilon - 1)\frac{\partial(V_M + V_\sigma)_{in}}{\partial \mathbf{n}}, \tag{1.21}$$

where V_M is the electric potential generated by the solute charge distribution; $\mathbf{n}$ is the unit vector perpendicular to the cavity surface and pointing outward and "*in*" designates the potential inside the cavity (Tomasi et al., 2005). Once $\sigma(\mathbf{r}_s)$ is determined, the associated potential V_σ is added to the Hamiltonian operator

$$H = H_0 + V_\sigma, \tag{1.22}$$

where H_0 is the Hamiltonian without the presence of solvent. As we can see, the potential V_σ from the surface charge is obtained by knowing the molecular charge distribution through V_M but also enters the Hamiltonian and thus influences the molecular wave function, which determines V_M. The procedure is therefore iterative, leading to self-consistence between the solute wave function and the solvent polarization.

Specifically, in the conducting polarizable continuum model (CPCM) implemented in GAUSSIAN03 (Barone and Cossi, 1998; Cossi et al., 2003), the apparent polarization charges distributed on the cavity surface are determined by imposing the condition that the total electrostatic potential goes to zero on the surface. This boundary condition, suited for cavities in the conducting media, can describe the solvation in polar liquids. It is computationally simpler, especially for the expression of the energy gradients, which can be efficient to allow geometry optimizations in solution.

Chapter 2

Solvent modeling

To our knowledge, bulk solvent effects have not been considered in chemical shielding calculations for proteins. It is acknowledged that since solution NMR measurements in proteins are typically conducted in the presence of water, a highly polar solvent, the effects of bulk solvent molecules can be crucial. To study solvent effects on chemical shifts, one can utilize continuum models (discussed in detail in the previous chapter). A recent review (Tomasi et al., 2005) points out two aspects important for the applications of continuum models to QM studies of chemical shifts. The first aspect is the perturbation effect of the solvent on the electronic wave function of the solute and the geometric distortion of the solute molecule. The second aspect concerns the importance of both short-range and long-range solute-solvent interactions in determining the solvent effect on the nuclear shieldings. The current concept is that short-range interaction can be effectively handled by a number of explicitly treated solvent molecules from the first solvation shell, while the long-range effects can be described effectively by continuum methods. Tomasi et al. also noted that because the characteristic time scale in NMR spectroscopy is milliseconds and longer, whenever explicit solvent molecules are used, it is necessary to correctly account for the statistical picture inherent in the dynamic nature of the solvation shell (picosecond time scale). Recent years have seen continuum methods being ap-

plied to study the solvent effect on chemical shifts of small molecules (Aidas et al., 2007; Mennucci and Martinez, 2005; Mennucci et al., 2001). For example, Mennucci and Martinez (2005) compared continuum-only description, discrete description in terms of solute-solvent clusters, and mixed discrete/continuum description in order to identify and characterize different aspects of solvation.

In this chapter we report density-functional theory calculations for N-methylacetamide (NMA) to analyze in detail the specific and bulk effects of the solvent water on ^{15}N chemical shielding. Figure 2.1(a) shows the chemical structure of NMA. We then apply the continuum model in both structure optimization and chemical shielding calculations of N-formyl-alanyl-X amides, where X is one of the 19 naturally occurring amino acids excluding proline. For each compound, calculations were carried out for two backbone conformations, corresponding to a standard α-helix ($\phi = -58°$, $\psi = -47°$) and a standard β-sheet ($\phi = -139°$, $\psi = 135°$) respectively, with the latter illustrated for N-formyl-alanyl-Ala in Figure 2.1(b). The results are compared with the available experimental data as well as with the previous approach that does not account for solvent effects. We also compare the effect of polarizable continuum and structure-related hydrogen bonding on ^{15}N chemical shielding for several residues from protein GB3.

2.1 Computational details

All calculations reported in this dissertation were performed using the GAUSSIAN03 suite of programs (Frisch et al., 2004). We use density-functional theory

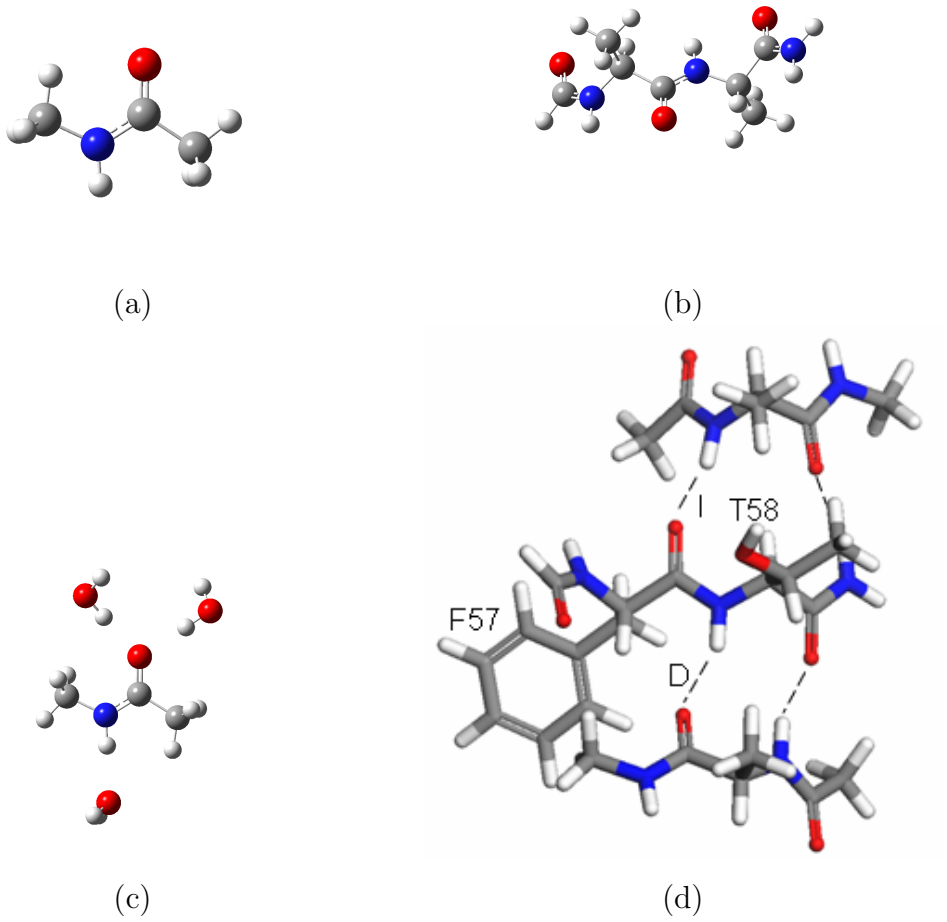


Figure 2.1: Molecules considered in this study: (a) N-methylacetamide (NMA); (b) N-formyl-alanyl-X amide (where X is alanine); (c) N-methylacetamide with three water molecules from the first coordination shell; (d) Illustration of the structural model of a β-sheet fragment (corresponding to Thr58 in 1IGD (Derrick and Wigley, 1994)) together with direct (D) and indirect (I) hydrogen-bonding partners, as well as other hydrogen bonds in this cluster included in the ^{15}N (Thr58) calculation.

(DFT) with three-parameter Becke-Lee-Yang-Parr (B3LYP) exchange-correlation functional (Becke, 1988, 1993; Lee et al., 1988).

In this chapter where the solvent effect is considered, the solvent was taken into account approximately, by employing the conducting polarized continuum model (CPCM) (Barone and Cossi, 1998; Cossi et al., 2003). In this model, the solute

molecule is placed into a cavity surrounded by the solvent considered as a continuum medium with certain dielectric constant. The charge distribution of the solute polarizes the dielectric medium, which generates surface charges around the cavity and hence in turn polarizes the solute. In our calculations, the dielectric constant of water (78.39) was used and the cavity was chosen to be built up by the simple united atom topological model (UA0), in which the van der Waals surface was built by placing a sphere around each solute heavy atom while hydrogen atoms were enclosed in the sphere of the atom to which they are bonded. The number of surface elements for each sphere was 60, and an area of 0.2 Å^2 was set for each surface element. Gauge-invariant atomic orbitals (GIAO) were employed to compute NMR properties (Ditchfield, 1974; Wolinski et al., 1990) as implemented in Gaussian03. We used the 6-311+G(2d,p) basis set for all our calculations except for the geometry optimization of the dipeptides which was performed with the 6-31+G(d) basis set.

2.2 Results and discussion

2.2.1 NMA calculations: disentangling solvent contributions

NMA serves as a simple model representing the amide linkage in proteins. It enables us to eliminate the conformational complexity of peptides and to concentrate only on the solvent's influence on ^{15}N chemical shielding. To this end we performed a series of calculations for gas phase, continuum-only model, and cluster-continuum model. Figure 2.1(c) shows NMA with the first water coordination shell represented by three water molecules that make direct hydrogen bonds with NMA. The param-

eters used for these models are shown in Table 2.1, and the results are presented in Table 2.2.

Table 2.1: Bond lengths (in Å) used for calculations in gas phase (in vacuo), continuum-only, and cluster/continuum models for N-methylacetamide

	NMA_vac[1]	NMA_cont[2]	$NMA + 3w$[3]
NH bond	1.00	1.02	1.02
CN bond	1.36	1.35	1.33
CO bond	1.22	1.24	1.25

1. NMA_vac refers to the NMA's molecule structure optimized in vacuo.
2. NMA_cont refers to NMA's molecular structure optimized in continuum-only model.
3. $NMA + 3w$ refers to the optimized structure of NMA and three water molecules from the first coordination shell within the polarized continuum model.

To elucidate the role of various solvent effects we define the following differences of the various chemical shieldings:

$$\triangle\sigma^N_{geom} = \sigma^N_{NMA^*/vac} - \sigma^N_{NMA\_vac/vac} \quad (2.1)$$

$$\triangle\sigma^N_{solv} = \sigma^N_{NMA+3w/cont} - \sigma^N_{NMA^*/vac} \quad (2.2)$$

$$\triangle\sigma^N_{total} = \sigma^N_{NMA+3w/cont} - \sigma^N_{NMA\_vac/vac} \quad (2.3)$$

$$\triangle\sigma^{N1}_{cont} = \sigma^N_{NMA*/cont} - \sigma^N_{NMA^*/vac} \quad (2.4)$$

$$\triangle\sigma^{N1}_{3w} = \sigma^N_{NMA+3w/cont} - \sigma^N_{NMA^*/cont} \quad (2.5)$$

$$\triangle\sigma^{N2}_{3w} = \sigma^N_{NMA+3w/vac} - \sigma^N_{NMA^*/vac} \quad (2.6)$$

$$\triangle\sigma^{N2}_{cont} = \sigma^N_{NMA+3w/cont} - \sigma^N_{NMA+3w/vac} \quad (2.7)$$

Here the subscripts in chemical shieldings consist of two parts separated by a slash. The first part refers to the geometry of the model and the second part indicates if the chemical shift calculation for the given geometry is performed in gas

Table 2.2: Characteristics of the calculated ^{15}N chemical shielding tensor for N-methyl-acetamide using gas phase (in vacuo), continuum-only, and cluster/continuum models.

Model[1]	σ_{11}[2]	σ_{22}[2]	σ_{33}[2]	σ_{iso}[3]	CSA[4]	η[4]	α[5]	β[5]	γ[5]	A_1[6]	A_2[6]	A_3[6]
$NMA*/cont$	18.69	157.75	207.13	127.86	-163.75	0.45	-0.78	18.04	0.48	89.52	0.57	89.7
$NMA*/vac$	44.40	149.81	220.00	138.07	-140.51	0.75	-0.93	17.31	0.53	89.47	0.60	89.74
$NMA+3w/vac$	21.54	160.61	195.18	125.77	-156.35	0.33	-1.88	20.15	0.69	89.31	1.14	89.09
$NMA+3w/cont$	7.33	163.13	185.61	118.69	-167.04	0.20	-2.10	20.47	0.62	89.38	1.46	88.68
NMA_vac/vac	38.59	148.55	229.03	138.72	-150.20	0.80	-0.37	17.14	-0.47	89.53	0.62	89.59
$NMA\_cont/cont$	16.85	156.04	210.42	127.77	-166.38	0.49	0.24	18.24	0.02	89.98	0.02	89.99

1.NMA_cont refers to NMA molecular structure optimized in continuum-only model

2.Principal components (in ppm) of the ^{15}N shielding tensor, ordered such that $\sigma_{11} \leq \sigma_{22} \leq \sigma_{33}$

3.Isotropic chemical shielding, $\sigma_{iso} = \mathrm{tr}(\boldsymbol{\sigma})/3$

4.Anisotropy, CSA $= \sigma_{11} - (\sigma_{22} + \sigma_{33})/2$, and asymmetry, $\eta = (\sigma_{22} - \sigma_{33})/(\sigma_{11} - \sigma_{iso})$ of the chemical shielding tensor

5.Euler angles (in degrees) determining the orientation of the principal axes of the chemical shielding tensor, defined as shown in 2.5

6.A_1, A_2, and A_3 are the angles between the normal to the peptide plane (defined by the peptide bond and the $N_i - C\alpha_i$ bond) and the orientation of the principal axes of the tensor corresponding to its principal components σ_{11}, σ_{22}, and σ_{33}, respectively

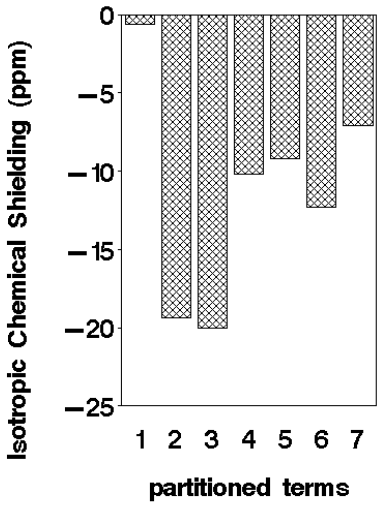


Figure 2.2: Partitioning of the ^{15}N chemical shielding in NMA into various contributions. The numbers 1-7 on the x-axis refer to Eqs. 2.1-2.7,respectively.

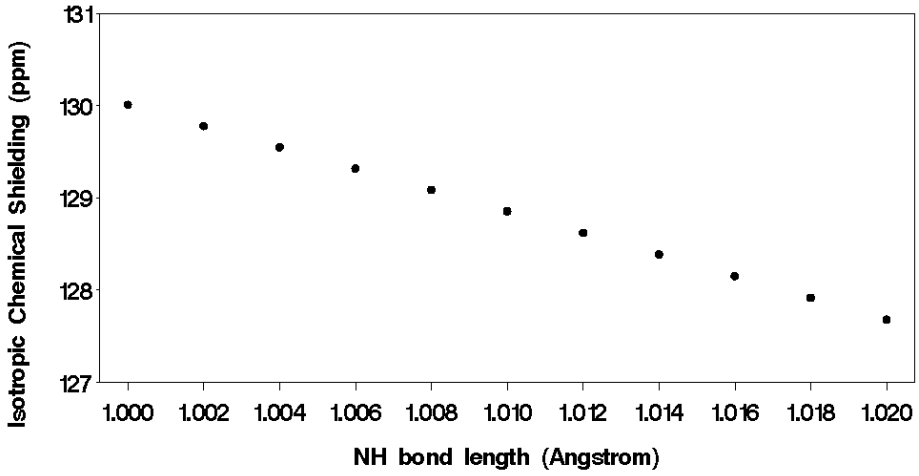
phase (*vac*) or in polarizable continuum (*cont*). Three geometries have been employed in this analysis: NMA_vac refers to the NMA molecule structure optimized in vacuo; $NMA + 3w$ refers to the optimized structure of NMA and three water molecules from the first coordination shell within the polarized continuum model; NMA* refers to the molecular geometry obtained by extracting NMA atoms from the $NMA + 3w$ structure. These $\triangle\sigma$ terms partition the solvent effects into several important contributions. The first difference, $\triangle\sigma^N_{geom}$, shows how the shielding constant calculated in vacuum is changed due to the distortions in the geometry of NMA caused by its aggregation with three water molecules in continuum. $\triangle\sigma^N_{solv}$ represents the effect of solvation of NMA by the hydrogen-bonded water molecules and by more distant water surrounding modeled by a continuum reaction field. $\triangle\sigma^N_{total}$ accounts for the total shift due to both the geometrical distortions and the solvation. Our results (Figure 2.2) show that $\triangle\sigma^N_{geom}$, being -0.65 ppm as a deshielding effect, is small compared to the deshielding effect of solvation, $\triangle\sigma^N_{solv}$, which came out to

be -19.38 ppm. Thus we conclude that the solvation is the dominating solvent effect in ^{15}N chemical shielding in NMA. To further analyze the roles played by the three hydrogen-bonded water molecules and the more distant water reaction field, we tried to partition the solvation shift, $\triangle\sigma^{N}_{solv}$, as $\triangle\sigma^{N}_{solv} = \triangle\sigma^{N1}_{cont} + \triangle\sigma^{N1}_{3w} = \triangle\sigma^{N2}_{3w} + \triangle\sigma^{N2}_{cont}$, with the superscripts "1" and "2" representing two artificial paths to account for the two sources of contributions. The two paths differ in the order in which the contributions from explicit and continuum waters are taken into account (see Eqs. 2.3-2.7). For either path, the two contributions are additive. As calculated, it is not possible to quantitatively separate the contributions this way as $\triangle\sigma^{N1}_{cont}$ (-10.21 ppm) and $\triangle\sigma^{N2}_{cont}$(-7.08 ppm) are not equal, and by definition neither are $\triangle\sigma^{N1}_{3w}$ (-9.17 ppm) and $\triangle\sigma^{N1}_{3w}$ (-12.30 ppm). Although $\triangle\sigma^{N}_{3w}$ takes on different values depending on which water molecules (hydrogen-bonded or distant) are considered first, it is qualitatively clear that the effect of the bound water molecules on ^{15}N chemical shielding in NMA is at least as important as, if not more important than the effect of the more distant water molecules. This is understandable since hydrogen bonding is expected to influence the electronic environment of ^{15}N greatly in the case of NMA with saturated hydrogen bonds. We also applied a similar analysis to ^{15}N chemical shielding anisotropy (CSA) values obtained from NMA calculations, namely:

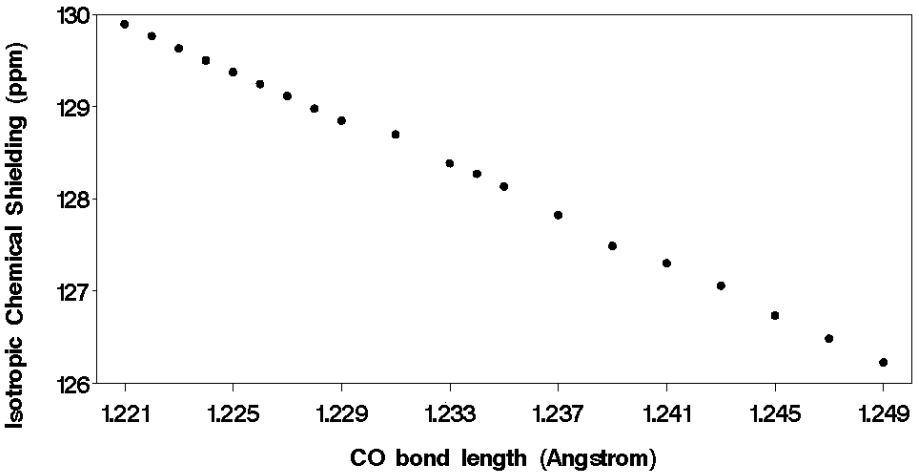
$$\triangle CSA^{N}_{solv} = \triangle CSA^{N1}_{cont} + \triangle CSA^{N1}_{3w} = \triangle CSA^{N2}_{3w} + \triangle CSA^{N2}_{cont} \qquad (2.8)$$

with $\triangle CSA^{N1}_{cont}$ being -23.24 ppm, $\triangle CSA^{N1}_{3w}$ -3.29 ppm, $\triangle CSA^{N2}_{3w}$ -15.85 ppm, and $\triangle CSA^{N2}_{cont}$ -10.69 ppm. This clearly shows that the continuum-only model calculation for NMA* yields ^{15}N CSA value that is already very close to that for NMA

clustered with three explicit waters in continuum, since a further correction of -3.29 ppm appears minor compared to -23.24 ppm. This suggests that the continuum-only model will produce a smaller relative error in the CSA values than in the isotropic chemical shift.



