Defining Local Order in the Unfolded State Using Short Peptide Model Systems and Spectroscopic Methods: Conformational Biases, Mediation by Solvation, and Nearest Neighbor Effects

A Thesis

Submitted to the Faculty

of

Drexel University

By

Siobhan Eileen Toal

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

August 2014

Dedications

For my Dad.

Acknowledgements

I would like to thank my research advisor and colleague Dr. Reinhard Schweitzer-Stenner. Not only did he advise me throughout this process as a scientific mentor, but also, through his own example, helped me become a better scientist and person. His support has been invaluable.

I'd also like to thank my family. Especially my uncles, aunts, and grandmother who made sure I had the support I needed in my personal life in order to accomplish this academic goal.

I would like to acknowledge the people with whom we collaborated and worked with to complete these studies. I would like to thank Dr. Harold Schwalbe and his research group for their collaboration, especially Dr. Daniel Matthieu, Dr. Christian Richter, and Marie Anders for their assistance in performing NMR experiments and for running additional NMR experiments in Frankfurt, Germany. I would like to thank Dr. Brigita Urbanc and Derya Meral for the very helpful collaboration and discussions, and specifically for conducting the MD experiments presented in this thesis.

I would also like to thank all of my colleagues, past and present, in the Biospectroscopy research group. Each person has added something to this research, whether it be scientific discussion or simply an open ear. In this regard I would especially like to thank Jonathan Soffer for instrumental advice, and Leah Pandiscia who, to her (mis)fortune, has had to sit next to me for the majority of 5 years.

Table of Contents

List of Tables

List of Figures

Abstract

Defining Local Order in the Unfolded State using Short Peptide Model Systems and Spectroscopic Methods: Conformational Biases, Mediation by Solvation, and Nearest Neighbor Effects

Siobhan Eileen Toal
Prof. Reinhard Schweitzer-Stenner

This thesis describes conformational ensembles of amino acid residues in unfolded peptides, and how their dependence on solvation and nearest-neighbor interactions can be obtained by a combination of vibrational and NMR spectroscopies. To this end, short peptide model systems may be chosen ranging from dipeptides to blocked and unblocked tripeptides. The question then arises whether one type of model system is more suitable for studying conformational propensities. Results from our spectroscopic studies suggest that the conformational ensemble of trialanine and its high pPII content are independent of the peptide's protonation state. In addition, we find that the conformational ensemble of the alanine dipeptide, a classic peptide model system, resembles the unblocked GAG model peptide, as expected in the absence of any end effects.

To explore the physical basis underlying residue-level conformational bias we utilized UVCD and NMR derived ^{3}J coupling constants to decompose the Gibbs free energy landscape. We found that the thermodynamics underling conformational propensities of (1) trialanine in different binary solvents and (2) GxG peptides in water

exhibit a near exact enthalpy-entropy compensation involving rarely observed isoequilibria. Their existence indicates peptide solvation as the common physical mechanism behind conformational preferences.

Contrary to the isolated pair hypothesis, an ingredient of the classical random coil model, amino acid residues can not be considered as isolated from their neighbors in the unfolded state. To explore nearest neighbor (NN) interactions, we chose a series of "GxyG" host guest peptides, where x/y={A,K,LV}. Utilizing a six different NMR J-coupling constants in conjunction with amide I' (IR, VCD, Raman) band profiles we extracted the conformational distributions of x and y residues in the GxyG peptides. Our data reveal large changes in conformational distributions due to neighbor interactions, contrary to the isolated pair hypothesis. Interestingly, residues that have large intrinsic biases towards specific sub-populations tend to loose these preferences upon interaction with a given neighbor, indicating a degree of conformational randomization. Strong effects induced by residues with bulky side chains suggests that the underlying mechanism is the the disruption of neighboring residues' hydration shells.

CHAPTER 1. INTRODUCTION

1.1 PROTEINS, THE PROTEIN FOLDING PROBLEM, AND THE UNFOLDED STATE

1.1.1 Protein Structure and Folding

Proteins are the major functional component of living cells and embody the quintessential relationship between structure and function at the bio-molecular level. There exists over 100,000 proteins of various sequences in the human genome that are involved in all aspects of the maintenance of life (1). At its most fundamental level, the structure of proteins consists of linear unbranched chains of amino acid residues attached through the peptide (C-N) bond. There are 20 natural amino acids common to all living organisms which comprise this primary structure of proteins. These amino acid compounds are composed of amine, carboxylic acid, and substituent side-chain functional groups (-R) attached to a common carbon atom referred to as the alpha carbon (Cα). The type of side-chain varies dramatically among amino acids, from a simple hydrogen atom in the amino acid glycine to much more complex aromatic-based structures such as tyrosine and tryptophan. Depending on the nature of the side-chain, an amino acid can be hydrophilic or hydrophobic, acidic or basic. A protein, then, can be depicted as a regularly repeating polypeptide backbone (…Ni—Cαi—Ci—Ni+1—Cαi+1—Ci+1…), with alpha carbons containing the variant R group of the 20 natural amino acids.

A protein's biological function is generally determined by its three-dimensional structure, which is attained through the process of "protein folding" in which the

unordered polypeptide chain arranges in space into its regularly ordered shape. The final thermodynamically stable and bio-functional state is generally referred to as the native state (N) of the protein. This (generally) reversible two-state process U↔N (where U denotes the unfolded state and N denotes the native state) is depicted in Figure 1.1. The process of protein folding is governed by the primary amino acid sequence as well as cellular interactions. The folding of a protein can be visualized as concerted process involving both local folding about the polypeptide backbone into regularly patterned secondary structures, as well as more global folding of protein sub-domains into the total tertiary 3D structure.