(a)



(b)

Figure 2.3: Isotropic ^{15}N chemical shielding in NMA as a function of (a) the NH bond length and (b) the CO bond length

We notice that in solvent, the NH bond (in the peptide plane) and CO bond tend to stretch while the CN bond becomes shorter (Table 2.1). This agrees with recent ab initio and DFT calculations (Selvarengan and Kolandaivel, 2004). When

the NH bond length was varied between its value in vacuum and in $NMA{+}3w$ structure, with the other structural parameters being optimized in the continuum-only model, the ^{15}N chemical shielding decreased as the bond stretches (Figure 2.3(a)), indicating a deshielding effect of bulk water which tends to correlate well with the NH bond length (Pearson's $|r| \approx 1.00$). A similar linear correlation exists between the ^{15}N chemical shielding and CO bond length in the continuum-only model (Figure 2.3(b)). This is similar to previously found correlations of ^{13}C and ^{17}O chemical shielding with bond length (Aidas et al., 2007; Oldfield, 2002). Our continuum-only calculations for NMA showed that elongation of the NH bond by 0.02 Å results in reduction of ^{15}N chemical shielding by 2.33 ppm ($\sim$ 1.8%), and the elongation of the CO bond by 0.03 Å reduces this shielding by 3.92 ppm ($\sim$ 3.1%). The variation of ^{15}N chemical shielding with the NH and CO bond lengths (accompanied by corresponding changes in the optimized geometry) is small compared to other nuclei (Aidas et al., 2007) or ^{15}N in a different chemical environment (Manalo and de Dios, 2002).

2.2.2 N-formyl-alanyl-X dipeptide calculations

All dipeptide structures, where appropriate, adopted χ_1 angles close to 180° after the geometry optimization, provided that the optimization started with such conformation. This might not represent the global energy minimum of the dipeptide though, since the energy barriers between the rotameric conformations could hinder the side chain's rotation to its energy minimum during optimization. The calculated

^{15}N shielding tensors are shown in Tables 2.3 and 2.4.

We assume that the differences between the DFT-calculated ^{15}N chemical shift and its true value are systematic and depend only on the local electron density around the nitrogen. Therefore these deviations should be the same for all nitrogens in the similar chemical environment. We chose methylamine (CH_3NH_2) as a reference compound, because it is the simplest molecule that has similar chemical bonding structure for the amide nitrogen, and reliable experimental ^{15}N chemical shielding in methylamine is available. Then to cancel out possible systematic errors, the chemical shift sample is computed as (Benzi et al., 2004)

$$\delta_{sample} = \sigma_0 - (\sigma^{comp}_{sample} - \sigma^{comp}_{ref} + \sigma^{exp}_{ref}), \tag{2.9}$$

where δ_{sample}, σ^{comp}_{sample} refer to the N-formyl-alanyl-X amides under study; $\sigma_0 = 244.6$ ppm is the absolute ^{15}N chemical shielding of liquid ammonia at 25°C (Jameson et al., 1981); σ^{exp}_{ref} is the experimental ^{15}N chemical shielding for methylamine, reported to be 249.5 ppm (Cramer, 2004) and $\sigma^{comp}_{ref} = 237.9$ ppm is the corresponding theoretical chemical shielding computed at B3LYP/6-311+G(2d,p) level of theory.

Figure 2.4(a)-(b) shows the values of ^{15}N chemical shift as a function of the residue type for the β-sheet and α-helical conformations of N-formyl-alanyl-X. The profile of ^{15}N chemical shift as a function of residue type (Figure 2.4(a)-(b)) agrees well with the statistically averaged experimental ^{15}N chemical shifts in the α-helices and β-sheets in proteins (Wang and Jardetzky, 2002). This indicates that a geometry-optimized structure in a water continuum provides a good model for chemical shifts in solution and the dynamics of local waters are effectively averaged out.

Table 2.3: Characteristics of the calculated ^{15}N chemical shielding tensor for N-formyl-alanyl-X in the β-sheet conformation

X	σ_{11}	σ_{22}	σ_{33}	σ_{iso}	CSA	η	α	β	γ	A_1	A_2	A_3
Ala	3.53	138.49	186.69	109.57	-159.06	0.45	11.27	14.28	0.37	89.63	0.39	89.87
Arg	2.25	138.62	187.23	109.37	-160.67	0.45	11.26	14.14	0.63	89.37	1.30	88.86
Asn	2.85	137.41	188.14	109.47	-159.92	0.48	7.55	14.54	-0.10	89.90	5.04	84.97
Asp	0.88	140.71	187.78	109.79	-163.37	0.43	13.82	14.32	1.93	88.07	5.49	84.86
Cys	5.38	137.24	188.51	110.38	-157.49	0.49	11.13	14.27	0.93	89.07	1.68	88.60
Gln	3.23	138.72	186.82	109.59	-159.54	0.45	11.86	14.31	0.94	89.06	2.09	88.14
Glu	0.25	138.74	184.70	107.9	-161.47	0.43	12.62	13.88	1.41	88.59	2.59	87.83
Gly	21.16	146.76	204.87	124.26	-154.66	0.56	11.94	19.42	1.11	88.89	5.48	84.63
His	6.98	139.09	189.33	111.80	-157.24	0.48	10.46	15.29	0.11	89.89	0.32	89.70
Ile	2.27	140.95	192.71	111.98	-164.56	0.47	7.72	15.42	-2.40	87.60	7.89	82.49
Leu	-0.66	137.87	184.63	107.28	-161.91	0.43	11.73	13.54	1.16	88.84	2.73	87.53
Lys	1.10	138.51	186.96	108.86	-161.64	0.45	9.45	13.88	0.30	89.70	2.47	87.55
Met	2.31	137.72	186.06	108.7	-159.59	0.45	8.22	14.20	-0.45	89.55	3.38	86.65
Phe	2.76	138.38	187.18	109.44	-160.02	0.46	9.95	14.32	0.39	89.61	1.76	88.29
Ser	13.46	141.82	197.46	117.58	-156.18	0.53	12.54	16.62	0.17	89.83	5.06	84.95
Thr	8.31	141.09	194.53	114.64	-159.50	0.50	7.35	15.96	-3.38	86.62	6.12	84.90
Trp	2.55	137.72	185.95	108.74	-159.29	0.45	11.27	14.17	0.91	89.09	2.04	88.17
Tyr	2.83	138.48	186.78	109.36	-159.81	0.45	10.22	14.34	0.52	89.48	1.72	88.36
Val	1.99	140.89	192.46	111.78	-164.69	0.47	7.98	15.06	-1.88	88.12	8.73	81.48
Mean	4.39	139.43	189.41	111.08	-160.03	0.47	10.44	14.84	0.14	89.00	3.49	86.73

The meaning of all parameters and the units are the same as in Table 2.2

Table 2.4: Characteristics of the calculated ^{15}N chemical shielding tensor for N-formyl-alanyl-X in the α-helix conformation

X	σ_{11}	σ_{22}	σ_{33}	σ_{iso}	CSA	η	α	β	γ	A_1	A_2	A_3
Ala	0.46	143.87	190.58	111.64	-166.76	0.42	-4.43	15.79	0.43	89.57	14.48	75.52
Arg	2.50	143.41	190.83	112.24	-164.62	0.43	-6.67	15.79	-0.64	89.36	15.55	74.47
Asn	3.51	146.59	192.11	114.07	-165.84	0.41	-6.78	16.17	-0.79	89.21	15.99	74.03
Asp	-0.20	144.06	189.24	111.03	-166.85	0.41	-5.46	15.85	-1.65	88.35	17.16	72.92
Cys	3.21	143.89	191.46	112.85	-164.46	0.43	-6.85	15.71	-0.69	89.31	14.27	75.75
Gln	2.69	143.39	190.37	112.15	-164.2	0.43	-8.37	15.79	-0.61	89.39	16.22	73.79
Glu	-0.10	141.61	188.49	110.00	-165.15	0.43	-6.46	15.63	-0.50	89.50	15.91	74.10
Gly	20.93	169.27	192.37	127.53	-159.9	0.22	-26.40	18.64	-1.29	88.71	31.83	58.21
His	2.55	148.34	192.67	114.52	-167.96	0.40	-11.22	15.86	-1.72	88.28	15.83	74.27
Ile	7.37	146.24	185.33	112.98	-158.42	0.37	11.10	19.79	3.25	86.75	23.88	66.37
Leu	-0.84	141.08	189.16	110.16	-166.02	0.43	-8.63	15.28	-1.35	88.65	16.08	73.98
Lys	1.96	143.49	190.40	111.95	-164.99	0.43	-5.67	15.74	-0.69	89.31	15.49	74.53
Met	2.07	143.46	189.978	111.84	-164.64	0.42	-8.71	15.6	-1.01	88.99	16.75	73.28
Phe	-0.47	146.35	189.56	111.81	-168.42	0.38	-7.42	15.39	-1.59	88.41	15.87	74.21
Ser	13.53	157.75	191.78	121.02	-161.24	0.32	-11.63	16.78	-1.26	88.74	19.30	70.74
Thr	16.03	157.26	189.21	120.83	-157.20	0.3	9.27	19.20	4.50	85.50	18.22	72.38
Trp	-0.57	146.68	188.84	111.65	-168.34	0.38	-7.72	15.35	-0.21	89.79	16.41	73.59
Tyr	-0.49	146.58	189.29	111.79	-168.43	0.38	-6.70	15.39	-1.42	88.58	16.00	74.07
Val	7.36	144.05	186.21	112.54	-157.77	0.40	10.47	19.36	4.00	86.00	23.81	66.57
Mean	4.29	147.23	189.89	113.82	-164.27	0.39	-5.70	16.48	-0.17	88.55	17.84	72.25

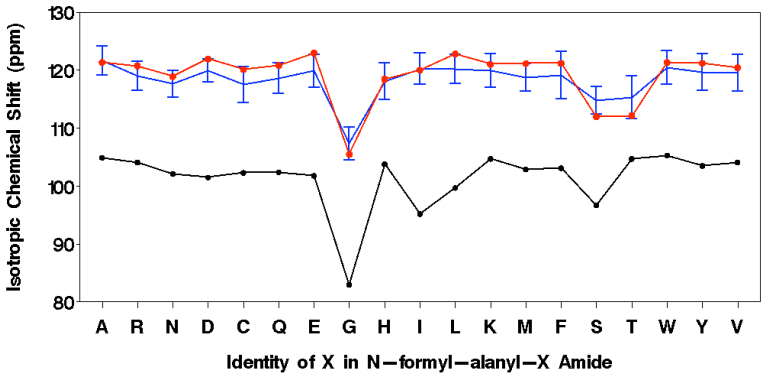
The meaning of all parameters and the units are the same as in Table 2.2

If we assume that the available protein chemical shift database samples all possible configurations, then a particular residue type in a particular secondary structure, averaged over the database, ought to produce a reasonable mean isotropic chemical shift for that residue in that secondary structure. Since our calculation used standard α-helical and β-sheet conformations, the isotropic chemical shifts we obtained may represent a mean as well. In fact, by comparison, our calculations overestimate the isotropic chemical shift by 2-3 ppm (on average) for the α-helical conformation for most N-formyl-alanyl-X amides (Figure 2.4(a)), but no over- or underestimation is obvious for the β-sheet conformation (Figure 2.4(b)). Considering the standard error associated with the statistical average, this overestimation may be even less significant.

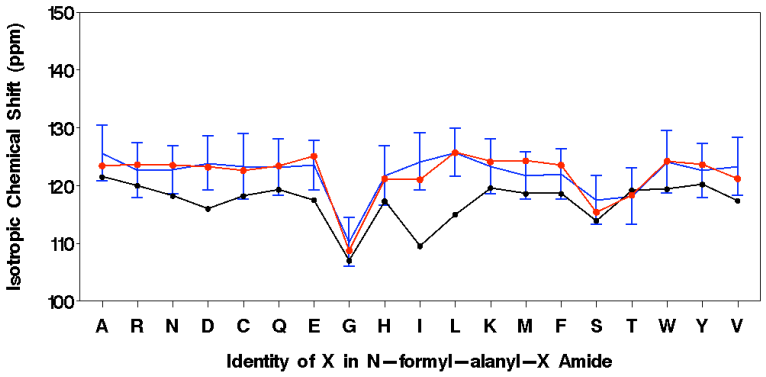
Now we turn our attention to the overall difference between ^{15}N chemical shifts for α-helical and β-sheet conformations. We define the difference in ^{15}N chemical shifts between the standard β-sheet and α-helical conformations as

$$\triangle\delta^X_{struc} = \delta^X_{sheet} - \delta^X_{helix}, \tag{2.10}$$

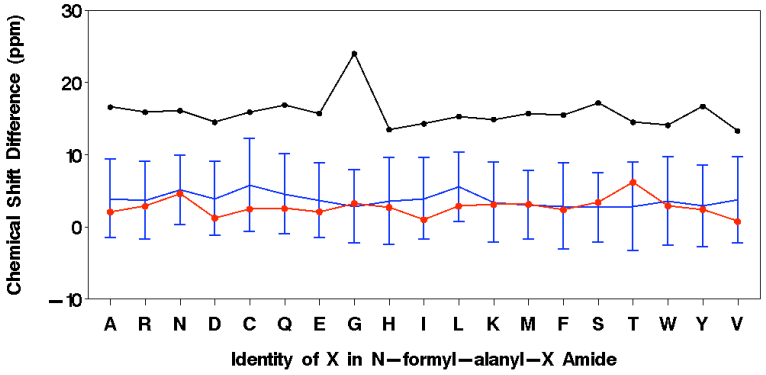
where X refers to one of the 19 amino acid residues. This difference in our calculation ranges from 0.76 ppm to 6.19 ppm with an average of 2.74 ppm. Now we would like to compare our results with the gas phase calculations of Poon et al. (2004) who used the same dipeptide model but without structure optimization and solvent consideration. Their gas phase calculations indicated that the difference in ^{15}N chemical shifts between the standard β-sheet and α-helix varies between 13.2 ppm and 24 ppm, with the average value of 15.8 ppm. Our calculation resulted in a



(a)



(b)



(c)

Figure 2.4: Comparison of the isotropic ^{15}N chemical shifts calculated in this study (*red*) with gas phase calculations (*black*) (Poon et al., 2004) and statistically averaged experimental data (*blue*) (Wang and Jardetzky, 2002) for each residue type. Panels (a) and (b) correspond to the α-helix and β-sheet conformations,respectively. (c) $\triangle\delta^X_{struc}$, the difference of the chemical shift values between the β-sheet and α-helix conformations

greatly reduced chemical shielding difference between the β-sheet and α-helix, (Figure 2.4(c)), which is now in agreement with the results from the statistical analysis of a database containing more than 6100 amino acid residues in proteins, where this difference ranged from 2.66 ppm to 5.80 ppm with an average of 3.74 ppm (Wang and Jardetzky, 2002). A close examination shows that the biggest discrepancy between our values and the experimental data is for threonine (about 3.4 ppm, see Figure 2.4(c)). However, it is well within the range of statistical uncertainty in the protein database, and we note that a variation in chemical shift due to side-chain conformation can be up to $\sim$5 ppm according to statistical analysis for amino acids Val, Ile, Thr, Phe, His, Tyr, and Trp (Wang and Jardetzky, 2004) (see also our calculations below). This discrepancy may be well due to the fact that side-chain configuration averaging is not considered in our calculation. Figure 2.4(c) demonstrates that by applying the continuum-only model the magnitude of $\triangle\delta^X_{struc}$ is reduced by $\sim$13 ppm from that of Poon et al. (2004). This reduction in the difference in shielding between the two secondary structures is due to the different deshielding effects the bulk water has for the two backbone conformations: the bulk water deshields ^{15}N by $\sim$18 ppm in the α-helical conformation but only by $\sim$5 ppm in the extended conformation of the β-sheet. This can be qualitatively understood based upon the following consideration. We considered the solvated dipeptide as a molecule embedded in a cavity in bulk dielectric with the dielectric constant of liquid water. The polarization of the surrounding dielectric continuum by the electrostatic potential of the dipeptide induces electric charges, which are distributed on the surface of the cavity. The dipeptide in the α-helical conformation is more "globular" and compact

than the dipeptide in the extended β-sheet conformation. Therefore, there are more cavity surface charges in proximity to the amide nitrogen in the helical conformation of the dipeptide than in the β-sheet conformation.

As noted above in continuum-only calculation of NMA, neglecting close-contact solvent can result in underestimation by about 9 ppm of the deshielding effect of solvent in the NMA model with saturated hydrogen bonds. Since secondary structure elements in proteins are often hydrogen-bonded, it is important to know the magnitude of possible underestimation for the ^{15}N chemical shielding in dipeptide caused by ignoring these (specific) interactions. In order to assess this effect, we performed ^{15}N chemical shielding calculations for selected residues from protein GB3. In these calculations, a fragment, C-Y-X-N, containing the residue of interest (X) and its preceding residue (Y) was taken directly from the crystal structure of GB3 (PDB code 1IGD) (Derrick and Wigley, 1994) and modified by replacing the end atoms to become N-formyl-Y-X-NH_2 (see Figure 2.1(d)). In addition, in the case of α-helix, the side chain of the residue preceding X was replaced with CH_3. The hydrogen-bonded residues (through the NH group of X and, where applicable, CO group of the preceding residue) were also taken from the GB3 structure (Figure 2.1(d)) and modified to become either CH_3-CO-NH-CH(CH_3)-CO-NH-CH_3 or CH_3-CO-NH-CH(CH_3)-CO-NH-CH(CH_3)-CO-NH-CH_3 (as detailed in Table 2.5). We then performed chemical shift calculations in vacuum and in continuum for the dipeptide alone and for this hydrogen-bonded cluster. The results (Table 5) show that the dipeptide in continuum model yields deshielding up to 15.3 ppm for α-helix and 9.5 ppm for β-sheet. Depending on the hydrogen bonding geometry, the clus-

Table 2.5: Isotropic ^{15}N chemical shielding (in ppm) calculated for selected residues in GB3 using dipeptide in gas phase (in vacuo), dipeptide in continuum-only, cluster in gas phase (in vacuo), and cluster in continuum models

Residue X	Conformation	Dipeptide in vacuum	Dipeptide in continuum	Cluster in vacuum	Cluster in continuum
Thr16	β-sheet	133.20	123.74	133.29[2,4]	122.60[2,4]
Thr18	β-sheet	127.58	122.68	128.14[2,4]	121.29[2,4]
Thr44	β-sheet	139.14	133.96	136.55[2,4]	129.86[2,4]
Thr53	β-sheet	124.78	118.73	122.27[3,4,6]	114.56[3,4]
Lys28	α-helix	144.84	129.52	/	125.85[3,5]
Ala29	α-helix	135.51	125.88	128.68[3,5]	123.49[3,5]
Gln32	α-helix	134.33	125.21	125.73[3,5]	121.37[3,5]

1. Residue numbering here corresponds to GB3 crystal structure (1IGD).
2. Only direct hydrogen-bonding partner exists
3. Both direct and indirect hydrogen-bonding partners exist
4. Hydrogen-bonding partner was modeled as CH_3-CO-NH-CH(CH_3)-CO-NH-CH_3
5. Side chain of the previous residue was replaced by $-CH_3$; the direct hydrogen-bonding partner was models as CH_3-CO-NH-CH(CH_3)-CO-NH-CH_3 and the indirect hydrogen-bonding partner as CH_3-CO-$NH^{(*)}$-CH(CH_3)-CO-NH-CH(CH_3)-CO-NH-CH_3 (with $NH^{(*)}$ hydrogen bonded to the CO group of the residue preceding X in the dipeptide)
6. This calculation was performed only with direct hydrogen-bonding partner

ter in continuum calculation can further deshield ^{15}N by about 1 to 4 ppm. This suggests that the polarizable continuum model can account for hydrogen bonding in a realistic protein secondary structure and, therefore, might be a reasonable first approximation for computing ^{15}N chemical shielding.

The chemical shielding tensor contains a wealth of potentially useful structural information, which could be lost when the tensor is reduced to isotropic shielding. Knowledge of individual components and orientation of ^{15}N shielding tensor could be important for many NMR applications, including accurate analysis of protein dynamics from ^{15}N relaxation data (e.g. (Fushman and Cowburn, 2001; Hall and Fushman, 2006)), TROSY-based experiments, cross-correlation effects involving ^{15}N

CSA, and the use of residual ^{15}N chemical shift anisotropy upon molecular alignment (e.g. (Lipsitz and Tjandra, 2003)) as restraints for structure refinement. It is therefore important to understand the dependence of particular components of the ^{15}N chemical shielding tensor on solvent, side chain, and conformation. In our calculations, the ^{15}N CSA values (Tables 2.3 and 2.4) fall in the range of experimental values reported for proteins (Cornilescu and Bax, 2000; Fushman et al., 1998, 1999; Kroenke et al., 1999; Kurita et al., 2003; Loth et al., 2005; Tjandra et al., 1996; Wylie et al., 2006; Wylie and Rienstra, 2008). The difference in CSA values between the α-helix and β-sheet (mean CSA's of -164.3 ppm and -160.0 ppm, respectively) is consistent with the observations for ubiquitin (Cornilescu and Bax, 2000) and GB1 (Wylie and Rienstra, 2008). It is worth pointing out that in our data this difference arises primarily from σ_{22}, which is systematically higher in α-helix (by 7.8 ppm on average), while the other two components of the ^{15}N shielding tensor (particularly σ_{11}) show a considerably smaller and less systematic difference between the β-sheet and α-helix conformations (see Tables 2.3 and 2.4). The calculated ^{15}N CSA values also agree with the solid state NMR measurements in short peptides (Hartzell et al., 1987; Hiyama et al., 1988; Mai et al., 1993; Oas et al., 1987; Shoji et al., 1989; Wu et al., 1995). A good agreement with the experimental data is also found for the angle between the least shielded component (σ_{11}) of the ^{15}N shielding tensor and the NH bond. The values of β obtained here, from 13.5° to 19.8°, are well in the range of the experimental values (12°-24°) obtained by different NMR techniques, both solution and solid-state (Cornilescu and Bax, 2000; Fushman et al., 1998; Hall and Fushman, 2006; Hartzell et al., 1987; Hiyama et al., 1988; Kurita et al., 2003;

Loth et al., 2005; Mai et al., 1993; Oas et al., 1987; Shoji et al., 1989; Vasos et al., 2006). There seems to be a weak correlation between the β angle and secondary structure, with slightly smaller angles for the β-sheet than for the α-helix (mean β angles are 14.8° and 16.5°, respectively). This also agrees with the experimental findings in GB3 (Hall and Fushman 2006) and ubiquitin (Fushman et al. 1998).

Our calculations show a considerable spread in ^{15}N CSA values, from -154.7 ppm to -168.4 ppm, depending on the residue type and the backbone conformation. This range, however, is smaller than the ^{15}N CSA dispersion observed by solution NMR in ubiquitin and GB3 (Fushman et al., 1998, 1999; Hall and Fushman, 2006; Kover and Batta, 2001) and by solid-state NMR in GB1(Wylie et al. 2006; Wylie and Rienstra 2008). This likely reflects the fact that these calculations do not take into account the complexity of local electronic environment in proteins, including interactions with neighboring atoms (e.g. hydrogen bonding, charge and ring-current effects), deviations of the backbone and side chain conformations from those considered here, averaging by anisotropic dynamics etc. Note, for example, that while the type of amino acid residue X varied in our calculations, the torsion angle χ_1 was close to 180° (where applicable) for both backbone conformations. In order to explore the effect of side chain's rotameric state on the ^{15}N chemical shielding tensor, we performed a set of calculations for glutamate (X=Glu) in N-formyl-alanyl-X in the β-sheet conformation, in which the angle χ_1 was fixed at -180°, -150°, 60°, -6° in the geometry optimization. The results showed a significant variation in the anisotropy of the shielding tensor (CSA = -161.38, -165.29, -164.94, and -153.98 ppm, respectively), which is bigger than for the isotropic shielding (107.96, 105.91,

114.05, and 111.18 ppm, respectively). These results emphasize the importance of side-chain conformation for the calculation of the ^{15}N shielding tensor, which could be one of the reasons for the wider spread in the experimental ^{15}N CSA values than that calculated here.

Also we notice that in the α-helical conformation, valine, isoleucine, and threonine have particularly low absolute ^{15}N CSA values. This can be due to the presence of branched side chains in these residues that disturbed the structure by tilting the NH bond out of the peptide plane by about 13°. The effect could result from a particular combination of the torsion angles (ϕ, ψ, χ_1) as it is not observed in the β-sheet conformation.

It is instructive to discuss the orientation of the individual components of the ^{15}N chemical shielding tensor. The expectation from solid state NMR measurements and planar symmetry of the peptide bond (see e.g. (Oas et al. 1987)) is that the least shielded component (σ_{11}) is lying in the peptide plane and tilted by a small angle (see Figure 2.5) from the NH-bond, while the intermediate component, σ_{22}, is orthogonal to the peptide plane. Interestingly, while the least shielded component σ_{11} of our calculated ^{15}N chemical shielding tensor lies almost in the peptide plane for all residues, independent of the backbone conformation, the orientations of the other two components differ between the two conformations (Tables 2.3 and 2.4). Our calculations show that, in the β-sheet, σ_{22} is almost orthogonal to the peptide plane (the biggest tilt is $\sim$ 9° for valine), which automatically places the most shielded component, σ_{33}, close to the peptide plane. The deviations from "ideal" picture are more dramatic for the α-helix. Here the σ_{22} component is tilted by as

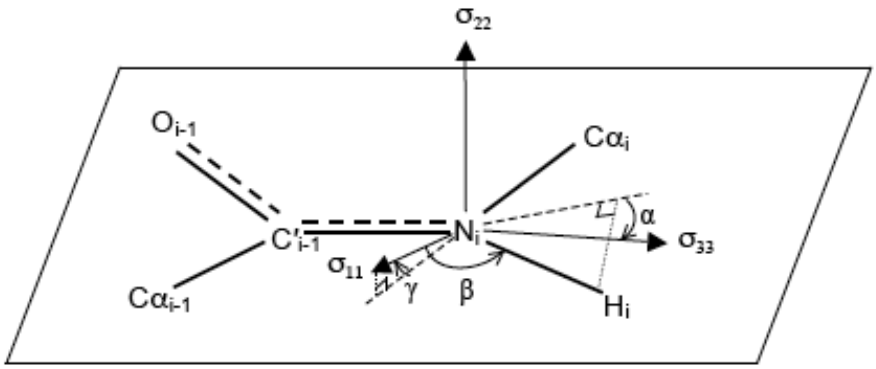


Figure 2.5: Schematic illustration of the orientation of the principal components of the ^{15}N chemical shielding tensor with respect to the peptide plane defined by $C' - N - C_\alpha$. σ_{11} is the least shielded component, tilting out of the peptide plane by an angle γ and forming an angle β with the NH bond. σ_{22} is the next least shielded component and stands roughly perpendicular to the peptide plane. σ_{33} is the most shielded component and lies approximately in the peptide plane; the projection of NH bond onto the plane of σ_{22} and σ_{33} forms an angle α with σ_{33}. This definition of the angles is taken from Brender et al. (2001).

much as 32° for glycine and 24° for valine and isoleucine, and the σ_{33} component is also significantly tilted away from the peptide plane. These results demonstrate the effect of the backbone conformation on the orientation of the ^{15}N shielding tensor.

Another important aspect is the asymmetry of the shielding tensor, as it is often assumed (e.g. in ^{15}N relaxation analysis) that the tensor is axially symmetric, although solid state NMR data on short peptides indicated that deviations from axial symmetry could be substantial (Hiyama et al., 1988; Oas et al., 1987). For the N-formyl-alanyl-X examples considered here the asymmetry of the ^{15}N shielding tensor, defined as

$$\eta = (\sigma_{22} - \sigma_{33})/(\sigma_{11} - \sigma_{iso}), \tag{2.11}$$

ranges from 0.43 to 0.56 for the β-sheet and from 0.22 to 0.43 for the α-helix.

The higher ^{15}N shielding asymmetry in the β-sheet conformation is consistent with the data reported for ubiquitin (Cornilescu and Bax, 2000). These differences in the asymmetry of ^{15}N shielding between the two backbone conformations could be related to the differences in the orientation of the tensor. The absolute values of the asymmetry are somewhat higher than the experimentally observed in solution (Cornilescu and Bax, 2000; Loth et al., 2005) but comparable to solid-state NMR data (Wylie et al., 2006), which likely reflects motional averaging expected to be more pronounced in proteins in solution.

2.3 Summary and conclusions

To examine the effects of solvation, backbone conformation, and the side chain on ^{15}N chemical shielding in proteins, we performed density-functional theory calculations with the polarizable continuum solvent model for NMA and N-formyl-alanyl-X amides, where X is one of the 19 naturally occurring amino acids excluding proline. The main results of our calculations can be summarized as follows:

1) Solvent considered as the polarizable continuum model with the explicit water molecules in the first solvation shell has a considerable effect on the isotropic chemical shift but not as much on the anisotropy of the chemical shielding tensor.

2) The calculations for the dipeptides demonstrated that the averaged over all 19 types of residues difference in isotropic ^{15}N chemical shifts between the standard β-sheet and α-helical conformations is 2.7 ppm, in good agreement with the experimentally observed difference of 3-4 ppm in proteins.

3) The orientation of the ^{15}N chemical shielding tensor as well as its anisotropy and asymmetry are overall in the range observed for peptides and proteins. Our calculations show that for both backbone conformations, the least shielded component, σ_{11}, of the tensor lies approximately in the peptide plane and makes an angle of 13.5° to 19.8° with the NH bond. In the β-sheet, the intermediate component, σ_{22}, is almost orthogonal to the peptide plane and the most shielded component, σ_{33}, lies almost in the peptide plane. However, in the α-helix the σ_{22} component is tilted by as much as 32° for Gly and 24° for valine and isoleucine, and the σ_{33} component is also significantly tilted away from the peptide plane.

4) The anisotropy of the ^{15}N chemical shielding tensor varies among amino acids in the range from -154.7 ppm to -168.4 ppm with the mean value of -160 ppm.

5) The asymmetry of ^{15}N chemical shielding tensor varies from 0.43 to 0.56 for the β-sheet and from 0.22 to 0.43 for the α-helix.

6) Our calculations for selected fragments from protein GB3 suggest that the polarizable continuum model could serve as a reasonable approximation for the effect of protein environment on ^{15}N chemical shielding.

Chapter 3

Chemical shielding calculations of ^{15}N in real proteins: classical α-helical residues

3.1 Computational details

All calculations reported in this chapter were performed using the GAUSSIAN03 suite of programs (Frisch et al., 2004). We used density-functional theory (DFT) with three-parameter Becke-Lee-Yang-Parr (B3LYP) exchange-correlation functional (Becke, 1988, 1993; Lee et al., 1988). The atom coordinates were taken from the best representative conformer (PDB code: 2OED) of the ensemble of GB3 solution NMR structures (Ulmer et al., 2003). The bond lengths were used directly from the experimental structure without geometry optimization.

Only "classical" helical residues, A26 through Y33, of GB3's α-helix (that spans residues D22-N37) are examined in this disseration. According to the definition used here, a classical helical residue i is hydrogen bonded through its NH group to residue i-4 and through its CO group to residue i+4. Thus the backbone nitrogen of residue i has a direct hydrogen bonding partner i-4 and an indirect hydrogen bonding partner i+3 (Figure 3.1). To avoid terminal artifacts, four residues at either end of the α-helix were not examined here because they are missing either a direct or an indirect hydrogen bonding partner.

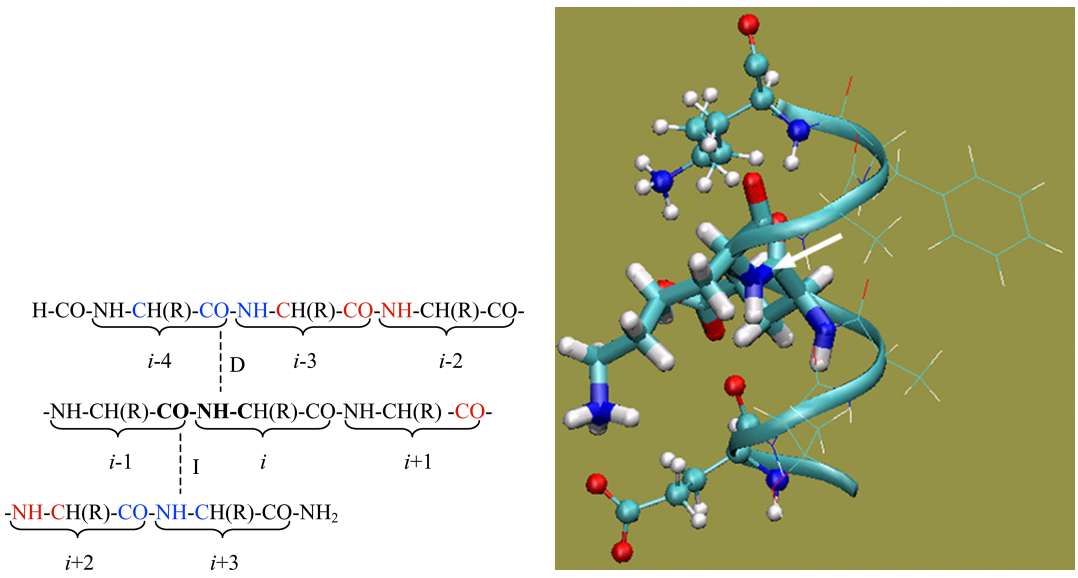
Several models with different levels of complexity are investigated here (illustrated in Figure 3.1(b) for the case of amide nitrogen of K28):

Model A: a simple dipeptide Model containing only residues i and i-1. If not indicated directly, the ionizable side chains are assumed to be in their charged state. In some cases, as indicated in the text, the ionizable side chains were altered to be neutral.

Model B: a main fragment containing residues i and i-1 (as in Model A) and two additional fragments of the hydrogen bonding partners, both direct and indirect. This results in a "three-fragment" model. In some cases, as specified in the text, the side chains of the hydrogen bonding partners were altered to be neutral or modified to be $-CH_3$.

Model C: a "long-chain" Model containing a stretch of residues from the direct hydrogen bonding partner to the indirect hydrogen bonding partner as one main fragment. That is, for classical helical residues the "long chain" includes residues from i-4 to i+3. In some cases, as specified in the text, some side chains of the "long chain" were altered to be neutral or changed to $-CH_3$, or the "long chain" was extended to include an additional residue at one end.

In our calculations, the main fragments in various models had the N-terminus capped by a formyl group (-COH) and the C-terminus capped by an amino group ($-NH_2$), and the hydrogen bonding partners in Model B were modified to be CH_3-CO-(NH-CH(R)-CO)-NH-CH_3. When the charges from the atoms of the rest of the protein were included in the calculation, we refer to it as Charge Field Perturbation (CFP) calculation. If the charges are not included, we call it "vacuum" calculation.



(a) (b)

Figure 3.1: (a) Illustration of the basis set assignment. The dashed lines represent the direct (D) and indirect (I) hydrogen bonds for the peptide plane containing the amide group of residue *i*. (b) Illustration of various peptide fragment models used in the calculations. In the case of amide nitrogen of K28 shown here, Model A (a dipeptide model) includes E27 and K28 (thick sticks); Model B includes residues from Model A (thick sticks) and the two hydrogen bonding partners (ball-and-stick); Model C includes all residues from E24 to K31 (ribbon). Residues shown in thin-line representation are included in Model C but not in Model B. The molecular image was generated using VMD (Humphrey et al., 1996). The arrow points to the amide nitrogen whose chemical shielding is being calculated.

In CFP calculations, the point charges were all taken from AMBER charge set (Cornell et al., 1995) with the overall non-zero charge for ionizable residues. The calculations were performed using local dense basis sets (Chesnut et al., 1993). For the residue of interest *i*, we applied a 6-311+G(2d,p) basis set for the N_i, H_i, $C_{\alpha i}$, C_{i-1}, and O_{i-1} atoms (shown in bold in Figure 3.1(a)). Where applicable, the 6-311+G(2d,p) basis set was also applied to similar atoms of its direct and indirect hydrogen bonding partners (shown in blue in Figure 3.1(a)), as well as the second pair (Xu and Case, 2002) of hydrogen bonding partners (shown in red in Figure 3.1(a)). A 4-21G basis set was applied to the remaining atoms in the model.

The charge field perturbation gauge-including atomic orbital method (de Dios and Oldfield, 1993; Ditchfield, 1974; Wolinski et al., 1990) was used.

3.2 Results and discussion

3.2.1 Model Building

Fragment Size. It was suggested (Xu and Case, 2002) that a 7-9 residue sequence should be used for chemical shift calculations in the α-helix conformation in order to eliminate the terminal artifacts. The terminal effects may be due to (i) the lack of hydrogen bonding partners for the residues at either end; or (ii) the underestimation of the interaction with the dipole moment of the helix. In order to analyze the magnitude of the hydrogen bonding and helical dipole effects without side chains complicating the picture, all non-alanine residues other than residues i and i-1 were altered to be alanine (See Table 3.1), and we compared isotropic chemical shieldings from Model A to Model C vacuum calculations. As shown in Figure 3.2(a), the first pair of hydrogen bonding partners deshield ^{15}N . The deshielding ranges from 4.75 ppm to 9.89 ppm depending on the strength of hydrogen bonding. The effect is the strongest for K31 because the peptide plane containing the amide group of K31 makes the strongest hydrogen bonds. In fact, the distance between K31's amide proton and E27's carbonyl oxygen is 1.83 Å, compared to the 1.89 to 2.86 Å range for the rest of the hydrogen bonds in the GB3 helix. In addition, the O-H distance for the corresponding indirect hydrogen bond is 1.89 Å. Model C deshields ^{15}N further from Model B, ranging from 4.10 ppm to 6.59

ppm (Figure 3.2(a)). This is due to the inclusion of atoms from residues i-3, i-2, i+1, and i+2 in Model C, which introduced new aligned NH and CO groups, namely NH groups from residues i-2 and i+2, and CO groups from residues i-3 and i+1 (see Figure 3.1). These groups contribute to the helical dipole moment and bring about electrostatic field contribution to ^{15}N chemical shielding, in agreement with previous findings (Le and Oldfield, 1996; Xu and Case, 2002). In fact, our calculation showed on average an increase of 18.5 Debye in dipole moment from "Model A neutral" to "Model B neutral $-CH_3$"; and on average an additional increase of 10.2 Debye from "Model B neutral $-CH_3$" to "Model C neutral $-CH_3$" (Table 3.2). Since the "long chain" model further deshields the ^{15}N comparing to the "three-fragment" model and this deshielding differs for different residues (Table 3.1), we conclude that the "long chain" model includes some new nontrivial contributions, which are not present in the simpler models.

We note here that in the calculations with charged ionizable side chains (Figure 3.2(b)), the contributions from the first pair of hydrogen bonds and from the additional helical dipole are similar to those in the calculations with neutral ionizable side chains (Figure 3.2(a)). This indicates that the side chain charges do not interfere with hydrogen bonding effects and helical dipole effects.