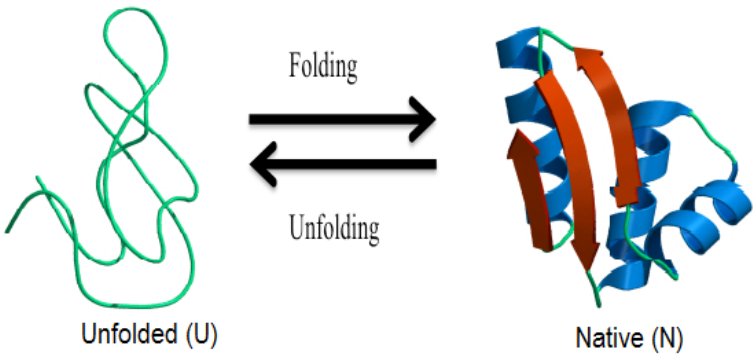

Figure 1.1: Scheme illustrating protein folding from unfolded (U) to folded native state (N).

This linear sequence of residues contains certain rotational degrees of freedom that lend flexibility to the chain and allow local folding into distinct three dimensional

conformations. The peptide group (H-Ni-Ci-O) itself is nearly planar due to electronic resonance and delocalization of the carbonyl π electrons over the CO and adjacent Ci—Ni bond. Due to the π-like molecular orbital that extends over the three atoms (O-C-N), the peptide bond has partial double bond character. This has been evidenced by X-ray crystallographic studies showing the C-N bond length is 1.33Å, only slightly larger than the 1.27Å for average C=N bond lengths (2). Hence, the angle, Ω, defined as the angle between carbonyl and amino groups, is generally fixed in a 180° trans-conformation. Local folding of the polypeptide chain then results in a regular "secondary structure" that can be fully defined through the remaining backbone dihedral angles, herein denoted as ϕ and ψ (Figure 1.2). These angles represent the only rotational degrees of freedom in the chain. The angle ϕ is defined as the torsional angle between the nitrogen and Cα (ϕi=∠Ci-1—Ni—Cαi—Ci), whereas the angle ψ represents the torsional angle between Cα and carbonyl carbon Ci (ψi=∠Ni—Cαi—Ci—Ni+1).

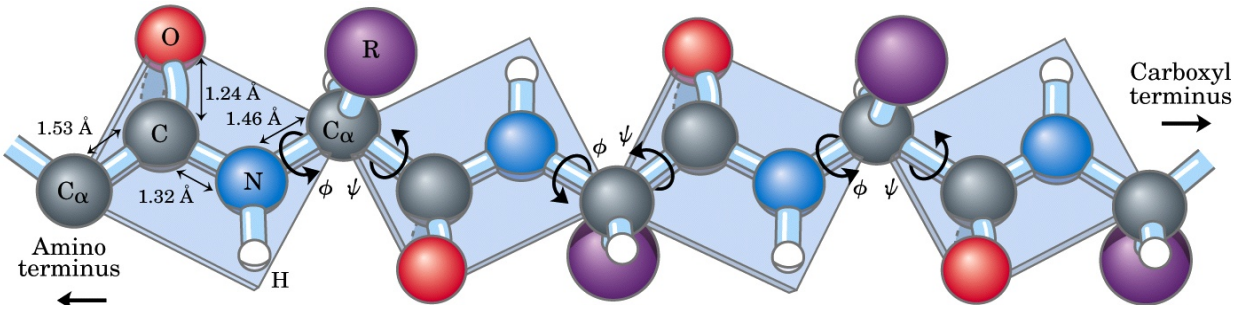

Figure 1.2: Polypeptide chain with dihedral angles ϕ and ψ denoted as well as typical bond lengths along the chain. Taken from reference (3) and modified.

In folded proteins, certain types of secondary structures predominate. Among these, the two most commonly found structures are α-helices and β-sheets (3). The right handed α-helix is spiral-like conformation with canonical (ϕ, ψ) = (-60°, -45°) which is stabilized by backbone (i→i+4) hydrogen bonding between N-H and C=O groups. In contrast, β-sheet structures are composed of H-bonded pairs of adjacent β strands which have relatively extended dihedral angles (ϕ , ψ) = (–140°, 135°) in antiparallel sheets and (ϕ , ψ) = (–120°, 115°) in parallel sheets (3). In contrast to the intra-strand H-bonding seen in alpha helical conformations, these structures are stabilized by inter-strand H-bonding between N-H and C=O groups. The conformations adopted by amino acid residue which constitute the loop between two beta-strands is generally referred to as β-turns. Depending on the position of amino acid within the loop, there can be various (ϕ , ψ) coordinates. Turns and loops play special roles in protein chemistry as they are thought to serve as initiation sites for the folding process (4).

1.1.2 Ramachandran Plots

In the early 1960's G. N. Ramachandran contrived a relatively simple method for analyzing and conceptualizing regular secondary structure motifs in proteins (5). Ramachandran investigated the conformations available to the so-called "alanine

dipeptide" (AdP), which is a single alanine residue with methyl-blocked termini (N-acetyl-alanyl-N′-methylamide) (Figure 1.3).

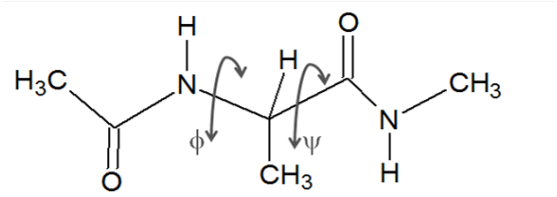

Figure 1.3: The structure of the alanine dipeptide (N-acetyl-alanyl-N-methylamide) showing dihedral angles ϕ and ψ.