Charged or neutral side chain? Previously, when building the database for SHIFTS (Xu and Case, 2002), all ionizable amino acids took neutral forms because it appeared that in gas-phase calculations, neutral side chains provided a better model for solution chemical shifts than the charged side chains did. More recently, it was suggested (Vila and Scheraga, 2007) that the whole protein and

Table 3.1: Calculated isotropic chemical shieldings (in ppm) for selected α-helical residues from GB3 with various models

Residue	Direct HB partner	Indirect HB partner	Model A neutral[1]	Model B neutral[1]	Model B $-CH_3$[2]	Model B neutral[1]	Model C neutral[1]	Model C $-CH_3$[2]	Model C neutral[1]
A26	D22	A29	131.50	126.75	128.46		121.31		122.21
E27	A23	F30	139.12	130.13	130.65		124.19		125.17
K28	E24	K31	136.88	130.37	131.31		123.84		125.06
A29	T25	Q32	137.51	131.70	131.98		125.41		125.74
F30	A26	Y33	137.00	131.07	130.23		125.56		126.55
K31	E27	A34	134.22	124.33	123.80		120.23		119.27
Q32	K28	N35	136.10	128.96	130.39		124.17		125.53
Y33	A29	D36	133.22	126.64	126.47		120.05		120.96

Residue	Model A *	Model A CFP[3]	Model B $-CH_3$[2]	Model B *	Model B partial CFP[4]	Model B CFP[3]	Model C $-CH_3$[2]	Model C *	Model C CFP[2]
A26	131.5	126.52	126.75	128.28	120.33	121.58	121.31	120.07	120.57
E27	127.38	120.64	119.59	120.12	113.06	110.49	112.96	113.98	112.65
K28	147.65	136.05	141.32	134.65	127.07	127.86	135.63	128.42	127.58
A29	134.53	130.89	129.06	129.45	122.01	125.63	123.42	122.20	124.26
F30	137.00	130.40	131.07	130.23	122.80	123.34	125.56	125.01	124.27
K31	140.56	127.09	130.10	125.80	120.89	117.52	125.99	123.17	118.08
Q32	132.26	129.49	125.69	127.96	121.33	121.86	122.65	122.70	122.23
Y33	133.22	132.59	126.64	132.02	123.70	125.18	120.05	124.78	125.96

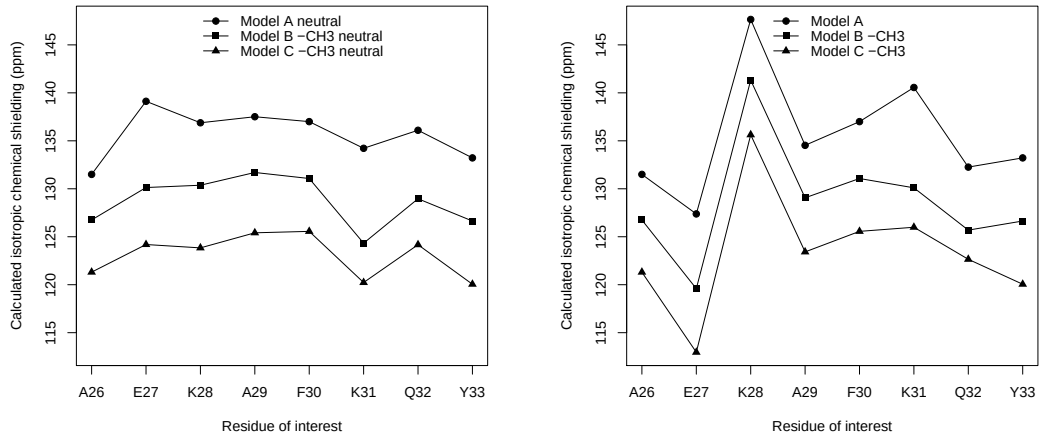
* "generic" model (A, B, or C) with no additional modifications.

1.neutral: all ionizable side chains were made neutral.

2. $-CH_3$: all nonalanine residues other than residues i and i -1 have the side chains modified to be $-CH_3$ groups.

3.CFP: the calculation is performed in presence of AMBER charges representing atoms of GB3 that are not included in the model

4."Model B partial CFP": this calculation is performed in presence of AMBER charges representing atoms that are not included in the Model B but are included in Model C.



(a) (b)

Figure 3.2: Isotropic chemical shieldings for selected α-helical residues calculated using various models. Where applicable, the ionizable side chains of residues i and i -1 are neutral in (a) but charged in (b).

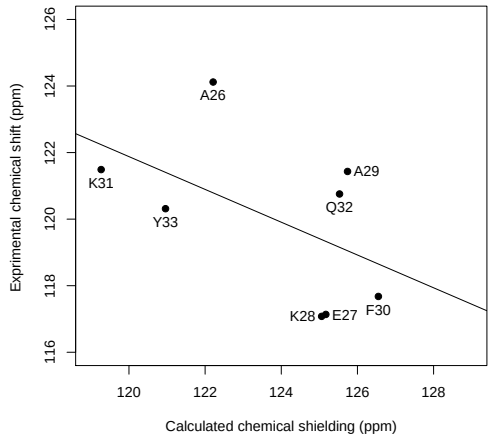
its chemical environment should be taken into account by considering the proton binding/release equilibria and then the chemical shifts should be calculated as a weighted average. That study also found that for C_α the results obtained for Asp and Glu are almost systematically worse when using charged rather than neutral side chains, presumably because these two amino acids have much shorter side chains compared to Arg and Lys. When charged, these side chains may have a bigger influence on the C_α chemical shielding. Here we explore how the ^{15}N chemical shift can be affected by the ionization form of the side chains.

As shown in Table 3.1, the side chain charge on the residue of interest has an apparent effect on its ^{15}N chemical shielding (compare the results from "Model A" and "Model A neutral" calculations for E27 and K31). Deprotonation of the side chains introduces deshielding while protonation introduces shielding, similarly to what was observed for C_α (Vila and Scheraga 2007). Thus, the negative charge

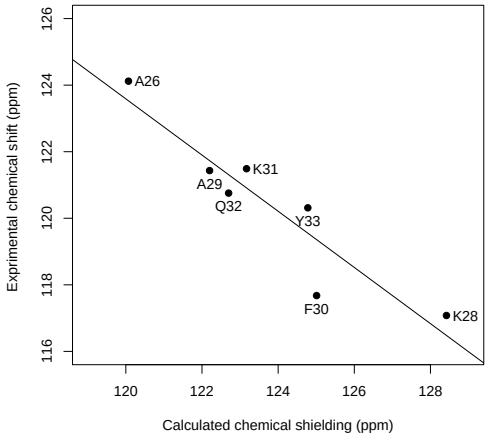
on the side chain of E27 deshields its ^{15}N by 11.74 ppm while the positive charge on the side chain of K31 shields its ^{15}N by 6.34 ppm in Model A.

Unlike C_α, surrounding side chain charges on other residues can influence amide nitrogen chemical shielding. Take K31 as an example. Its "Model B $-CH_3$" calculation (130.10 ppm) yields a more shielded value than its "Model B neutral $-CH_3$" calculation (124.33 ppm) for ^{15}N . This shielding effect of 5.77 ppm is due to the charge on K31. In the Model B calculation where K31's hydrogen bonding partner E27 is also charged, the amide nitrogen of K31 has a chemical shielding of 125.80 ppm. This 4.3 ppm deshielding from the "Model B $-CH_3$" calculation is mainly due to the charge on E27. This example illustrates that the charges on other side chains can make ^{15}N chemical shielding calculation more interesting but also more challenging because it introduces another source of variation. In the case of neutral side chains, on the contrary, their identity does not play a significant role, such that they can be safely replaced by a $-CH_3$ group. This is supported by the close resemblance of the results from "Model B neutral" and "Model B neutral $-CH_3$" calculations and, likewise, of the results from "Model C neutral" and "Model C neutral $-CH_3$" calculations.

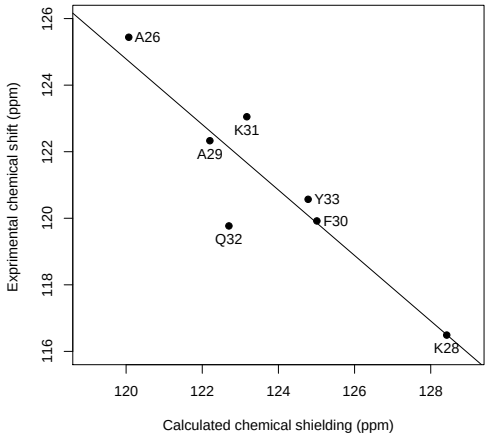
We used two approaches when calculating ^{15}N chemical shielding with Model C. The first approach makes all ionizable side chains neutral in the Model C vacuum calculation. The comparison between experimental solid-state chemical shifts and the calculated chemical shieldings is shown in Figure 3.3(a). The correlation is poor ($|r| = 0.52$), and the regression line has an intercept of 181.00 ± 41.00 ppm and a slope of -0.49 ± 0.33. The slope is far from the ideal value of -1. No obvious outlier is



(a)



(b)



(c)

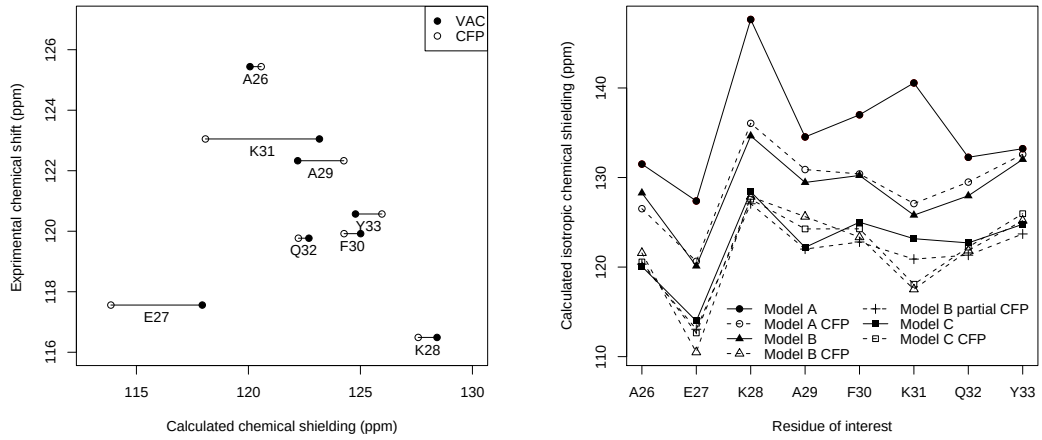
Figure 3.3: Comparison of the calculated isotropic ^{15}N chemical shieldings with experimental data for GB3. (a) "Model C neutral" and (b) Model C versus solid-state NMR data (BioMagResBank accession number 15283 (Nadaud et al., 2007)). (c) Model C versus solution NMR data (Gronenborn et al., 1991).

identified. The second approach represents all ionizable side chains in their charged states. This yields an obvious outlier, E27 (see Figure 3.4(a)). By excluding E27, a better correlation is achieved ($|r| = 0.93$) and the linear regression yields a slope of -0.84 ± 0.15 and an intercept of 224.91 ± 18.83 ppm (Figure 3.3(b)). A comparison with solution NMR data yields similar results: $|r| = 0.91$, slope $= -0.98 \pm 0.20$, and

intercept = 242.87 ± 24.70 ppm (Figure 3.3(c)). In this approach it is important to include all charges that are close to the ^{15}N site of interest (Bader, 2009). Consider E27 as an example. In the Model C vacuum calculation that includes residues from A23 to F30, we obtained an isotropic chemical shielding of 113.98 ppm for the amide nitrogen. This calculation did not include K31, which makes a salt bridge with the side chain of E27. However, extending the long chain to include K31 in the Model C vacuum calculation yielded a more shielded value of 117.94 ppm, bringing it closer to the experimental value (Figure 3.4(a)). This shielding effect of 3.96 ppm is due to the charge on K31's side chain. In comparison, a proton unit charge placed on the ϵ-amine of K31 produced a 4.77 ppm shielding effect with Model A vacuum calculation (Table 3.4). Interestingly, having E27 in its neutral form can significantly shield its amide nitrogen, as suggested by the difference (11.74 ppm) between its Model A and "Model A neutral" calculations (Table 3.1). Indeed, making E27 neutral while keeping other ionizable side chains charged (in Model C) yields an isotropic chemical shielding of 124.25 ppm. Thus, protonation of the side chain can make E27 not an outlier.

In principle, there can be other possible combinations of ionization states of D22, D24, D27, K28, K31, and D36 that are involved here (Table 3.1). The two approaches shown here emphasize the complexity of calculating backbone ^{15}N chemical shielding when dealing with uncertainties of the charge states of ionizable side chains. On the other hand, a detailed comparison with experimental NMR data might provide useful information on the ionization state of these residues.

Counterions were not explicitly included in our calculations, primarily because



(a) (b)

Figure 3.4: (a) Comparison of the isotropic chemical shielding from Model C vacuum (solid circles) and Model C CFP (empty circles) calculations with experimental (solution NMR) data. Note that the E27 data in this plot represent the results from Model C extended to include K31.(b) Comparison of the isotropic ^{15}N chemical shieldings for selected α-helical residues calculated using various models considered in this study.

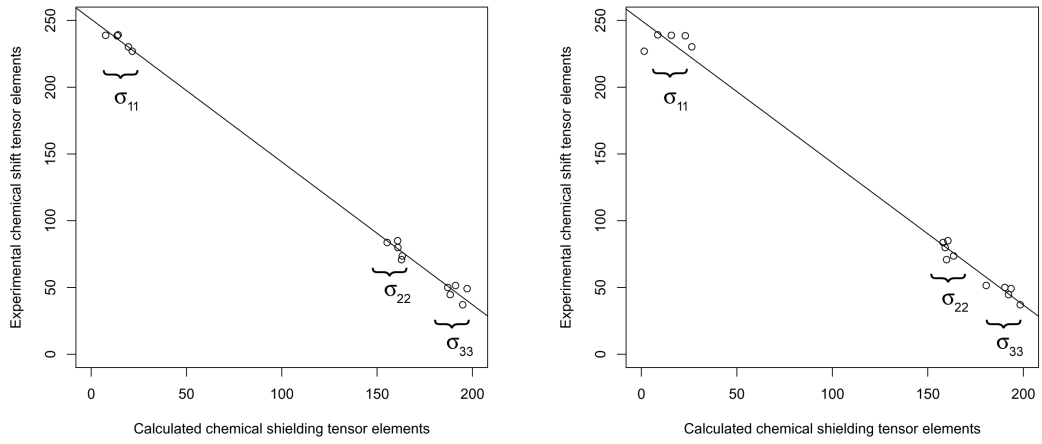
the amino acid composition and the structure of GB3's α-helix are such that every charged side chain (except for D22 at the very N-terminus and D36 at the very C-terminus) has a salt-bridge partner (K31-E27, K28-E24), which could naturally balance its charge effect. Of the long-chain fragments used in Model C calculations, the one for K28 was neutral (as was the extended fragment for E27, see above), while the rest of the fragments had one or two unbalanced charges. Including the extra residues necessary for charge balancing would have further increased the size of the fragment. However, the effect of those unbalanced charges is generally small, as can be seen from Table 1 (compare, for example, Model C with Model C -CH_3), perhaps because of the rather long distances from the charged groups to the amide nitrogen. To explore this effect further, we performed a dipeptide-model calculation

in which we included point charges mimicking the side chains that could serve as "counterions" but were not present in the corresponding Model C fragments. The results (Table 3.4) show that the effect of these additional charges on ^{15}N chemical shielding is generally small, except for E27, where including the side chain charge of K31 introduces a significant shift in the shielding, as discussed above.

3.2.2 The tensor

Comparing the shielding tensor's principal values between the calculations with "Model C neutral" and "Model C" (Table 3.3), the variation mainly exits in the least and most shielded components, σ_{11} and σ_{33}, while the intermediate component, σ_{22}, appears less susceptible to model selection. Residue E27 showed the biggest difference in all three principal components between these two calculations, with a 20 ppm difference in σ_{11}. There is a generally good agreement between these two models in the orientation of the principal components of the shielding tensor (Table 3.3). Experimental data for all three principal components of GB3's ^{15}N chemical shift tensors are not available. However, these data exist for most residues of protein GB1 (Wylie et al., 2007), which is highly homologous to GB3. In fact, all helical residues of GB3 are present in GB1, except for A29, which is a valine in GB1. Therefore we compared our calculated principal values of the ^{15}N chemical shielding tensor for A26, E27, K28, K31, and Y33 in GB3 with the corresponding components of the chemical shift tensor of these residues in GB1, obtained by solid-state NMR measurements (Figure 3.5). We found that the individual components of

the calculated chemical shielding tensor are in good agreement with the experimental data. Over the whole range of variation in the tensor's principal components, the difference between using "Model C neutral" and "Model C" is small (Figure 3.5).



(a) (b)

Figure 3.5: Comparison of the principal components of the calculated chemical shielding tensor for GB3 with the corresponding components of the experimental chemical shift tensor for GB1 (from solid-state NMR measurements, kindly provided by Chad Rienstra and Benjamin Wylie (Wylie et al., 2007)).The model used and the parameters of the regression line are as follows: (a) Model C, intercept = 250.88 ± 2.24 ppm, slope = -1.07 ± 0.02; (b) "Model C neutral", intercept = 249.98 ± 4.11 pp m, slope = -1.07 ± 0.03.

3.2.3 CFP calculations

The difference between a vacuum calculation and a CFP calculation is a useful indicator of whether all necessary electrostatic effects are taken into account in the vacuum calculation. For example, when using a "three-fragment" model (Model B) there are large differences between the CFP calculation and the vacuum calculation (Table 3.1). This suggests that the fragment that is used in the vacuum calculation is too small. When using Model C which includes every residue from the direct

hydrogen bonding partner to the indirect hydrogen bonding partner, not only the difference between vacuum and CFP calculations becomes smaller in general (Table 3.1 and Figure 3.4(b)) but also the correlation between the vacuum calculation and the experiment becomes better with a $|r|$ value of 0.93, compared to 0.68 for Model B vacuum calculation and 0.54 for "Model B CFP" calculation (E27 was not included in this analysis). From Table 3.1 and Figure 3.4(b), we observe that a simple CFP calculation for one model is close to a vacuum calculation for the next-complexity-level model. That is, Model A CFP calculation is very similar to Model B vacuum calculation and, likewise, Model B CFP is in general very close to Model C vacuum calculation. Although the CFP calculations include all point charges representing the atoms from the rest of the protein, only certain charges that are close enough to the fragment under consideration can influence the distribution of the electrons by perturbing the wave function and hence potentially affect the ^{15}N chemical shielding. For example, Model B CFP calculation shows on average $\sim$7 ppm deshielding of ^{15}N compared to the Model B calculation, but including only point charges for the remaining helical residues in Model C yields similar deshielding effect already (see "Model B partial CFP" calculation in Table 3.1).

3.3 Summary and conclusions

We performed density functional calculations of backbone ^{15}N chemical shielding tensor for selected α-helical residues in protein G (GB3) and compared the isotropic chemical shielding and the principal values of the shielding tensor with ex-

perimental chemical shift data. We explored the effect of electrostatic interactions in a protein on the calculated ^{15}N chemical shielding and found that :

1) To calculate the chemical shielding of backbone ^{15}N on residue i, the hydrogen bonding partners and the residues that contribute to the helical dipole need to be included. Their side chains, when not charged, do not influence much the chemical shielding of ^{15}N of interest; when charged, however, they can introduce an additional source of variation to the calculated ^{15}N chemical shielding.

2) To accurately predict ^{15}N chemical shielding when adopting charged forms of the ionizable side chains, it is important to include all charges that can potentially influence the ^{15}N chemical shielding.

3) Tensor elements calculation yields a good correlation with experiments in general and the slope is close to the ideal of -1. The variation exists mainly in the σ_{11} and σ_{33} components while σ_{22} is less affected by the charge states of the ionizable side chains that surround the amide nitrogen site.

4) CFP calculation can serve as an indicator of whether all reasonable long range electrostatic effects have been taken into consideration in the vacuum calculation. In the example of GB3, the rest of the protein does not seem to exert much effect, and the "long chain" model (Model C) in vacuum seems to be adequate for computing ^{15}N chemical shielding in the α-helix.

Note that this study focused on an α-helix, where the spatial relationship between various atoms is different from that in, for example, extended conformation. Therefore these conclusions apply specifically to helical residues and may or may not be applicable to other secondary structures. Similar calculations (see next chapter)

for the other parts of the protein are expected to address this issue. Although this study covered different residues in the helix, it is inevitably limited by the amino acid composition of GB3. Comparison with experimental data for a broad range of proteins is required in order to reach a better understanding of the accuracy of the computational approaches used here and the ways to improve them.