The choice of this compound is motivated by the fact that alanine has the simplest side chain (-CH_3) other than glycine (-H) and hence the alanine dipeptide was thought to offer a baseline for identifying regions in (ϕ, ψ)-space in which repulsions are minimized. Ramachandran utilized simple models that assumed standard atomic radii and local hard sphere repulsions between atoms that are at least third neighbors (i.e. 1-4 interactions), to calculate sterically accessible ϕ and ψ combinations. For this analysis the author constructed 2D correlation plots reflecting the probability of specific secondary conformations. This type of plot, widely termed a "Ramachandran plot", is depicted in Figure 1.4 and is clearly dominated by "extended structures" including β-strand conformations in the upper left quadrant, right-handed helical structures in the lower left

quadrant, and left-handed helical conformations in the upper right quadrant. The obtained regions accurately reflect experimentally known secondary structures (i.e., (ϕ, ψ)-positions) sampled by folded proteins. Ramachandran also reported plots for dipeptides of residues other than alanine, which were found to be nearly identical to that of alanine, and hence, alanine's conformational sampling was for a long time thought to be representative of all naturally occurring amino acid residues except glycine and proline.

Ramachandran plots are now regularly used as a tool to express and visualize secondary structures found in proteins and peptides or conformations derived through experimental/theoretical means. Indeed, as increasing numbers of high-resolution protein structures become available through X-ray crystallography, residue conformations in folded proteins have been found to distribute according to the outlines in the Ramachandran plot, with most residues in favored regions.

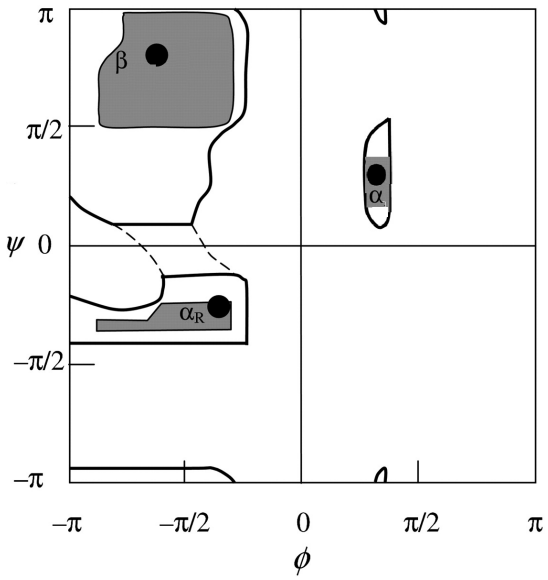

Figure 1.4: Generalized Ramachandran plot which shows the allowed (black lines) and favored (grey filled) ϕ,ψ conformations obtained for the alanine dipeptide with the three major secondary structures (β: β-strand, $\alpha_{R:}$ right handed α helical, $\alpha_{L:}$ left handed α helical) depicted.

1.1.3 Anfinsen's Thermodynamic Concept

The entire protein folding process is inherently coupled to the conformational sampling of individual amino acid residues and hence the energy landscape dictated by the amino acid sequence. The now classical experiments performed by Anfinsen et al. on ribonuclease showed that proteins can spontaneously (ms), reproducibly, and reversibly fold into their native bio-functional state, indicating that all the information required for protein folding is encoded in the primary amino acid sequence (6). Anfinsen's own explanation for this spontaneous re-folding was the so-called "thermodynamic model," which postulates that the native state is the most thermodynamically stable, and as such

the folding process must be a consequence of the drive to minimize the Gibbs free energy(7). In other words, a protein achieves it native state as a result of the constraints of the polypeptide backbone as modified by the chemical and physical properties of amino acid side chains. To test this hypothesis, Anfinsen unfolded the ribonuclease enzyme under extreme denaturing conditions (high guanadinium chloride concentrations) and observed that the protein re-folded spontaneously back into its native structure when returned to non-denaturing conditions.

1.1.4 Levinthal's Paradox

Although the above described simple thermodynamic model explains the overall driving force for protein folding, it leaves many questions unanswered. In the well known "Levinthal paradox," it was countered that proteins fold on time scales far too short (μs-s) to allow for a stochastic search of conformational space (8). Specifically, if one considers that each residue can sample three possible conformations (a modest estimate), statistics predicts that there exists about 3^{100} possible configurations of the chain. If the protein is able to sample new configurations at the rate of one picosecond, it would take longer than the age of the universe for it to find the native state through a stochastic search of the energy landscape. Since we know that proteins fold reproducibly in short time scales, there must be some unifying driving forces that enable a more biased search for the native state. However, considering Levinthal's calculation of the vast number of

conformational states possible, the question still remained how a protein navigates pathways to this global minimum in energy.

1.1.5 Energy Landscapes and Folding Funnels

Anfinsen's thermodynamic approach has been extensively adopted and expanded over the years. Most importantly, Wolynes, Onuchic and colleagues reconsidered Levinthal's conformational search problem along with Anfinsen's findings, and took into account that not all conformations are equally likely (9-11). In fact, Levinthal's paradox implicitly states that a folding protein starts from one specific conformation, and that all available unfolded states are isoenergetic. However, by the very definition of the denatured state of a protein, it can be concluded that it is not a single point on the energy landscape, but rather an ensemble of such disordered states. The so called "new view" of folding describes the folding process in terms of statistical ensembles of states, and hence, folding is depicted as a flow process of an ensemble of chain molecules that navigate a complex multidimensional potential energy funnel (Figure 1.5). The free energy surface of the polypeptide chain can then be viewed as a funnel-like landscape with many local minima corresponding to small energy traps (metastable states) on the order of RT and a single global minimum corresponding to the native state.