Table 3.2: Computed dipole moments (in Debye) for various models in the vacuum calculation

Residue	Model A neutral	Model B neutral $-CH_3$	Model C neutral $-CH_3$
A26	10.47	21.67	31.26
E27	10.40	29.41	41.95
K28	11.28	29.55	39.49
A29	9.10	28.52	39.04
F30	9.73	29.40	39.32
K31	8.22	28.52	38.58
Q32	5.73	26.24	36.86
Y33	5.71	25.23	33.98

Table 3.3: Characteristics of the calculated ^{15}N chemical shielding tensor for selected residues. The definitions of the Euler angles α, β, γ and A_1, A_2, A_3 are in (Cai et al., 2008). The principal values of the tensor are in ppm and the angels are in degrees.

Using Model C neutral

Residue	σ_{11}	σ_{22}	σ_{33}	α	β	γ	A_1	A_2	A_3
A26	14.14	155.28	197.2	-4.42	15.06	-3.19	86.81	12.83	77.58
E27	21.57	162.81	191.13	-7.67	17.16	-2.74	87.26	15.65	74.61
K28	19.52	160.76	194.91	-7.63	17.19	-0.35	89.65	7.67	82.34
A29	19.64	161.51	196.08	-14.22	14.97	-3.48	86.52	12.38	78.14
F30	25.06	160.73	193.85	-9.48	19.04	0.51	89.49	15.48	74.53
K31	7.59	163.19	187.04	-8.93	18.10	-1.69	88.31	24.32	65.75
Q32	19.23	160.88	196.49	0.02	15.73	-1.71	88.29	18.67	71.41
Y33	13.61	160.89	188.36	-6.12	17.22	-1.80	88.20	11.95	78.19

Using Model C

Residue	σ_{11}	σ_{22}	σ_{33}	α	β	γ	A_1	A_2	A_3
A26	8.60	158.04	193.56	-11.27	15.17	-2.55	87.45	17.03	73.17
E27	1.53	159.82	180.58	-17.78	16.26	-2.38	87.62	24.82	65.31
K28	26.40	160.47	198.37	-5.28	18.16	-0.42	89.58	5.53	84.49
A29	13.18	161.31	192.12	-17.44	15.04	-3.58	86.42	15.87	74.56
F30	23.32	160.14	191.58	-10.35	18.50	0.81	89.19	15.63	74.39
K31	15.74	163.45	190.33	-6.32	18.74	-1.38	88.62	20.20	69.85
Q32	14.80	159.47	193.82	-1.56	15.81	-2.29	87.71	22.35	67.78
Y33	22.96	159.13	192.26	-4.68	16.28	-1.29	88.71	9.06	81.03

Table 3.4: The effect of the side chain charges on the chemical shielding of the amide nitrogens of interest in GB3. Shown are the distances (in Å) between the center of the side chain charge (C_γ of Asp, C_δ of Glu, ϵ-amine N of Lys) of all ionizable helical residues in GB3 (indicated in the left column for each row) and the amide nitrogens of the residues under examination (A26 to Y33, indicated on top of each column). Highlighted in blue for each amide is the distance to the charge (two for Q32) that was not included in the corresponding 8-residue long-chain fragment but would be necessary in order to neutralize the charge of its salt-bridge partner. The bottom row shows the change in ^{15}N chemical shielding (in ppm) caused by the inclusion of such charges into a dipeptide model calculation. The fragment used for K28 already was neutral. Point charges of -1 (for Asp and Glu) or 1 (for Lys) were placed at the coordinate of the charge center atom in the dipeptide model. Doing so allowed us to estimate the effect of these side-chain counterions on the chemical shielding of the amide nitrogen of interest. As discussed in the main text, no significant effect was observed except for E27, in which case K31 was finally included in the (extended) Model C calculation

	A26	E27	K28	A29	F30	K31	Q32	Y33
D22	5.74	7.59	8.83	10.04	11.79	13.29	14.66	16.09
E24	7.45	7.95	7.70	9.79	11.85	12.28	13.35	15.51
E27	7.3 0	4.97	4.82	7.45	7.73	6.83	8.84	10.99
K28	7.44	7.72	6.18	7.77	10.24	10.45	10.80	13.07
K31	10.41	7.82	6.87	8.84	8.49	6.36	7.95	10.17
D36	14.91	13.90	12.15	10.14	9.4	8.81	6.46	4.88
$\triangle\delta$, ppm	0.21	4.77	N/A	1.08	0.12	-0.89	1.96	0.2

Chapter 4
Chemical shielding calculations of ^{15}N in real proteins: β-sheet residues

4.1 Computational details

The twenty-four β-sheet residues of GB3 that are under consideration were divided into three groups according to how many hydrogen bonding partners they have (see Figure 4.1). To calculate the ^{15}N chemical shielding of residue i, we included the main fragment consisting of residues i and i-1. The HB partners, both direct and indirect if applicable, were included as well. The modification of the end groups of the main fragment and the HB partners were the same as specified for α-helical residues in Ch. 3. The assignments of the basis sets[1] were the same too. We also used the same exchange-correlation functional (B3LYP) and the same atomic charges (AMBER) as in the calculations of α-helical residues in Ch. 3.

4.2 Results and discussion

Overall, the calculated chemical shielding is slightly better correlated with solid state NMR data than with solution NMR data (Figure 4.2). This can be explained by the residues whose chemical shifts differ by more than 2 ppm between solid state NMR and solution NMR measurements. The calculated isotropic chemical shielding of these residues correlate much better with solid state NMR data

[1] An exception is made for the calculation for Q2 where a 6-311+G(2d,p) basis set was used for all atoms because a sulfur atom was involved that can not be assigned a 4-21G basis set in GAUSSIAN03. For T51, since the calculation with both direct and indirect HB partners could not be completed, we omit this residue in the study

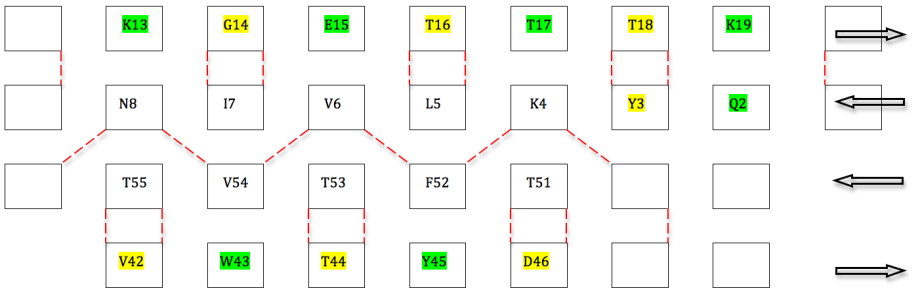
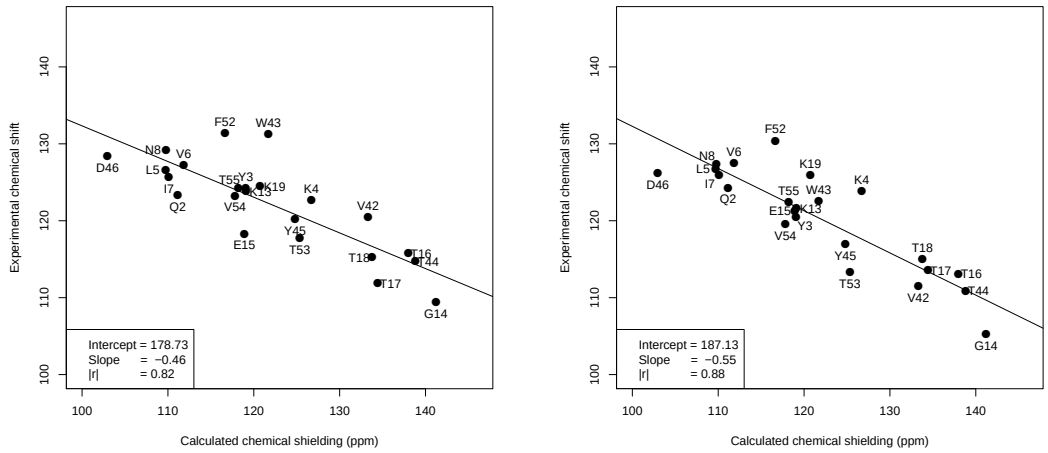


Figure 4.1: Schematic drawing of β-sheet residues in GB3 to illustrate the hydrogen bonding network. The arrows indicate the direction of each β-strand. The residues highlighted in yellow have a direct HB partner only; those highlighted in green have an indirect HB partner only; those not highlighted have both a direct and an indirect hydrogen bonding partner. Each empty box represents a hydrogen bonding partner that is not in β-sheet conformation and not included in this part of study.

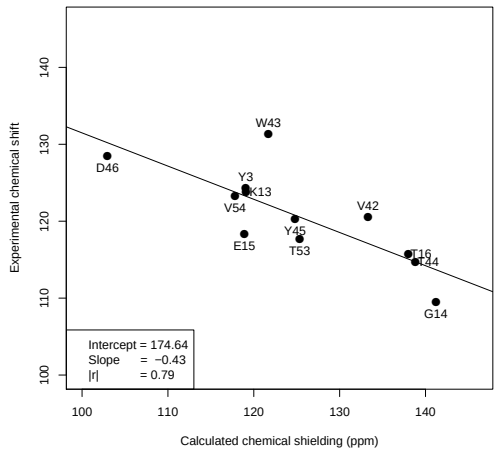


(a) Solution NMR vs calculation (b) Solid state NMR vs calculation

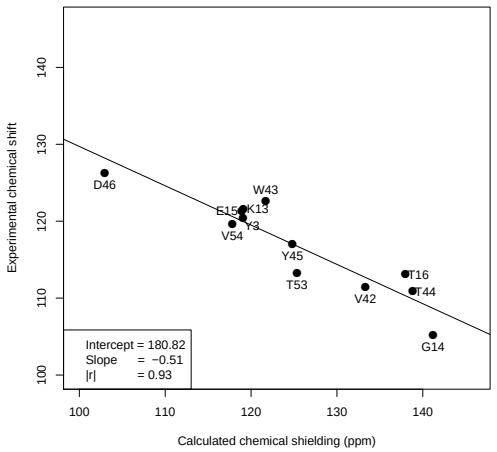
Figure 4.2: Comparison between experimental chemical shifts and calculated chemical shieldings for the β-sheet residues. The regressed line of the least squares regression is shown along with its intercept and slope in the legend box; the third number in the legend box is the absolute value of Pearson's correlation $|r|$.

(Figure 4.3 (b)) than with the solution NMR data (Figure 4.3(a)). There probably exits a more profound structural perturbation of these residues in solution so that the coordinates taken from X-ray structure provide a less valid representation of structure in solution. The complementary subset of residues do not show this kind

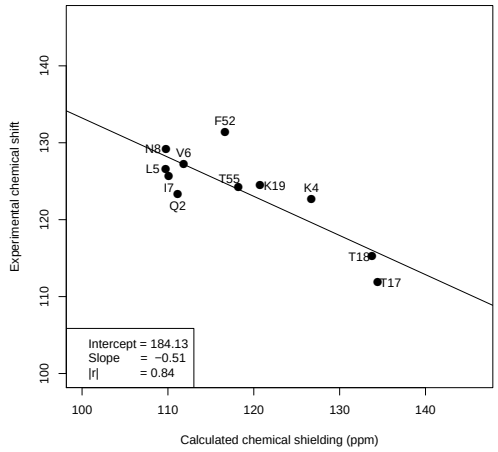
of differential correlation (Figure 4.3 (c) and (d)).



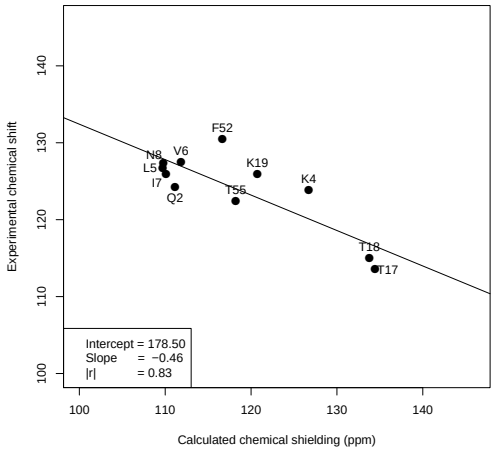
(a) Solution NMR vs calculation



(b) Solid state NMR vs calculation



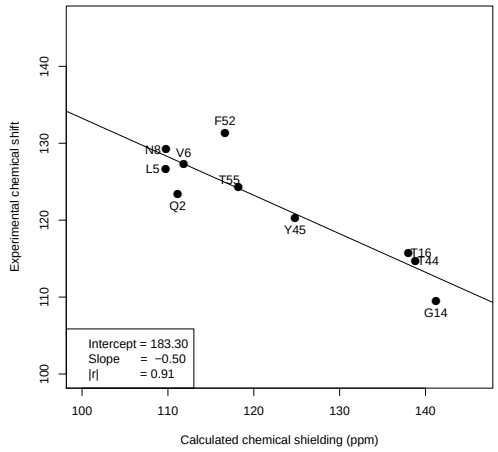
(c) Solution NMR vs calcuation



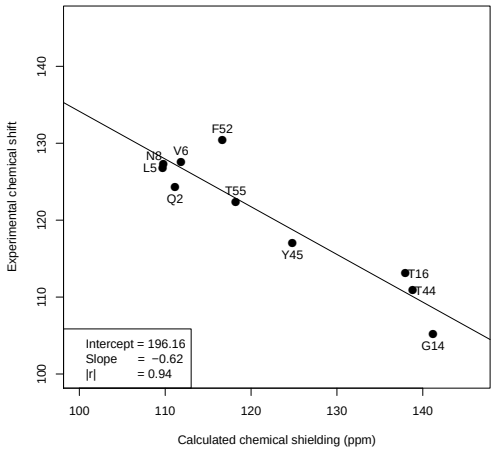
(d) Solid state NMR vs calculation

Figure 4.3: Comparison between calculated chemical shieldings and experimental chemical shifts for subsets of β-sheet residues in GB3: residues that have > 2 ppm difference ((a) and (b)) and that have < 2 ppm difference ((c) and (d)) between solid and solution NMR measurements.

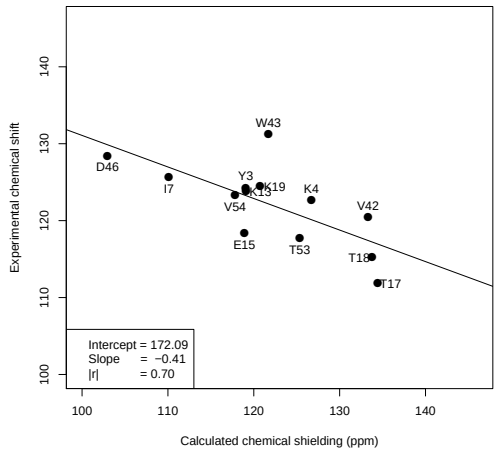
We also look at the residues whose difference between vacuum and CFT calculation is smaller than 3 ppm (Figure 4.4 (a) and (b)) and the complementary subset separately (Figure 4.4 (c) and (d)). As discussed with α-helical residues, a smaller difference between vacuum and CFT calculation is a good indication that we have



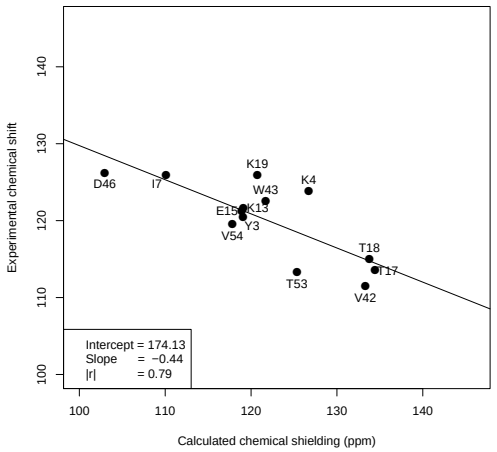
(a) Solution NMR vs calculation



(b) Solid state NMR vs calculation



(c) Solution NMR vs calcuation



(d) Solid state NMR vs calculation

Figure 4.4: Comparison between calculated chemical shieldings and experimental chemical shifts for subsets of β-sheet residues in GB3: residues that have < 3 ppm difference ((a) and (b)) and that have > 3 ppm difference ((c) and (d)) between vacuum and CFP NMR calculations.

included enough close groups in calculation and hence gives us more confidence in the results. As shown in Figure 4.4, a better correlation is indeed obtained with these residues whose difference between vacuum and CFP calculations is small.

We do notice that the slope and intercept of the regression lines are much different from what we obtained with the helix analysis. Nominally this is due to

a bigger range produced by calculation than by experiments for the residues under consideration. The structure that we take from X-ray structure is probably not a good average structure and are not able to represent well the thermal mobility of atoms (even in solid state). Whether this mobility is correlated with secondary structure may be worthy of further investigation. It has been previously suggested by many researchers that the poorly described χ torsion angles of the side chains are responsible for less satisfactory agreement between experimental and quantum mechanical theoretical chemical shifts for ^{13}C (Villegas et al., 2007). The arguments hold similarly for ^{15}N as Le and Oldfield (1996) demonstrated the effects of torsion angles of the residues *i* and *i*-1 on the ^{15}N chemical shielding of residue *i*.

Another reason that the sheet structure does not produce a more accurate agreement is probably because the model has not included enough close contacts. In an earlier paper (Xu and Case, 2002), when building a model for calculating residues in β-sheet structure, a system was designed as a 3-5 residue sequence along with their duplex (with one β-strand running on one side) and triplex forms (with two β-strands running on both sides). This is reasonable because usually one β-strand is accompanied by other β-strands in parallel and/or antiparallel fashion (also see Figure 4.1). Apparently this design of model would have more electrostatic effects taken into account. As demonstrated in our calculation for T17, after bringing another residue K4 into calculation, a deshielding of 6 ppm was induced. In this case, K4 is neither directly or indirectly hydrogen bonded with T17 but rather opposing T17 from another sheet strand. Modifying the side chain of K4 to be neutral yields similar results so it was not the charge of K4 that was causing this

effect.

4.2.1 The tensor

The principal chemical shielding components of the β-sheet residues were also plotted against the experimental chemical shift counterparts. The overall agreement with the GB1 data is still good (Figure 4.5). The regression line has a slope close to an ideal of -1 and the intercept is about 10 ppm less than what was obtained above for α-helical residues.

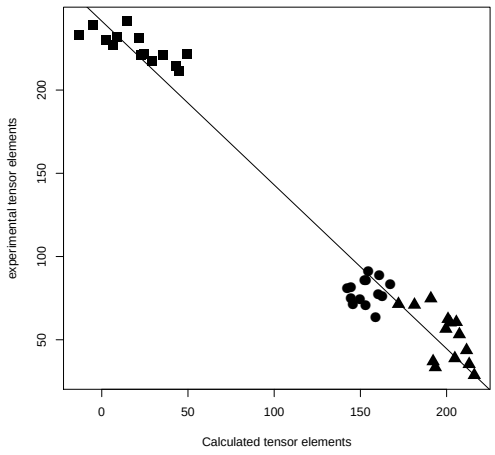


Figure 4.5: Comparison of the principal components of the calculated chemical shielding tensor for GB3 with the corresponding components of the experimental chemical shift tensor for GB1 (Wylie et al., 2007). Squares, circles, and triangles are for the principal elements σ_{11}, σ_{22} and σ_{33} respectively. The parameters of the regression line are: intercept $= 241.43 \pm 4.32$ ppm, slope $= -0.98 \pm 0.03$. Plotted here are residues L5, G14-T18, W43-D46 and F52-T55 for which GB1 shares the sequence and meanwhile the experimental data for GB1 is available

4.3 Summary and conclusions

We performed calculations for twenty-three β-sheet residues of the protein GB3 both in vacuum and using CFP. Comparing the isotropic chemical shieldings with experimental chemical shifts, we found that overall, the calculated chemical shielding is slightly better correlated with solid state NMR data than with solution NMR data. It is informative to divide these β-sheet residues in subsets and then compare their corresponding correlations between experiments. Hence it was found that the subset of residues whose experimental shift difference between solid-state and solution is significant tends to correlate better with solid-state NMR data. Arguably this is because these residues undergo more profound structural perturbation in solution. Also it was found that the subset of residues whose vacuum calculation has a smaller difference from its CFP calculation tends to have a better correlation with experimental data when using the vacuum results. This reinforces the message that the CFP calculation can help to determine whether or not enough electrostatic interactions have been taken into account. Lastly, we showed that the tensor elements correlate very well with the experimental counterparts. The regression slope is close to ideal, although the intercept is a little different than what we obtained from α-helical residues in the previous chapter. Statistics are necessary to get calibrated slope and intercept for the purpose of converting chemical shieldings to chemical shifts.