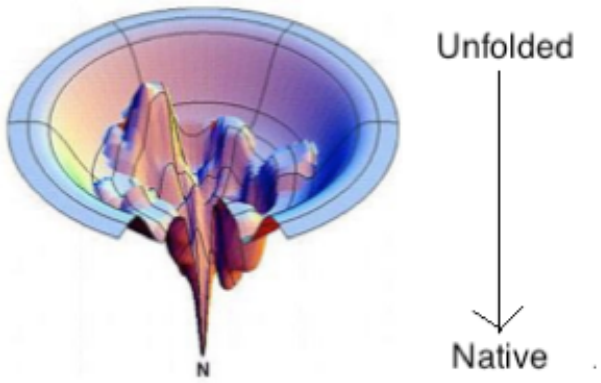

Figure 1.5. The funneled energy landscape representing the process of going from unfolded state to native state. Taken from ref. (10) and modified.

The depth of the funnel depicts the sum of all contributions to internal free energy such as hydrogen bonds, hydrophobic, and solvation free energies of a given chain conformation. The width of the funnel represents the configurational entropy of the polypeptide chain, i.e., the conformational degrees of freedom in (ϕ, ψ)-space. In going from the unfolded to native state on the funnel, there is a progressive reduction in dimensionality of the accessible conformational space (search volume), and concomitantly a decrease in the entropy and increase in favorable enthalpic interactions in the system. Formation of adventitious contacts and conformations lowers the free energy and increases the likelihood of formation of still further adventitious conformations. This way, the unfolded protein follows a pathway of minimal frustration

towards the absolute minimum in Gibbs energy (10). Rather then performing a random search for the native state, unfolded proteins fold by accumulating favorable protein-protein and protein-solvent interactions that lower energy and promote chain compaction.

1.1.6 The Unfolded State

At its most fundamental level, the "protein folding problem" asks how the amino acid sequence encodes a protein's final 3D structure. Since the unfolded state of proteins and peptides constitutes the starting point for this conformational search, a complete understanding of its local structure and dynamics is important. In addition, the discovery of Intrinsically Disordered Proteins (IDPs), which contain significant levels of disorder yet undergo complex biological functions, as well as unwanted aggregation, has motivated numerous experimental and theoretical studies aimed at describing local order in the unfolded state (12-15). These proteins challenge preconceived "lock and key dogma," which dictated that a protein must fold in order to be bio-functional. Some IDPs are prone to self-aggregation, which causes their involvement in numerous neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, the prion diseases, as well as the systemic amyloidoses (16-20). IDPs and unfolded proteins have in common that the structure of both are considered highly disordered relative to the the native state of proteins. Studying the unfolded state, therefore, should shed light on the key determinants of both the protein folding process and the structural distributions exhibited by various IDPs.

Unfortunately, for a long period of time the unfolded state attracted little scientific attention, in part, due to insufficient methods to study its inherent heterogeneity and dynamics, but also due to the wide-spread notion that it could be described as a "random coil" polymer, the residue conformations of which are governed solely by local steric interactions (5). Within the aforementioned funnel model, this random coil-like unfolded state can be visualized as a hyperplane of iso-energetic conformations at the top edge of the folding funnel, which are governed by large amounts of entropy (see Figure 1.5).

1.1.7 The Random Coil Model

The description of the unfolded state as a random coil stems originally from polymer theory, in which the unfolded state was described as a flexible polymer in a theta solvent. In a theta solvent, the effects of repulsive interactions exactly counterbalance the effects of attractive interactions, and hence the polypeptide chain is considered a freely jointed, freely rotating chain in which there are no significant intra-protein, non-local interactions (21, 22). The theta point marks the boundary between the good and bad solvents. In the theta solvent, the medium provides an exact compensation for the excluded volume effect. The random coil model thus assumes that unfolded polypeptides chains have large associated conformational and combinatorial entropies, which must be overcome by favorable enthalpic gains and entropic losses (i.e. H-bonding, hydrophobic interactions such as π-stacking, ete…) in order to fold into a stable state. It is not expected to exhibit any residual structure.

The validity of the random coil approach for unfolded and denatured states of proteins was first corroborated by the pioneering work of Tanford where it was reported that denatured proteins have a radius of gyration (R_g) that conforms to polymeric models of random coils (23). Specifically, it was shown that the intrinsic viscosity, [η], of denatured proteins varies with molecular weight, and hence the number of residues (n), according to a simple power law $[\eta]=n^a$, with the exponent "a" equal to approximately 0.68, well within the range predicted by Paul J. Flory (0.5-0.8) for random coils (22). In addition, Flory's model for a polymeric random coil assumes that each monomer sub-unit is randomly oriented with respect to neighboring monomer sub-units. This is the so-called "Isolated Pair Hypothesis" (IPH). The IPH suggests that the total number of possible conformations that can be adopted by a polypeptide can be determined as the product of the number of conformations sampled by each amino acid residue.

1.1.8 Local vs. Global Random Coils

The term random coil here deserves further clarification. Within the scope of polymer chemistry, a "random coil" generally suggests a long-chain polymer in which the entire backbone exhibits no well-defined structure. This notion stems from the Flory's description of freely jointed chains in theta solvents as described above and is supported by Tanford's work (22, 23). Herein we will denote this situation as a "global random coil." The "global random coil" applied to proteins thus assumes that unfolded

polypeptide chains adopt a multitude of conformations that lead to a Gaussian distribution of end-to-end distances. This distribution would be associated with large conformational and combinatorial entropies.

With regard to the sub-nanoscale (residue-level) scale, Flory's model for a polymeric random coil assumes that each monomer sub-unit is randomly oriented with respect to neighboring monomer sub-units (i.e., the IHP) (22). This hypothesis and Flory's reliance on an extensive sampling of the Ramachandran space by individual residues, constitutes what is referred to herein as a "local random coil (24)." Therefore, the term random in this context means that each amino acid residue can sample the entire sterically allowed, nearly ergodic region of the Ramachandran space without being influenced by its nearest neighbors. This type of maximal conformational sampling would encompass the entire shaded region displayed above in the generalized Ramachandran plot (Figure 1.4). In general, the work *in this thesis* focuses on experimental studies that challenge this local random coil model.