Chapter 5

Modeling chemical shift perturbations in diubiquitin

In this chapter, we study protein complexes. Since proteins perform their function via interaction with other molecules, structures and dynamics of protein complexes can greatly enhance our knowledge about the biological function of proteins. While x-ray crystallography is a very powerful tool for solving structures of monomeric proteins and tightly bound complexes, this method has very limited applicability in the case of weak interactions and transient complexes (Dominguez et al., 2003; McCoy and Wyss, 2000), which are however at the heart of a variety of cellular processes. In fact, these transient complexes are often not even amenable to crystallization. Traditionally, NMR approach to solve protein-protein complex requires the collection of intermolecular nuclear Overhauser effect (NOE) distances (Dominguez et al., 2003). This process is usually long and can be difficult in many unfavorable cases. In these difficult situations, chemical shift perturbation (CSP) data has long been a unique source of information because of the sensitivity of NMR chemical shifts to local electronic changes. Experimentally for example (McCoy and Wyss, 2000), amide proton and nitrogen chemical shift perturbations can be easily detected in heteronuclear single quantum coherence (HSQC) spectra of ^{15}N -labeled proteins interacting with unlabeled ligands or proteins. Chemical shift differences between free and bound protein are quantified as a weighed average of ^{1}HN and

^{15}N chemical shift perturbations. The precise value of these normalization factors (weights) is immaterial and we should view the quantitation as a statistical measure only: the results are most significant when a contiguous surface patch of shifting perturbation is obtained (Zuiderweg, 2002).

The so-called chemical shift perturbation mapping of the interaction interfaces not only allows mapping of the molecular surfaces involved in protein-ligand interactions, but also provides a means of quantifying the affinity of binding. A recent application of CSP data includes molecular docking based on the ambiguous constraints derived from CSP mapping (Dominguez et al., 2003) which offers unprecedented ability to determine structures of protein complexes in those cases (weak interactions) when traditional, NOE-based information on intermolecular contacts is scarce or even unavailable. More interestingly in terms of the quantitative use of CSP's, we have seen applications in the docking of protein-ligand complex using CSP's alone (McCoy and Wyss, 2000) and of protein-protein complex in combinations with residual dipolar coupling data (McCoy and Wyss, 2001). In these two situations, the proton differences between experimental and simulated CSP's were minimized to produce ligand alignment or to restrain protein-protein complex. The simulated proton chemical shifts were obtained by *SHIFTS* (David Case, Scripps), which computes proton chemical shifts from empirical formulas. Even though experimental ^{15}N CSP's can be large, it was not yet used in such procedures due to limited understanding of ^{15}N chemical shifts. Our ability to fully understand and accurately model the relationship between the structure and chemical shift will dramatically improve the use of chemical shifts and their changes upon forming complexes for

structure determination purposes.

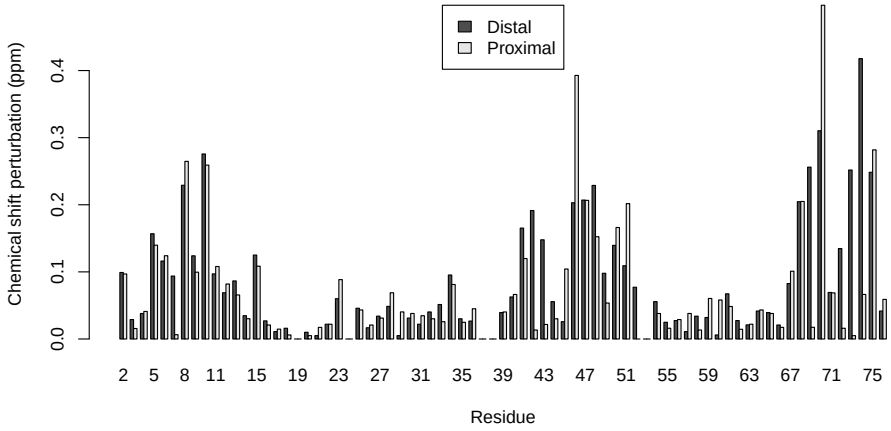
In this chapter, we try to model ^{15}N chemical shift perturbations of the protein ubiquitin, which can be linked with another ubiquitin by forming an isopeptide linkage between the C-terminus of one ubiquitin and a specific lysine in the other ubiqutin. Here we study the K48-linked diubiquitin. The second ubiquitin has a free C-terminus and is called the proximal domain, while the first ubiquitin is called the distal domain.

The backbone amide chemical shifts for monomeric Ub_1 and dimeric Ub_2 from the ^{1}H-^{15}N HSQC spectra were used to compute, for each residue, the combined amide chemical shift difference between Ub_1 and Ub_2:

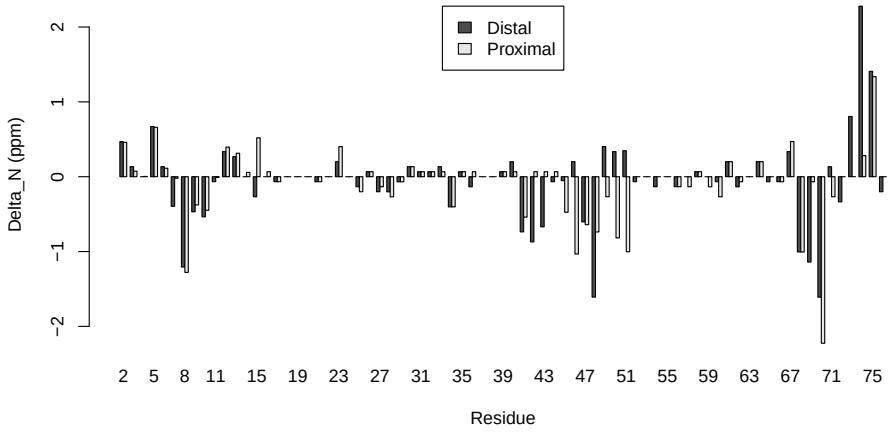
$$\delta \triangle \sigma = [(\triangle\delta_{\mathrm{H}})^2 + (\triangle\delta_{\mathrm{N}}/5)^2]^{1/2}, \tag{5.1}$$

where $\triangle\delta_{\mathrm{H}} = \delta_{\mathrm{H}}(\mathrm{Ub}_2) - \delta_{\mathrm{H}}(\mathrm{Ub}_1)$ and $\triangle\delta_{\mathrm{N}} = \delta_{\mathrm{N}}(\mathrm{Ub}_2) - \delta_{\mathrm{N}}(\mathrm{Ub}_1)$ are the corresponding ^{1}H and ^{15}N chemical shift perturbations (plotted in Figure 5.1). This is the length of the vector that connects the two-dimentional end points on the ^{1}H-^{15}N HSQC spectra for each residue, normalized by typical ^{1}H and ^{15}N chemical shift range. The easily spotted contiguous surface patch around residues L8, I44 and V70 for both domains is significant (Figure 5.1(a)). These residues are also hypothesized to form hydrophobic interactions in the closed dimer form as illustrated in Figure 5.2.

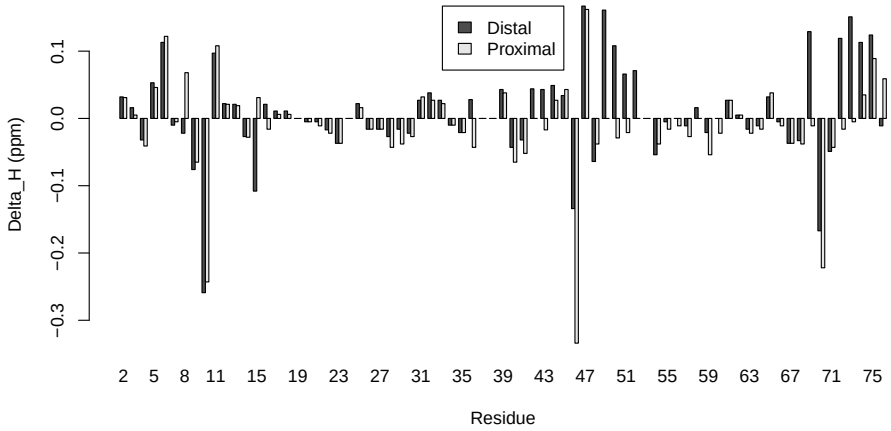
The solution NMR study of Varadan et al. (2002) showed that the ubiquitin dimer conformation is dynamic and pH sensitive. It can switch from open state to closed state with increasing pH. At pH 4.5 the open conformation is fully populated,



(a)



(b)



(c)

Figure 5.1: The chemical shift differences (a), ^{15}N (b) and ^{1}H (c) chemical shift perturbations between dimeric and monomer ubiquitin for all residues. The experimental NMR data for both the monomer and the K48-linked dimer was obtained in pH 6.8 solution and provided kindly by Ming-Yi Lai. Please note: the data for: P19, E24, P37, P38 and G53 are not available.

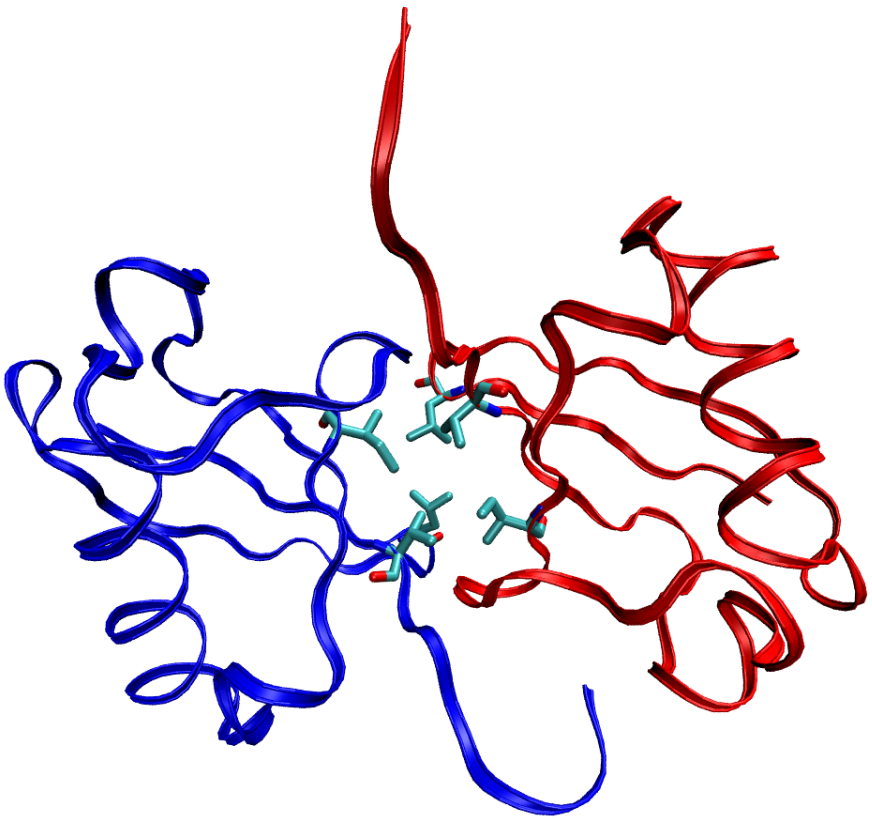


Figure 5.2: Hydrophobic interface of 1AAR. The blue ribbon represents the distal domain while the red one represents the proximal domain.

and at pH 7.5 the closed conformation is almost fully populated. The chemical shift perturbations near the hydrophobic residues L8, I44 and V70 are bigger in neutral to basic solution where closed conformation is supposed to dominate (Ryabov and Fushman, 2006; Varadan et al., 2002). It was also estimated that the population of the open conformation obtained for the individual amide groups at pH 6.8 ranges from less than 1% to 25% (Varadan et al., 2002).

Currently, one cannot use chemical shift changes to predict what exactly happens at the interface, but with theoretical calculations, one may hope to gain some perspective. To be able to calculate the chemical shift perturbation at pH 6.8, ideally we should take structures of monomer ubiquitin and dimeric ubiquitin obtained at the same pH condition and then compute ^{15}N chemical shielding differences using these structures. Instead, we only have them from different pH: the crystal structure of monomer ubiquitin 1UBQ from pH 5.2-5.8 and the crystal structure of dimeric ubiquitin 1AAR from pH 4.0-4.4 (Table 5.1). On the other hand, there are related tetraubiqutin structures available (Table 5.1). Among them, 2O6V can be viewed as a dimer of diubiquitin and each of the "diubiquitin" is nearly identical to the structure of 1AAR with hydrophobic patches buried in the same fashion. Other tetramer structures, 1F9J and 1TBE, were both obtained at acidic pH. For 1TBE, the hydrophobic patches are exposed while for 1F9J, the hydrophobic patches are buried in a different arrangement than 2O6V.

Table 5.1: The available PDB entries of monomeric, dimeric and tetrameric ubiquitin

PDB ID	Resolution(Å)	pH	State	Reference
1UBQ	1.8	5.2-5.8	monomer	Vijay-Kumar et al. (1987)
1D3Z	N/A	6.6	monomer	Cornilescu et al. (1998)
1AAR	2.3	4.0-4.4	dimer(closed)	Cook et al. (1992)
1TBE	2.4	5.0	tetramer(open)	Cook et al. (1994)
1F9J	2.7	4.8	tetramer(closed)	Phillips et al. (2001)
2O6V	2.2	6.7	tetramer(closed)	Eddins et al. (2007)

5.1 Computational details

Here, we try to model the ^{15}N chemical shift perturbations using density functional theory. The computational methods are the same as in previous chapters Ch. 3 and Ch. 4. Only vacuum calculations were performed.

Calculations were performed for the α-helical residues from A26 to D32 in both monomer and dimeric ubiquitin. The monomer crystal structure of 1UBQ and the dimer crystal structure of 1AAR were used.

Calculations were also performed for residues L8, I44 and V70 from both domains of 1AAR. A dipeptide model calculation was performed, along with a more complex model calculation that includes not only the dipeptide but also fragments of some residues from the hydrophobic pocket. The residues whose closest side chain heavy atoms are within 4.5 Å range of at least one side chain heavy atom of the residue of interest are included. The details of what residues are included in the model calculation for L8, I44 and V70 are listed in Table 5.2.

Table 5.2: The diagonal matrix designating the residues (other than the residue of interest itself) included as extra fragments in addition to the dipeptide fragment for the calculation of the hydrophobic residues in the closed form of diubiquitin. 1 represents inclusion while 0 the opposite.

	D:L8	D:I44	D:V70	P:L8	P:I44	P:V70
D:L8	0	0	1	1	1	0
D:I44	0	0	1	1	0	1
D:V70	1	1	0	0	1	0
P:L8	1	1	0	0	0	1
P:I44	1	0	1	0	0	1
P:V70	0	1	0	1	1	0

5.2 Results and discussion

5.2.1 α-helical residues

Monomer ubiquitin. The correlation of calculated chemical shieldings with solution NMR data from either pH 5.7 or pH 4.5 is good when excluding D32, an obvious outlier (see Figure 5.3(a) and (b)). But the slope of the regression line is far from the ideal of -1. The comparison with the solid state NMR data identifies two outliers of D32 and K27. Otherwise, the rest of residues show a good correlation ($|r| = 0.95$) and the slope of regression is improved to be -0.67 (Figure 5.3(c)). This improvement in slope is due to a wider range (by $\sim$ 3 ppm) of isotropic chemical shifts in solid state NMR over these α-helical residues (see Table 5.3 and Figure 5.3). In solid state, protein motions are more restricted while in solution, there is the effect of protein dynamics that may somehow average out the solution chemical shifts. It is worth noting, however, that over all the residues of ubiquitin, the shielding ranges obtained by solution and solid state NMR are comparable (Figure 5.3(d)).

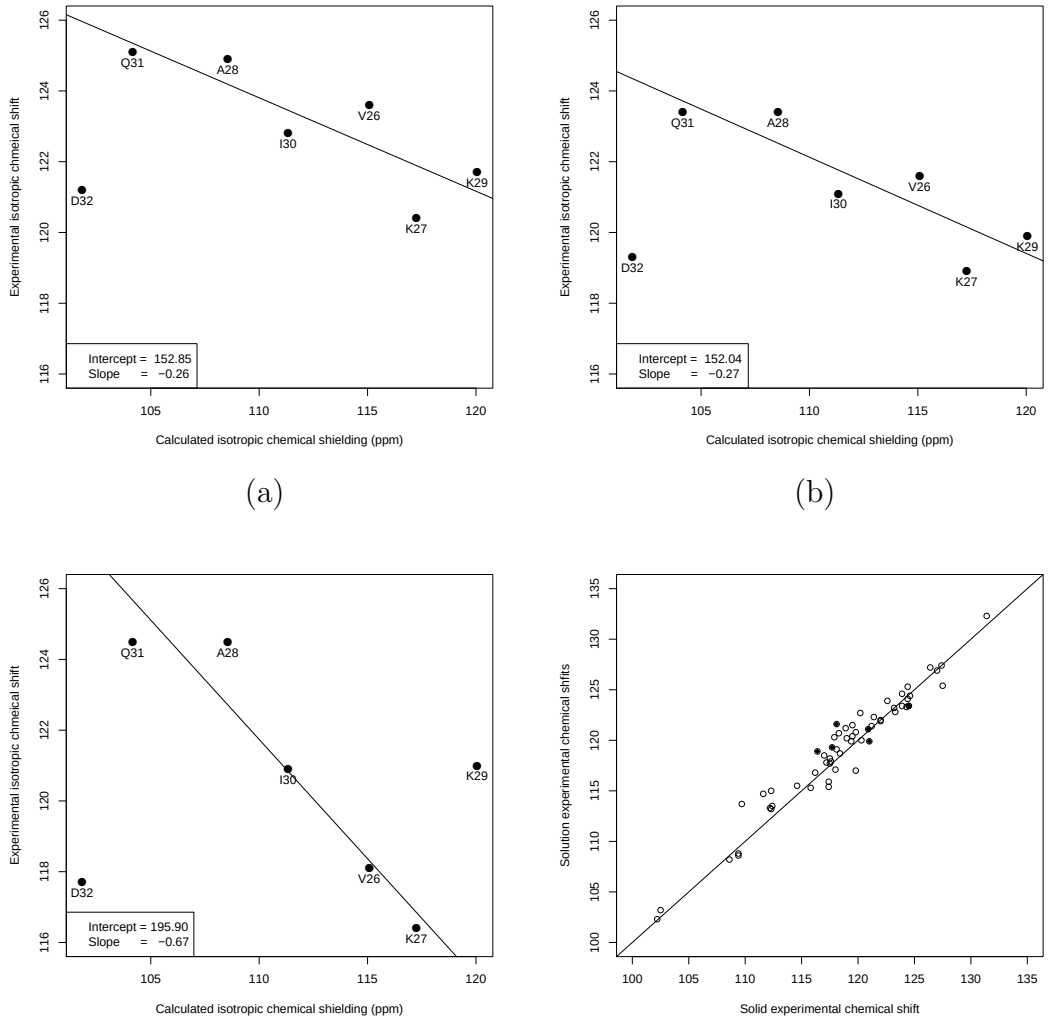


Figure 5.3: Comparison of experimental ^{15}N chemical shifts for ubiquitin measured in: (a) pH 5.7, (b) pH 4.5 and (c) solid state vs calculated chemical shieldings. The regressed line of the least squares regression is shown along with its intercept and slope in the legend box. In panel (d) is shown the comparison of the ^{15}N solid state NMR and solution NMR at pH 4.5. The diagonal line is plotted to guide the eye. The α-helical residues from A26 to D32 are plotted with solid circles.

Diubiquitin. Although experimentally, the measured ^{15}N chemical differences between the distal and proximal domain are quite small for the six α-helical residues in consideration (see Figure 5.1), the difference can be as big as 7 ppm for calculations (Table 5.3). This is probably because the crystal structure is not

representative of the ensemble in solution.

Table 5.3: Calculated isotropic chemical shiedings and various experimental isotropic chemical shifts for selected α-helical residues (in ppm).

	Calculated results			Experimental data for monomer		
Residue	1UBQ	1AAR (Distal)	1AAR (Proximal)	Solid	Solution (pH5.7)	Solution (pH4.5)
V26	115.08	113.95	111.88	118.1	123.6	121.6
K27	117.25	109.13	109.39	116.4	120.4	118.9
A28	108.55	99.33	96.23	124.5	124.9	123.4
K29	120.05	122.52	125.34	121.0	121.7	119.9
I30	111.33	105.03	110.40	120.9	122.8	121.1
Q31	104.16	104.33	111.48	124.5	125.1	123.4
D32	101.83	107.15	113.09	117.7	121.2	119.3

5.2.2 The hydrophobic residues: L8, I44 and V70

Upon dimerization, the solvent effect on the interface residues may be attenuated as the water environment is replaced by the residues from the other ubiquitin. Since water usually deshields ^{15}N , as we discussed in Ch. 2, the replacement of the solvent will cause shielding of ^{15}N for these interface residues. This may be the reason that the interface residues near L8, I44 and V70 are more shielded (with a smaller chemical shift value) upon dimer formation (see Figure 5.1).

Generally, when we compare the results from the dipeptide model and the more complex model with selected surrounding hydrophobic residues, the difference is small, indicating that the hydrophobic residues in the pocket does not seem to affect the ^{15}N isotropic chemical shielding much. However, the exception is for P:V70 where the 3 ppm deshielding effect induced by the nearby residues of D:I44,

P:L8 and P:I44 stands out (Table 5.4)[1]. If we look closely at the ^{15}N chemical shift perturbations, P:V70 has the biggest absolute CSP among the six (while D:I44 and P:I44 have the smallest absolute CSP). This may be not a coincidence. There could be one reason for both.

Table 5.4: Calculated chemical shieldings (ppm) for the residues in the hydrophobic patch.

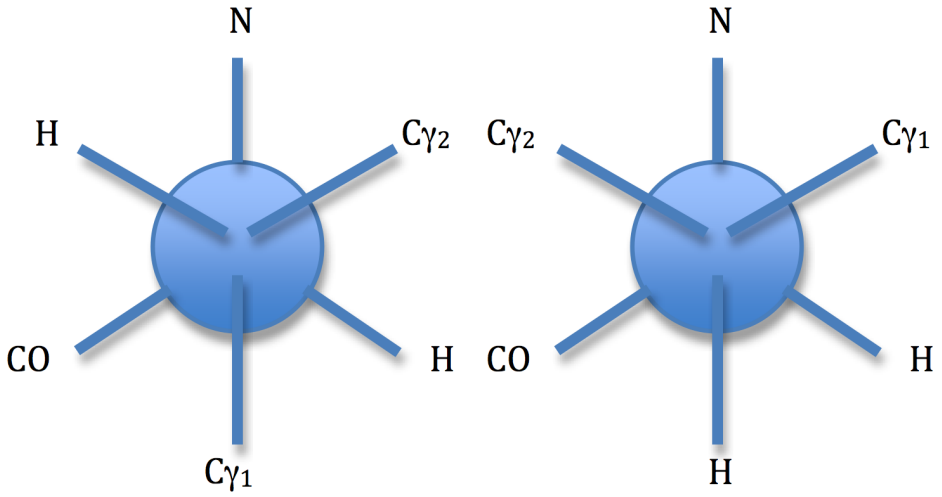
Residues	Dipeptide	plus more residues
D:L8	111.75	112.49
P:L8	106.91	106.75
D:I44	121.41	120.84
P:I44	131.85	131.36
D:V70	115.22	113.05
P:V70	129.70	126.57

The big calculated difference between D:V70 and P:V70 as seen from Table (5.4) is mainly due to different χ_1 angles. In 1AAR, the χ_1 angle of V70 is 152.07° in the distal domain and −67.09° in the proximal domain. Using a dipeptide model and modifying the χ_1 angle of D:V70 to be the same as P:V70 introduces about 17 ppm shielding and brings the isotropic chemical shielding closer to that of P:V70. The χ_1 angle effect calculated here is much bigger than the previous report of about 7 ppm by Le and Oldfield (1996).

In all ubiquitin crystal structures that we worked with, be it monomer, dimer or tetramer, V70's side chain adopts only two conformations (see Figure 5.4): (i) the favorable conformation where $C_{\gamma 1}$ is in *trans* position and $C_{\gamma 2}$ in the *gauche*(+) position; (ii) one unfavorable conformation where $C_{\gamma 1}$ is in *gauche*(+) and $C_{\gamma 2}$ is in

[1]P and D denotes proximal and distal domain respectively. Hence the notation P:V70 designates the residue V70 from the proximal domain; similarly, D:I44 designates the residue I44 from the distal domain

the *gauche*(−) position. The second conformation is not stable because the atom $C_{\gamma 2}$ is in close contact with the main chain CO and NH groups. The favorable conformation is observed in monomer ubiquitin structure (1UBQ and 1D3Z), tetramer open conformation structure (1TBE) and the distal domains of the closed dimer (1AAR) and closed tetramer (2O6V and 1F9J) structures. Interestingly, the unfavorable conformation is also observed. But it is only observed in the proximal domains of closed dimer (1AAR) and closed tetramer (2O6V and 1F9J) structures. It is hardly simply by chance to observe such unfavorable conformation in three crystal domains. A closer look into the structures reveals that the conformations which D:V70 and P:V70 adopt differentially actually enable both side chains to be better buried in the hydrophobic pocket. It is very likely that the unfavorable conformation of P:V70 is compensated by hydrophobic interaction.



(a) (b)

Figure 5.4: The favorable (a) and unfavorable (b) side chain conformations observed for V70 in monomeric, dimeric and tetrameric ubiquitin.

Taking into account protein dynamics, it is reasonable to think that in closed

conformation, D:V70 on average spends more time in the favorable conformation and P:V70, on the contrary, spends more time in the unfavorable conformation. The unfavorable conformation seems to have an effect of shielding the ^{15}N atom of V70 when we take the tetramer structure 2O6V for calculation (See Table 5.5), consistent with the calculation with 1AAR as discussed above. Even the closed form of diubiquitin is not 100% populated, we would still expect P:V70 to have a smaller shift value than D:V70 and this is indeed in accordance with the solution experimental data (See Figure 5.1). There is no significant difference between ^{15}N shielding calculations for D:I44 and P:I44 (Table 5.5), agreeing with experiment qualitatively as well.

Table 5.5: Calculated chemical shieldings (ppm) for I44 and V70 using 2O6V

Residues[1]	First dimer	Second dimer
D:V70	117.11	116.51
P:V70	123.95	125.10
D:I44	123.24	122.49
P:I44	120.26	121.78

1. The calculation was performed using a dipeptide and a fragment from the direct HB partner

5.2.3 New hydrogen bonds upon dimerization

One structural change upon dimerization is that of forming hydrogen bonds between the two domains. Observing the structure of 1AAR, we found some inter-domain hydrogen bonds and illustrate them in Figure 5.5. For example, Q49 is directly hydrogen bonded to L71 of the other domain, which makes it an indirect HB partner of R72.

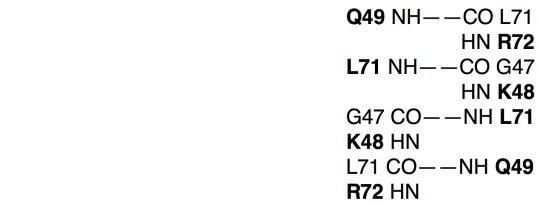


Figure 5.5: Schematic drawing for the hydrogen bonding network between the the distal and proximal domains in closed form diubiqutin. Shown in bold are the residues whose ^{15}N chemical shieldings were calculated.

The calculated results of Q49 and L71 with and without their direct HB partners L71 and G47 respectively from the other domain are shown in Table 5.6, along with the calculations of K48 and R72 with and without their indirect HB partners L71 and Q49 respectively from the other domain. The HB deshielding effect is most notable for P:K48 (about 7ppm). Meanwhile, the difference between P:K48 and D:K48 is as big as 18.92 ppm. From Figure 5.1(b), we can see that the experimental difference between the K48's from the two domains are also notable but at a significantly smaller scale. This is hard to interpret as some solvent is replaced by weak hydrogen bonding as the dimer opens and closes.

Table 5.6: Calculated chemical shieldings (ppm) for the residues that are hydrogen bonded with the other domain

Residue	dipeptide	with direct HB	Residue	dipeptide	with indirect HB
D:Q49	106.42	106.08	D:K48	119.14	118.45
D:L71	107.57	107.07	D:R72	118.38	114.88
P:Q49	106.06	104.52	P:K48	106.78	99.53
P:L71	114.19	113.44	P:R72	127.35	122.23

5.3 Summary and conclusions

In this chapter, we try to account for the ^{15}N chemical shift difference between monomer and dimeric ubiquitin. As the monomer and dimer structures used were obtained from different pH conditions, we feel that it makes better sense to look at chemical shift differences between the distal and proximal domains. We learned the following:

1) Although solid-state NMR and solution NMR of ubiquitin covers the same shielding range over all residues, zooming in on the classical α-helical residue region (from A26 to D32), we observed the shift range is shortened in solution NMR measurements and hence a poorer correlation to theoretical calculations resulted.

2) Calculations were not able to reproduce ^{15}N chemical shift differences between monomer and dimeric ubiquitin for the α-helical residues, by using the crystal structures 1UBQ and 1AAR.

3) Among the residues in the hydrophobic interface L8, I44 and V70, V70 from the proximal domain seems to sense the presence of other hydrophobic residues on the interface most strongly. It also happens to be the one with the biggest ^{15}N chemical shift perturbation.

4) Calculations were not quite able to reproduce the chemical shift differences between the distal and proximal domains by using static crystal structures. In addition, the nature of the protein being dynamic makes occupancy of certain conformations a problem as well. However, the chemical shift difference between the two domains for the residues investigated here quantitatively agrees with the experiment

in some cases even though they are significantly larger than what were observed experimentally. This can help to understand the reason underlying chemical shift perturbations, such as flexible side chain orientation or formation of inter-domain hydrogen bonds.

Appendix A

Ring effects

In assessing the potential ring effect that can be exerted on ^{15}N chemical shielding, we can work with a model of NMA and benzene cluster. First, we can perform a chemical shielding calculation with an optimized geometry of the cluster, then we can isolate the NMA molecule from the cluster and perform another chemical shielding calculation on the NMA molecule only. The difference of the two calculations for ^{15}N isotropic chemical shielding is assumed to be caused by the ring effect.

Two optimized cluster geometries (Bendova et al., 2007) were considered here, with **starting** geometries to be T-shaped and stacked respectively. The optimized geometry are presented in Figure A.1.

(a) (b)

Figure A.1: Optimized geometries of NMA+Benzene clusters. The starting geometry for (a) was T-shape orientation and that for (b) was stack orientation

In each optimized geometry, the amide nitrogen is located right above the center of the benzene molecule. In the T-shape orientation, the peptide plan is

almost perpendicular to the benzene ring while in the "stack" orientation (which is not stacked any more after geometry optimization), the peptide plan is slightly tilted. The distance of the nitrogen atom to the center of the benzene ring is 3.23 Å and 3.19 Å respectively for the two orientations. Since the two geometries are very similar, we present here only the calculations for the T-shape oriented cluster.

We performed calculations with different levels of theory (see Table A.1). Although MP2 calculations for the clusters using bigger basis sets were not able to be carried out on our computing facility due to memory limits, it was however easy to induce from the calculations which did carry out, that the ring effect on ^{15}N isotropic shielding is marginal, regardless of the level of theory used (see also Figure A.2).

Table A.1: Calculated ^{15}N chemical shieldings (ppm) for NMA and NMA+benzene from various levels of thoery

	NMA			NMA+Benzene (T-shape)		
Basis sets	HF	B3LYP	MP2	HF	B3LYP	MP2
1). 6-31G	181.00	158.79	194.01	183.18	159.47	194.57
2). 6-31G(d)	182.19	159.37	190.7	184.74	160.16	191.23
3). 6-31G(d,p)	183.95	161.41	191.05	186.13	161.93	NA
4). 6-31+G(d,p)	182.67	160.76	190.31	184.87	161.13	NA
5). 6-311+G(d,p)	165.94	142.78	171.76	167.79	142.48	NA
6). 6-311+G(2d,p)	164.59	141.75	169.02	166.42	141.49	NA
7). 6-311++G(2d,p)	164.62	141.88	169.10	166.48	141.58	NA

In GB3, we take the peptide plane that the amide nitrogen of E27 belongs to and modified it to be a NMA molecule; and take the benzene group of residue F52's side chain to build a benzene molecule. This way, we formed a new cluster and examined the ring effect. In this case, the amide nitrogen is 4.91 Å away from the center of the benzene ring. One would not anticipate any ring effect on amide

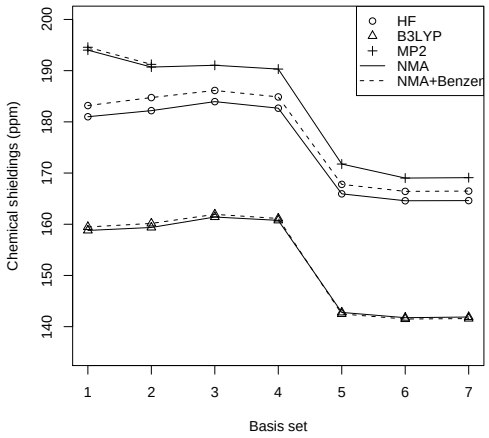


Figure A.2: Calculated ^{15}N isotropic chemical shieldings in NMA and NMA+Benzene cluster from various levels of theory

nitrogen from a benzene ring at such distance based on the model calculations listed above. Indeed, the calculation of an isolated NMA molecule and of a NMA+Benzene cluster yielded almost identical isotropic chemical shielding for the amide nitrogen. Furthermore, statistics show that in GB3, any amide nitrogen that is adjacent to the side chain of either a Phe or a Trp and hence potentially subject to ring effect, has a distance further away (usually >4 Å) from the ring center. Based on this, we can safely ignore the ring effect when calculating amide nitrogen chemical shielding in protein GB3.

This theoretical calculation supports the common treatment where ring effect was not considered (de Dios et al., 1993b). On a side note, an observed 2.7 ppm shielding effect from the ring was observed for amide hydrogens, which are about 2.3 Å away from the center of the benzene ring. No ring effect was observed for hydrogens from the two methyl groups of the NMA molecule, which are further out.

Appendix B

MATLAB code

The following is the MATLAB code that we wrote for calculating Euler angles as defined in (Brender et al. 2001) and some characteristic angles as defined in (Cai et al. 2008).

```
% Calculating the tensor priciple components:
sym = (tensor + tensor')/2; % Symmetrize the calculated tensor
[V,D] = eig(sym);
% Sxx > Syy > Szz in chemical shielding:
Sxx = V(:,3); Syy= V(:,2); Szz= V(:,1);
% Define vectors with N, H, CA, CO being coordinates for N, H, alpha-
% carbon and CO  respectively:
NH = H - N;
NCA = CA - N;
NCO = CO - N;
norm = cross(NCA,NCO); % The normal to the peptide plane.
% Make the principle directions comply to that defined in
% (Brender et al. 2001),reversing vector direction if necessary

if dot(norm,Syy)<0
```

```
        Syy = -1*Syy;
end
if dot(Sxx,NCO)>0
        Sxx = -1*Sxx;
end
% The third axis is defined after the first two principle axes:
        Szz = cross(Sxx,Syy);
% Euler angels defined as in (Brender et al. 2001):
beta = angle1_ref(Szz,NH,norm);
gamma = angle2(NCA,NCO,Szz);
alpha = angle3(Sxx,Syy,NH);
% Characteristic angles defined as in (Cai et al. 2008)
A1 = angle1(norm,Szz);
A2 = angle1(norm,Syy);
A3 = angle1(norm,Sxx);
function ang = angle1(x,y) % output range is [0,90]  degrees for the
% angle between the two lines defined by the two vectors x and y
a = 180/pi* acos (dot(x,y)/dot(x,x).^0.5/dot(y,y).^0.5);
if a > 90
    ang = 180 - a;
elseif a <= 90
    ang = a;
end
```

```
function ang=angle2(x,y,z)  % output range is [-90,90] degrees for the
% angle between vector z and the plane defined by vectors x and y.
% If above the plane, output is positive; otherwise, output is negative.
  norm = cross(x,y);
  ang = 90 - angle1(norm,z);
  if dot(norm,z) < 0
     ang = -1* ang;
  end
function ang = angle3(x,y,z) % Orthogonal vectors x and y define a
% plane XY. This function returns the angle that the projection of
% vector z onto the plane XY need to turn clockwise in order  to overlap
% with the axis x. The projection must be in either the first or fourth
% quadrant for this particular Euler angle calculation: if first
% quadrant, the output angle is positive; if fourth quadrant, the
% output angle is negative
norm = cross(x,y);
      proj = z - dot(z,norm)/dot(norm,norm)*norm;
      a = dot(proj,x); b = dot(proj,y);
      if sign(a)>0
       ang = sign(b)*angle1(x,proj);
      else ang=0; % giving an error message signaled by a zero value
% if projection is not in the first or fourth quadrant
      end
```

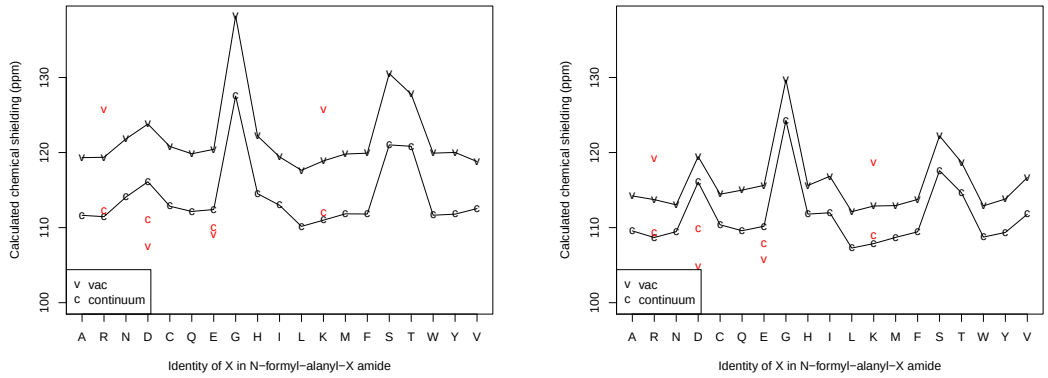
```
function ang = angle1_ref(x,y,ref) % The angle vector x will

% turn clockwise to overlap with vector y if seen from the end of

% vector ref

ang = 180/pi* acos (dot(x,y)/dot(x,x).^0.5/dot(y,y).^0.5);

if dot(cross(x,y),ref) < 0

   ang = -1*ang;

end
```

Appendix C

E27 in GB3: a difficult case for DFT calculations

It is observed from Figure C.1 that with a dipeptide model, for Lys (K) and Arg (R), the shielding effect of protonation is apparent (about 6 to 7 ppm) in the vacuum calculation but not in the continuum-only calculation; for Asp (D) and Glu (E), the deshielding effect of deprotonation is profound in the vacuum calculation (with effect in D >E), and yet the continuum-only calulation still shows such deshielding effect (with the effect in D > E as well). In vaccum, the deshielding effect of the negative charges on Asp and Glu are bigger than the shielding effects of the positive charges of Lys and Arg; in the continuum, the charge effects from Lys and Arg's side chains are more effectively screened by the solvent water. These are all probably due to the longer side chain of Lys and Arg which makes the charge center and their amide nitrogen further away from each other.

Since there is only a slight shielding effect due to the solvent when the Glu's side chain is in the charged state (compare "charged vacuum" and "charged continuum" calculations in Figure C.1), it is probably unrealistic to expect E27 in GB3 to be charged while counting on the solvent to counteract the deshielding effect of its negative charge. However as shown in Ch. 3, the residue E27 needs about 10 ppm more shielding to be not outlying. This analysis hence also suggests that E27 in GB3 take on a neutral side chain.