1.2 Evidence for non-Randomness in the Unfolded State

Reproduced in part, from:

I. Toal, S., Schweitzer-Stenner, R. Local Order in the Unfolded State: Conformational Biases and Nearest Neighbor Interactions. Biomolecules 2014, 4(3), 725-773

1.2.1 Overview

The main question concerning unfolded states is whether or not "unfolded" necessarily means "random" or "devoid of secondary structure." For a long period of time since the work of Tanford (23) and Flory (21) the notion that the sterically accessible conformational space is randomly sampled by each residue within an unfolded protein was considered textbook dogma. This greatly suppressed any deeper interest in unfolded state exploration. More recently, however, exploring unfolded states has become increasingly relevant due to (1) the observation of residual structure in unfolded states as well as distinguishable conformational biases of individual amino acid residues that both suggest a breakdown of the local random coil model (25-31) and (2) the discovery of Intrinsically Disordered Proteins (IDPs), some of which lack any well-defined structure yet perform complex biological functions (12-15, 32-34).

The challenge of the classical random coil model is based on numerous experimental, bioinformatic, and, to a more limited extent, computational observations. Firstly, numerous NMR studies have revealed the existence of residual structure in both IDPs and denatured proteins. Works from the laboratories of Dobson and Shortle suggest the existence of some 'native-like topology' in the denaturated state of proteins, that is a result of non-local interactions (25-27, 35-38). In addition, work by Dyson and Wright

(39-42), and numerous NMR studies by Blackledge, Schwalbe and their associates (20, 28-30, 43-48), revealed residual structure in the form of local turns, as well as unstable helical and strand structures in unfolded peptides and IDPs. Secondly, it is now generally established that individual amino acid residues in short peptides do not conform to the basic assumptions of the local random coil model, in that they display different conformational preferences with a much less entropic Gibbs energy landscape (49-55). This is particularly true for alanine, for which most experimental evidence indicates an unusually high preference for the so-called “poly-proline II” conformation, which is at odds with the local random coil model (49-51, 54). For instance, Figure 1.6 displays the difference between conformational sampling in a local random coil model compared to recent experimental work reported by Toal et al. for the alanine dipeptide (56). Obviously, there is a large discrepancy between the two, with experiment revealing a much more restricted conformational sampling. The conformational analysis of alanine based peptides is one of the main subjects of this thesis.

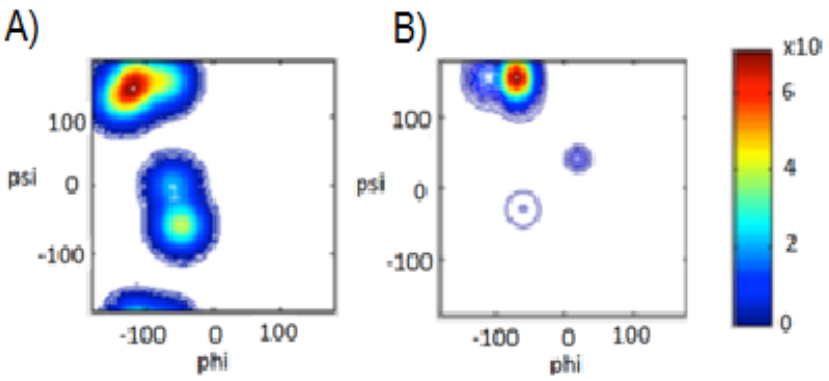

Figure 1.6: (A) Ramachandran plot depicting a local random coil distribution of conformations. (B) The experiential obtained Ramachandran plot of the alanine residue of the alanine dipeptide in water as reported by Toal et al. (56).

1.2.2 Model Systems for the Unfolded State

The presence of residual structure in the unfolded state has significant implications for protein folding theories, as the number of conformations needed to be sampled by a polypeptide chain would be greatly reduced. This type of investigation is, however, complicated by experimental limitations and the inherent dynamics of the unfolded/ disordered state system, which in some cases can yield inconsistent results. As previously stated by Kallenbach and associates (57), recent advances have afforded a few systems to become the favored way of studying the unfolded state, namely by investigation of denatured proteins, intrinsically disordered proteins, charged oligopeptides, short peptides with no long range order, and/or coil libraries.

Short peptides have been extensively utilized as model systems for studying unfolded/disordered states. Generally, such peptides are unable to form the stable hydrogen bonding and long range interactions needed to fold into well defined secondary structures. Thus, one can study the conformations of residues in an unfolded state without the necessity of applying denaturing agents which are likely to affect the backbone structure. In particular, short alanine based peptides have been increasingly used in this regard for a variety of reasons. Alanine is the most abundant amino acid residue in nature and exhibits the highest helical propensity in folded proteins (3). In terms of the unfolded state, alanine has also proven special with regard to its unusually high preference for the so called polyproline II region in the Ramachandran plot, which is discussed in detail in the following chapters. In addition, alanine is one of the simplest amino acids due to the small methyl side chain, which provides a basic model for comparing experimental and computational results.

1.2.3 The pPII Conformation and the TK Hypothesis

Historically, the local random coil model did not remained unchallenged. Tiffany and Krimm were among the first who questioned its applicability for proteins (58). In a series of experiments using ultraviolet circular dichroism (UVCD), they reported that the spectrum of unfolded, fully ionized poly-L-glutamic acid and poly-L-lysine are remarkably similar to that of the conformationally restricted trans-L-polyproline (Figure 1.7) (58, 59). In proline rich proteins, such as collagen, the trans-L-polyproline (pPII)

structure is a rather rigid structural motif with backbone dihedral angles (-78°, 146°) located in the upper-left quadrant of the Ramachandran map (60). The similarity of UVCD spectra between *trans*-L-polyproline and these oligopeptides was hence surprising since charged oligopeptides were thought to adopt a true random coil due to the high degree of electrostatic repulsion along the polypeptide chain.

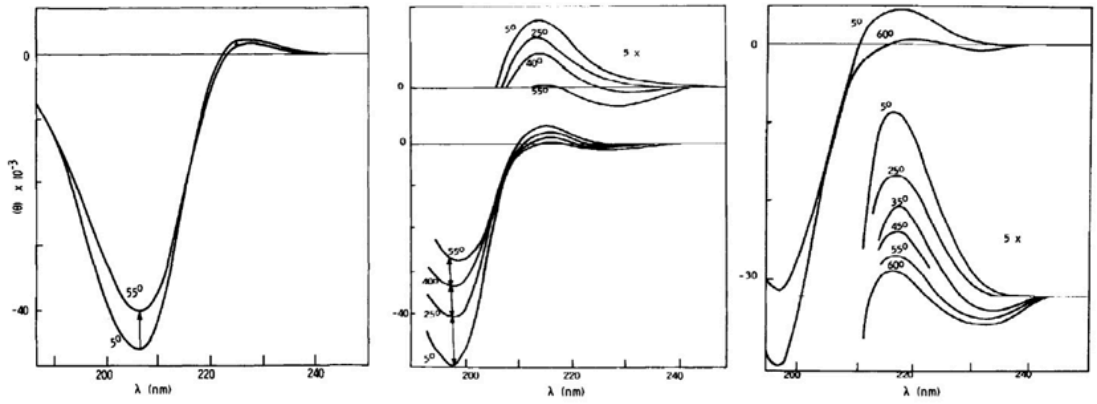

Figure 1.7. Left Panel: UVCD spectra of poly-L-proline II, Middle Panel: poly-L-glutamic acid, and Right Panel: poly-L-lysine measured as a function of temperature, which show characteristic pPII signals. Taken and modified from (59).

The typical pPII UVCD spectra of non-proline peptides exhibits a pronounced asymmetric couplet with a large negative maximum at approximately 195 nm (π->π*) and small positive maximum at approximately 218 nm (n->π*) (Figure 1.7). The theory of CD and UVCD are discussed in further detail in *Chapter 2 of this thesis*. Previous to Tiffany and Krimm's studies, this type of UVCD spectrum was strictly associated with

disordered chains with no regular secondary structure (i.e., random coils). Woody and coworkers have since shown that the adaptation of pPII conformations in peptides and proteins indeed gives rise to this far UVCD spectrum (61-64), although many in the scientific community still mis-interpret this signal as indicative of a random coil.

Tiffany and Krimm hence concluded that charged polypeptides assume, at least locally, a rather ordered pPII conformation . These authors also observed the resemblance between UVCD spectra of proline-based peptides (which may be expected to form a stable pPII structure) and proteins unfolded by denaturing agents, which led them to hypothesize that the conformational manifold of unfolded peptides and proteins are dominated by pPII-like conformations (59).

Tiffany and Krimm's initial challenge of the random coil model did not at first meet acceptance. The main argument against interpreting the negative couplet in the UVCD spectra of homopeptides as being indicative of pPII was based on the observation that similar couplets were also observed for very short (unionized) dipeptides in water (65). Since it was assumed at this time that individual amino acid residues in such short peptides could not exhibit any structural preference, it was concluded that the interpretation of Tiffany and Krimm was questionable.

It was only after Dukor and Keiderling presented similar findings as Tiffany and Krimm, using Vibrational Circular Dichroism (VCD) spectroscopy on the same polypeptides, that the random coil theory became a central point of contention again (66).

Dukor and Keiderling showed that the amide I mode of both, poly-prolines $(Pro)_n$ of different lengths (n=3,4) and ionized poly-L-glutamic acid (PLG) give rise to very pronounced negative-positive couplets in the respective VCD spectra, the intensity of which increases with the length of the peptide. Based on this evidence, it was concluded that the conformational ensemble sampled by PLG must in fact have large fractions of pPII, in agreement with Tiffany and Krimm's work. The VCD of the amide I mode is extremely conformational sensitive, which will be more fully discussed in *Chapter 2 of this thesis*, and continues to be used as marker by protein biochemists for disentangling various degrees of conformational order (67).

1.2.4 Alanine Studies: Experimental

1.2.4.1 Evidence for pPII

The preferential sampling of pPII-like conformations by individual amino acids has been identified many times since the work of Tiffany and Krimm, and Dukor and Kiederling. Particular in this regard are studies on alanine and short alanine-based peptides. One of the most notable, and widely debated, studies in this context was performed by Shi et al., who utilized HNMR and UVCD spectroscopy to investigate a peptide containing a sequence of seven alanine residues, i.e., X_2-A_7O_2-NH_2 ("XAO"), where O is ornithine and X is diaminoisobutyric acid (49). Figure 1.8 shows the structurally sensitive $^3J(H^NH^\alpha)$ values for all alanine residues within the XAO peptide

obtained by Shi et al. as a function of temperature. The monotonous increase of $^{3}J(H^{N}H^{\alpha})$ with rising temperatures and the clearly detectable temperature dependence of the UVCD spectrum indicate that the conformational distribution of the peptide changed as a function of temperature. This would not be the case if most conformations of the ensemble were nearly iso-energetic, as dictated by the local random coil model.