(a) (b)

Figure C.1: Calculated isotropic chemical shieldings for each residue type in vacuum and continuum with a dipeptide model. (a) is for α-helix conformation and (b) is for β-sheet conformation. In red are calculated results with ionizable side chains charged. The dipeptide geometry is the same as used in Ch. 2.

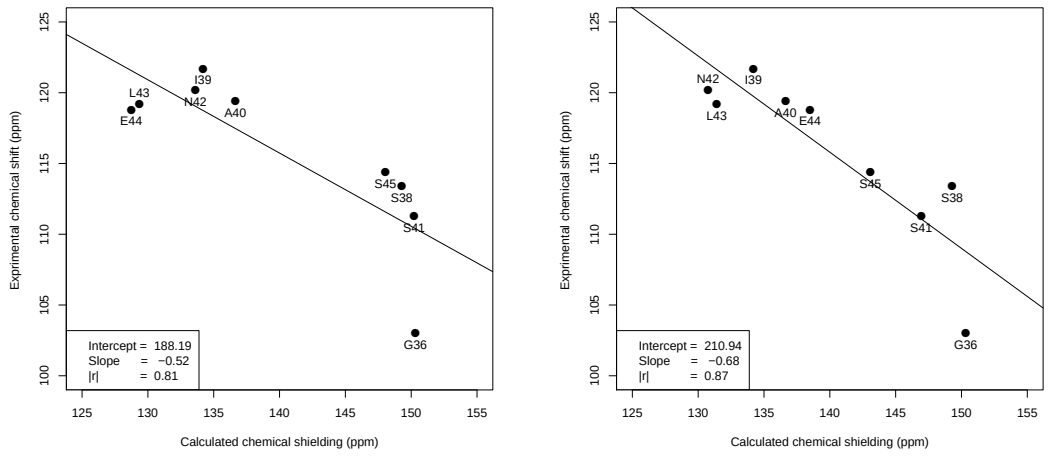
Although here it shows that Lys and Arg's charges are screened effectively by solvent in the continuum calculation, this is an extreme case where Lys and Arg, in a small dipeptide, are fully exposed to solvent with dielectric constant ϵ=80. The analysis here is not contradicting K28 and K31 being charged in GB3. In the protein environment of GB3, these two residues are probably not as effectively screened by solvent.

However, if we believe that E27 is neutral and modify it to be so in our calculations, it affects some helical residues whose "long-chain" fragment inevitably includes E27. The calculated ^{15}N isotropic chemical shieldings of residues A26-Y33, then, would not correlate well with the experimental data. A consistent picture so far is not available.

Appendix D

A helical calculation illustrated with 1CEX

The crystal structure 1CEX (Longhi et al., 1997) with good resolution (1.0 Å) for cutinase is another good example of illustrating the α-helical residue calculation. The sequence of classical residues from G36 to S45 was considered. This sequence was selected also because that it only has one ionizable residue, which is E44. So only E44's charge state has to be considered. In Model C, E44 is involved in the calculations of residues from S41 to S45. It is noted that with E44 neutral, we get a slightly better correlation, intercept and slope (see Figure D.1).



(a) (b)

Figure D.1: The comparison of calculated chemical shieldings (using model C) and experimental chemical shifts for selected helical residues from 1CEX when (a) making E44 charged and (b) making E44 neutral. The regressed line of the least squares regression is shown along with its intercept and slope in the legend box; the third number in the legend box is the absolute value of Pearson's correlation $|r|$. P37 is not plotted because of missing experimental data.

Bibliography

Aidas, K., Mogelhoj, A., Kjaer, H., Nielsen, C. B., Mikkelsen, K. V., Ruud, K., Christiansen, O., Kongsted, J., 2007. Solvent effects on NMR isotropic shielding constants. A comparison between explicit polarizable discrete and continuum approaches. Journal of Physical Chemistry A 111 (20), 4199–4210.

Anonymous, June 1, 2001. Dihedral angle. http://en.wikipedia.org/wiki/Dihedral_angle.

Atkins, P., Friedman, R., 1997. Molecular quantum mechanics. New York: Oxford University Press.

Bader, R., 2009. Utilizing the charge field effect on amide N-15 chemical shifts for protein structure validation. Journal of Physical Chemistry B 113 (1), 347–358.

Barone, V., Cossi, M., 1998. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. Journal of Physical Chemistry A 102 (11), 1995–2001.

Becke, A. D., 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. Physical Review A 38 (6), 3098–3100.

Becke, A. D., 1993. Density-functional thermochemistry. 3. the role of exact exchange. Journal of Chemical Physics 98 (7), 5648–5652.

Bendova, L., Jurecka, P., Hobza, P., Vondrasek, J., 2007. Model of peptide bond-aromatic ring interactoin: correlated Ab Initio quantum chemical study. Journal of Physical Chemistry B 111 (33), 9975–9979.

Benzi, C., Crescenzi, O., Pavone, M., Barone, V., 2004. Reliable NMR chemical shifts for molecules in solution by methods rooted in density functional theory. Magnetic Resonance in Chemistry 42 (S1), S57–S67.

Brender, J. R., Taylor, D. M., Ramamoorthy, A., 2001. Orientation of amide-15N chemical shift tensors in peptides: A quantum chemical study. Journal of the American Chemical Society 123 (5), 914–922.

Cai, L., Fushman, D., Kosov, D. S., 2008. Density functional calculations of N-15 chemical shifts in solvated dipeptides. Journal of Biomolecular NMR 41 (2), 77–88.

Chesnut, D. B., Rusiloski, B. E., Moore, K. D., Egolf, D. A., 1993. Use of locally dense basis sets for nuclear magnetic resonance shielding calculations. Journal of Computational Chemistry 14 (11), 1364–1375.

Cook, W., Jeffrey, L., Carson, M., Chen, Z., Pickart, C., 1992. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). Journal of Biological Chemistry 267 (23), 16467–71.

Cook, W., Jeffrey, L., Kasperek, E., Pickart, C., 1994. Structure of tetraubiquitin shows how multiubiquitin chains can be formed. Journal of Molecular Biology 236 (2), 601–9.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A., 1995. A second generation force-field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society 117 (19), 5179–5197.

Cornilescu, G., Bax, A., 2000. Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. Journal of the American Chemical Society 122 (41), 10143–10154.

Cornilescu, G., Delaglio, F., Bax, A., 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of Biomolecular NMR 13 (3), 289–302.

Cornilescu, G., Marquardt, J., Ottiger, M., Bax, A., 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. Journal of American Chemical Society 120 (27), 6836–6837.

Cossi, M., Rega, N., Scalmani, G., Barone, V., 2003. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. Journal of Computational Chemistry 24 (6), 669–681.

Cramer, C. J., 2004. Essentials of computational chemistry: theories and models. Wiley.

de Dios, A. C., Laws, D. D., Oldfield, E., 1994. Predicting Carbon-13 nuclear magnetic resonance chemical shielding tensors in zwitterionic L-threonine and L-tyrosine via quantum chemistry. Journal of the American Chemical Society 116 (17), 7784–7786.

de Dios, A. C., Oldfield, E., 1993. Methods for computing nuclear magnetic resonance chemical shielding in large systems. Multiple cluster and charge field approaches. Chemical Physics Letters 205 (1), 108–116.

de Dios, A. C., Pearson, J. G., Oldfield, E., 1993a. Chemical-shifts in proteins - an Ab-Initio study of C-13 nuclear-magnetic-resonance chemical shielding in glycine, alanine, and valine Residues. Journal of the American Chemical Society 115 (21), 9768–9773.

de Dios, A. C., Pearson, J. G., Oldfield, E., 1993b. Secondary and tertiary structural effects on protein NMR chemical-shifts - an Abinitio approach. Science 260 (5113), 1491–1496.

Derrick, J. P., Wigley, D. B., 1994. The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. Journal of Molecular Biology 243 (5), 906–918.

Ditchfield, R., 1974. Self-consistent perturbation-theory of diamagnetism.1. Gauge-invariant LCAO method for NMR chemical-shifts. Molecular Physics 27 (4), 789–807.

Dominguez, C., Boelens, R., Bonvin, A., 2003. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. Journal of American Chemical Society 125 (7), 1731–1737.

Eddins, M., Varadan, R., Fushman, D., Pickart, C., Wolberger, C., 2007. Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. Journal of Biological Chemistry 367 (1), 204–11.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., Pople, J. A., 2004. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT.

Fushman, D., Cowburn, D., 2001. Nuclear magnetic resonance relaxation in determination of residue-specific N-15 chemical shift tensors in proteins in solution: Protein dynamics, structure, and applications of transverse relaxation optimized spectroscopy. Methods Enzymol 339, 109–126.

Fushman, D., Tjandra, N., Cowburn, D., 1998. Direct measurement of N-15 chemical shift anisotropy in solution. Journal of the American Chemical Society 120 (42), 10947–10952.

Fushman, D., Tjandra, N., Cowburn, D., 1999. An approach to direct determination of protein dynamics from 15N NMR relaxation at multiple fields, independent of variable 15N chemical shift anisotropy and chemical exchange contributions. Journal of the American Chemical Society 121 (37), 8577–8582.

Gronenborn, A., Filpula, D., Essig, N., Achari, A., Whitlow, M., Wingfield, P., Marius Clore, G., 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science 253 (5020), 657–661.

Hall, J. B., Fushman, D., 2006. Variability of the 15N chemical shielding tensors in the B3 domain of protein G from 15N relaxation measurements at several fields.

Implications for backbone order parameters. Journal of the American Chemical Society 128 (24), 7855–7870.

Hartzell, C. J., Whitfield, M., Oas, T. G., Drobny, G. P., 1987. Determination of the 15N and 13C chemical shift tensors of L-[13C]alanyl-L-[15N]alanine from the dipole-coupled powder patterns. Journal of the American Chemical Society 109 (20), 5966–5969.

Hiyama, Y., Niu, C. H., Silverton, J. V., Bavoso, A., Torchia, D. A., 1988. Determination of 15N chemical shift tensor via 15N-2H dipolar coupling in Boc-glycylglycyl[15N glycine]benzyl ester. Journal of the American Chemical Society 110 (8), 2378–2383.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD - Visual Molecular Dynamics. Journal of Molecular Graphics 14 (1), 33–38.

Jameson, C. J., Jameson, A. K., Oppusunggu, D., Wille, S., Burrell, P. M., Mason, J., 1981. 15N nuclear magnetic shielding scale from gas phase studies. Journal of Chemical Physics 74 (1), 81–88.

Jensen, F., 2007. Introduction to computational chemistry. Wiley.

Kover, K. E., Batta, G., 2001. Separating structure and dynamics in CSA/DD cross-correlated relaxation: A case study on trehalase and ubiquitin. Journal of Magnetic Resonance 150 (2), 137–146.

Kroenke, C. D., Rance, M., Palmer, A. G., 1999. Variability of the 15N chemical shift anisotropy in Escherichia coli ribonuclease H in solution. Journal of the American Chemical Society 121 (43), 10119–10125.

Kurita, J., Shimahara, H., Utsunomiya-Tate, N., Tate, S., 2003. Measurement of 15N chemical shift anisotropy in a protein dissolved in a dilute liquid crystalline medium with the application of magic angle sample spinning. Journal of Magnetic Resonance 163 (1), 163–173.

Le, H. B., Oldfield, E., 1996. Ab initio studies of amide-N-15 chemical shifts in dipeptides: Applications to protein NMR spectroscopy. Journal of Physical Chemistry 100 (40), 16423–16428.

Lee, C. T., Yang, W. T., Parr, R. G., 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B 37 (2), 785–789.

Lipsitz, R. S., Tjandra, N., 2003. 15N chemical shift anisotropy in protein structure refinement and comparison with NH residual dipolar couplings. Journal of Magnetic Resonance 164 (1), 171–176.

Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A., Cambillau, C., 1997. Atomic resolution (1.0 å) crystal structure of fusarium solani cutinase: stereochemical analysis. Journal of Molecular Biology 268 (4), 779–799.

Loth, K., Pelupessy, P., Bodenhausen, G., 2005. Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. Journal of the American Chemical Society 127 (16), 6062–6068.

Mai, W., Hu, W., Wang, C., Cross, T. A., 1993. Orientational constraints as 3-dimensional structural constraints from chemical-shift anisotropy: the polypeptide backbone of Gramicidin A in a Lipid Bilayer. Protein Science 2 (4), 532–542.

Manalo, M. N., de Dios, A. C., 2002. An ab initio study of solvent polarity and hydrogen bonding effects on the nitrogen NMR shieldings of N,N-dimethylacetamidine. Magnetic Resonance in Chemistry 40 (12), 781–785.

McCoy, M. A., Wyss, D. F., 2000. Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. Journal of Biomolecular NMR 18 (3), 189–198.

McCoy, M. A., Wyss, D. F., 2001. Structures of protein-protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. Journal of American Chemical Society 124 (10), 2104–2105.

Mennucci, B., Martinez, J. M., 2005. How to model solvation of peptides? Insights from a quantum mechanical and molecular dynamics study of N-methylacetamide. 2. 15N and 17O nuclear shielding in water and in acetone. Journal of Physical Chemistry B 109 (19), 9830–9838.

Mennucci, B., Martinez, J. M., Tomasi, J., 2001. Solvent effects on nuclear shieldings: Continuum or discrete solvation models to treat hydrogen bond and polarity effects? Journal of Physical Chemistry A 105 (30), 7287–7296.

Nadaud, P., Helmus, J., Jaroniec, C., 2007. C-13 and N-15 chemical shift assignments and secondary structure of the B3 immunoglobulin-binding domain of streptococcal protein G by magic-angle spinning solid-state NMR spectroscopy. Biomol. NMR Assignments 1 (1), 117–120.

Oas, T. G., Hartzell, C. J., Dahlquist, F. W., Drobny, G. P., 1987. The amide nitrogen-15 chemical shift tensors of four peptides determined from carbon-13 dipole-coupled chemical shift powder patterns. Journal of the American Chemical Society 109 (20), 5962–5966.

Oldfield, E., 1995. Chemical-Shifts and 3-Dimensional Protein Structures. Journal of Biomolecular NMR 5 (3), 217–225.

Oldfield, E., 2002. Chemical shifts in amino acids, peptides, and proteins: From quantum chemistry to drug design. Annual Review of Physical Chemistry 53, 349–378.

Phillips, C., Thrower, J., Pickart, C., Hill, C., 2001. Structure of a new crystal form of tetraubiquitin. Acta Crystallogr D Biol Crystallogr 57 (Pt 2), 341–4.

Poon, A., Birn, J., Ramamoorthy, A., 2004. How does an amide-15N chemical shift tensor vary in peptides? Journal of Physical Chemistry B 108 (42), 16577–16585.

Ryabov, Y., Fushman, D., 2006. Interdomain mobility in di-ubiquitin revealed by NMR. Proteins 63 (4), 787–96.

Selvarengan, P., Kolandaivel, P. G., 2004. Molecular modeling of dipeptide and its analogous systems with water. Journal of Molecular Modeling 10 (3), 198–203.

Shen, Y., Bax, A., 2007. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. Journal of Biomolecular NMR 38 (4), 289–302.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G. H., Eletsky, A., Wu, Y. B., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D., Bax, A., 2008. Consistent blind protein structure generation from NMR chemical shift data. Proceedings of the National Academy of Sciences of the United States of America 105 (12), 4685–4690.

Shen, Y., Vernon, R., Baker, D., Bax, A., 2009. De novo protein structure generation from incomplete chemical shift assignments. Journal of Biomolecular NMR 43 (2), 63–78.

Shoji, A., Ozaki, T., Fujito, T., Deguchi, K., Ando, S., Ando, I., 1989. Nitrogen-15 nmr chemical shift tensors and conformation of some nitrogen-15-labeled polypeptides in the solid state. Macromolecules 22 (6), 2860–2863.

Tjandra, N., Szabo, A., Bax, A., 1996. Protein backbone dynamics and 15N chemical shift anisotropy from quantitative measurement of relaxation interference effects. Journal of the American Chemical Society 118 (29), 6986–6991.

Tomasi, J., Mennucci, B., Cammi, R., 2005. Quantum mechanical continuum solvation models. Chemical Reviews 105 (8), 2999–3093.

Ulmer, T. S., Ramirez, B. E., Delaglio, F., Bax, A., 2003. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. Journal of the American Chemical Society 125 (30), 9179–9191.

Varadan, R., Walker, O., Pickart, C., Fushman, D., 2002. Structural properties of polyubiquitin chains in solution. Journal of Molecular Biology 324 (4), 637–647.

Vasos, P. R., Hall, J. B., Kummerle, R., Fushman, D., 2006. Measurement of 15N relaxation in deuterated amide groups in proteins using direct nitrogen detection. Journal of Biomolecular NMR 36 (1), 27–36.

Vijay-Kumar, S., Bugg, C., Cook, W., 1987. Structure of ubiquitin refined at 1.8 A resolution. Journal of Molecular Biology 194 (3), 531–44.

Vila, J. A., Scheraga, H. A., 2007. Factors affecting the use of 13C(alpha) chemical shifts to determin, refine, and validate protein structures. Proteins 71 (2), 641–654.

Villegas, M., Vila, J., Scheraga, H., 2007. Effects of side-chain orientation on the C-13 chemical shifts of antiparallel beta-sheet model peptides . Journal of Biomolecular NMR 37 (2), 137–146.

Wang, Y. J., Jardetzky, O., 2002. Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Science 11 (4), 852–861.

Wang, Y. J., Jardetzky, O., 2004. Predicting 15N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from phi, psi(i-1), and chi(1) torsion angles. Journal of Biomolecular NMR 28 (4), 327–340.

Wishart, D. S., Nip, A. M., 1998. Protein chemical shift analysis: a practical guide. Biochemistry and Cell Biology 76 (2-3), 153–163.

Wishart, D. S., Watson, M. S., Boyko, R. F., Sykes, B. D., 1997. Automated H-1 and C-13 chemical shift prediction using the BioMagResBank. Journal of Biomolecular NMR 10 (4), 329–336.

Wolinski, K., Hinton, J. F., Pulay, P., 1990. Efficient implementation of the Gauge-independent atomic orbital method for NMR chemical shift calculations. Journal of the American Chemical Society 112 (23), 8251–8260.

Wu, C. H., Ramamoorthy, A., Gierasch, L. M., Opella, S. J., 1995. Simultaneous characterization of the amide 1H chemical shift,1H-15N dipolar, and 15N chemical shift interaction tensors in a peptide bond by three-dimensional solid-state NMR spectroscopy. Journal of the American Chemical Society 117 (22), 6148–6149.

Wylie, B. J., Franks, W. T., Rienstra, C. M., 2006. Determinations of N-15 chemical shift anisotropy magnitudes in a uniformly N-15, C-13-labeled microcrystalline protein by three-dimensional magic-angle spinning nuclear magnetic resonance spectroscopy. Journal of Physical Chemistry B 110 (22), 10926–10936.

Wylie, B. J., Rienstra, C. M., 2008. Multidimensional solid state NMR of anisotropic interactions in peptides and proteins. Journal of Chemical Physics 128 (5), 052207.

Wylie, B. J., Schwieters, C. D., Oldfield, E., Rienstra, C. M., 2009. Protein structure refinement using $^{13}C_\alpha$ chemical shift tensors. Journal of the American Chemical Society 131 (3), 985–992.

Wylie, B. J., Sperling, L. J., Frericks, H. L., Shah, G. J., Franks, W. T., Rienstra, C. M., 2007. Chemical-shift anisotropy measurements of amide and carbonyl resonances in a microcrystalline protein with slow magic-angle spinning NMR spectroscopy. Journal of the American Chemical Society 129 (17), 5318–5319.

Xu, X. P., Case, D. A., 2001. Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13' chemical shifts in proteins using a density functional database. Journal of Biomolecular NMR 21 (4), 321–333.

Xu, X. P., Case, D. A., 2002. Probing multiple effects on 15N, 13C alpha, 13C beta, and 13C' chemical shifts in peptides using density functional theory. Biopolymers 65 (6), 408–423.

Zuiderweg, E. R. P., 2002. Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry 41 (1), 1–7.