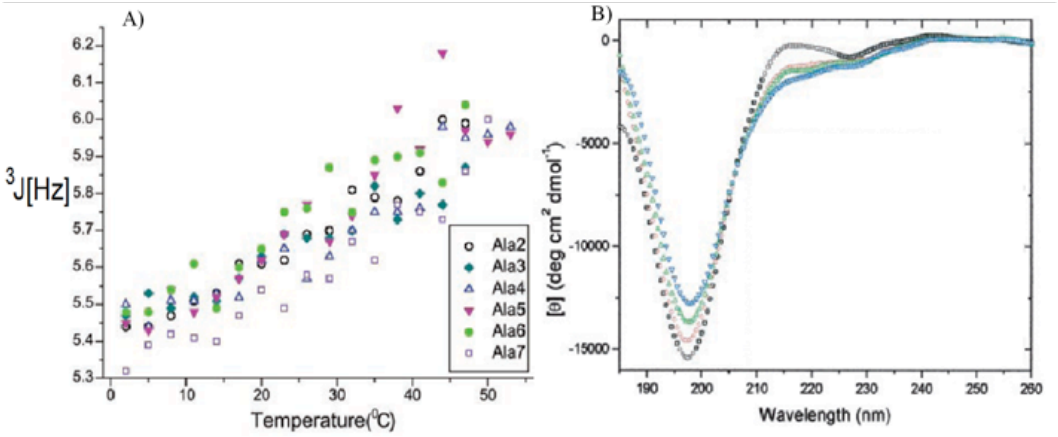

Figure 1.8: (a) The change in $^{3}J(H^{N}H^{\alpha})$ with temperature for Ala 2-7 residues of the $X_2A_7O_2$-NH_2 peptide. (b) The UV-CD spectra of XAO at ▫ (blue) 1°C; ○ (green) 35°C; △ (red) 45°C; and ▽ (black) 55°C. Taken from ref. (49) and modified.

The use of the empirical Karplus equation, which relates $^{3}J(H^{N}H^{\alpha})$ to backbone ϕ angles (see *Chapter 2 of this thesis*), led the authors to conclude that at room temperature the XAO peptide is predominantly in a conformation with an apparent backbone ϕ angle

of approximately 70° (49). Although this ϕ angle could theoretically correspond to both, pPII and α-helical conformations in the upper and lower left quadrants of the Ramachandran plot respectively, the measured UVCD spectra (Figure 1.8b) clearly resembles those reported by Tiffany and Krimm (59), thus leaving pPII dominance as the likely option. The monotonic increase in $^{3}J(H^{N}H^{\alpha})$ coupling constants with rising temperature suggests a redistribution towards conformations with more negative ϕ-values, consistent with the notion of an increasing population of β-strand-like conformations. This is corroborated by the temperature dependence of the UVCD spectrum which suggests that the conformation populated at high temperatures must exhibit a positive couplet with a positive signal below and a negative signal above 200 nm, which is diagnostic of β-strand like conformations (see *Chapter 2 of this thesis, Theory Behind Spectroscopic Techniques)*.

The preferential sampling of pPII by alanine in short and unfolded peptides proposed by Kallenbach, Shi, and colleagues have subsequently been corroborated numerous times by studies on different types of alanine-based peptides (56, 68-75). In this context, the alanine dipeptide (AdP), which is a single alanine residue flanked by two methyl-blocked peptide groups (see Figure 1.3), has played an important role owing to its traditional use as classical peptide model system, since Ramachandran et al. used it to construct (ϕ,ψ) maps 50 years ago (5). Even prior to the above study on XAO, Han et al. reported a strong pPII preference for the AdP peptide based on a comparison of experimental Raman, VCD, and ROA spectra with spectra calculated using DFT

approaches (76). In their study, they mimicked explicit water solvation by constructing AdP complexed to four water molecules (i.e., AdP-$(H_2O)_4$) and found that the presence of explicit water imposed a dominant preference for pPII, as well as right-handed helical conformations on the alanine residue of the peptide. Weise et al. later provided further experimental evidence for pPII dominance in AdP solvated in $CsPFO_ND_2O$ by rationalizing NMR derived dipolar coupling constants with a single representative pPII conformation (77, 78). Using 2D IR experiments, Hochstrasser and co-workers were able to derive the angle between the two amide I' transition dipoles of AdP, which correlates best to a representative pPII-like conformation (-70°, 120°) (79).

In addition to the classic AdP peptide, short unblocked oligoalanines have also been extensively subjected to conformational studies. Woutersen and Hamm, for instance, exploited the backbone sensitivity of the amide I mode in peptides by using non-linear time resolved 2D-IR spectroscopy to analyze cationic trialanine (AAA) in aqueous solution (68, 69). From their experiment they inferred the strength of the nearest neighbor coupling between the peptide's amide I modes and their relative orientation. By combining this with results from *ab initio* calculations on the (ϕ, ψ)-dependence of excitonic coupling, the authors were able to identify a representative pPII like conformation. In a later study from this group the results of MD simulations were utilized to re-analyze the results of the time resolved IR-experiment; this yielded a 80% fraction for pPII and 20% for a right-handed helical conformation (71). Barron and coworkers took a different approach to investigate the conformational preference of short alanine

oligomers (80). Using Raman optical activity (ROA) measurements, they showed that alanine oligomers from 3-7 residues were predominantly adopting pPII-like structures. To this end, they utilized the appearance of a very specific signal at 1314 cm^{-1} which was found to be diagnostic of pPII. Results from these experiments suggest that alanine oligomers ranging from tri- to hepta-alanine prefer pPII conformations in solution and that the pPII content per residue increases with the number of residues. In a separate study, Asher and coworkers used the ψ-sensitivity of UV-resonance enhanced amide III modes to show that a longer helical peptide, referred to as "AP" (AAAAA-$(AAARA)_3$-A) unfolds into a predominantly pPII-like structures (81, 82).

Prior work by Schweitzer-Stenner and co-workers has also indicated high pPII preferences for alanine based peptides (70, 73). Eker et al., for instance, utilized the excitonic coupling between local amide I modes to examine several polyalanines in water (72, 73). For unblocked and semi-blocked tripeptides the authors determined the intensity ratios of the two amide I bands in the respective IR, isotropic, and anisotropic Raman spectra and analyzed them with an excitonic coupling formalism to determine the (ϕ, ψ) values of a representative conformation of the peptide's central residue. Results from this analysis were checked by simulating the corresponding VCD signal for this conformation which was then compared with the experimentally determined profile. Additional analysis of this representative structure subsequently suggested that trialanine exhibits a 50:50 mixture of pPII and β. A similar investigation was performed on tetraalanine (AAAA) and revealed a higher pPII content (83), in agreement with McColl et al. (80).

1.2.4.2 Evidence for Statistical Coil

Although the above studies qualitatively agree in that they suggest a relatively high pPII propensity of alanine (though with different values for pPII fractions), it should be noted that a minority of results have challenged this notion. Particular in this regard are the conflicting reports concerning the pPII preference of the aforementioned XAO peptide studied by Shi et al. In a set of studies, Scheraga, Liwo, and colleagues performed MD simulations using the Amber 99 force field and simulated annealing (MD SA) to calculate the conformational ensemble of the XAO peptide (84, 85). The authors obtained a conformational ensemble, which they decomposed into 10 dominant conformational families, that satisfied experimental parameters. Figure 1.9 displays the Ramachandran map these authors obtained, which superimposes the entirety of the conformational families of all residues within the peptide. From this plot one may infer two major sub-populations within the families, namely one centered at ϕ = -160, which contains mainly extended β-strand populations, and a second centered at ϕ = -70, which contains pPII as well as β-turn-like conformations. These results suggest that the peptide exists in an ensemble of inter-converting structures, among which pPII is only one of many conformations sampled by its alanine residues. These results are supportive of the concept of a statistical coil, which the Scheraga group had earlier suggested as a modified version of the more simplistic random coil model (86, 87).

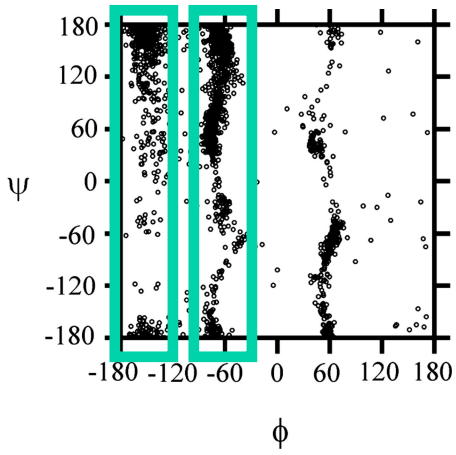

Figure 1.9: Ramachandran map superimposing the backbone distributions of all residues and all conformational families of the XAO peptide calculated by MD SA with NMR derived time-averaged restraints. The green boxes indicate the two dominant conformational clusters centered at ϕ = -160 and ϕ = -70. Taken from ref (84) and modified.

In an effort to resolve the conflict views of XAO experimentally, Zagrovic et al. conducted small angle X-ray scattering (SAXS) measurements on XAO, from which they derived a radius of gyration of 7.4 Å (88). This value seems to be inconsistent with a substantial sampling of pPII, for which one would expect a much more extended radius of gyration of 11.6 Å. Subsequently, by using their MD derived conformational ensemble for XAO, Makowska et al. (84) were able to reproduce this 7.4 Å radius of gyration obtained by Zagrovic et al. (88) as well as the $^{3}J(H^{N}H^{\alpha})$ constants of Shi et al. (49) with MD simulations, which led further credence to the notion that pPII is not dominantly sampled by the XAO peptide. However, the conformational manifold of alanine residues derived from Makowska's analysis was still atypical in that it indicated substantial

sampling of multiple turn-like conformations, which produced the rather compact structure of the peptide (as reflected by its small radius of gyration). Hence, their model was in fact still a departure from the classical random coil.

1.2.4.3 Reconciliation

Schweitzer-Stenner and Measey tackled these conflicting views concerning the conformation of the XAO peptide with a more extended experimental approach (89). They utilized the structural sensitivity of the amide I mode in polypeptides by simulating the IR, isotropic, anisotropic Raman, and VCD amide I band profiles as well as the $^{3}J(H^{N}H^{\alpha})$ constants obtained by Shi et al. (49) for XAO. They extended the two-state model of Shi et al. (pPII↔β), by constructing a statistical ensemble in which each residue was allowed to adopt a manifold of different representative conformations (pPII, β, helical and various turns structures), thus also considering the results of Makowska et al. (84, 85) Experimental spectra were fit using these distributions within an excitonic coupling model that exploits the conformationally sensitive coupling between the local amide I′ vibrational modes. As a result, they found that the best reproduction of all experimental data was achieved by assuming a mixture of conformations that contains various turn (26%) and β-strand conformations (23%) with a sizeable (50%) contribution from canonical pPII conformations. The inclusion of various turn structures at the XA and XO interface is in agreement with the MD simulations of Makowska et al. (84), but a high pPII contribution localized on the central alanine residues clearly indicate a pPII

preference for alanine, in line with Shi et al.(49). With their statistical model, the authors calculated a value of 19.1 Å for the peptide's end-to-end distance, which is nearly fully consistent with the R_g reported by Zagrovic et al. (88). Hence, Schweitzer-Stenner and Measey were able to reconcile both sides of the "XAO debate." Generally, the authors confirmed the notion that alanine residues exhibit more pPII sampling than predicted by any random or statistical coil models. However, their analysis also revealed a substantial nearest-neighbor influence of the hydrophilic terminal residues on the conformational manifold of alanine, which is at variance with the IPH.

1.2.4.4 Moving away from representative structures

Most of the above discussed spectroscopic studies have in common that they invoke representative conformations to reproduce experimental data, i.e.,, single (ϕ, ψ) point pairs for the entire conformational manifold or for subpopulations. Thus, the results are of limited use for (1) comparisons with conformational distributions obtained from computational studies, (2) for a reliable quantitative assessment of conformational propensities and (3) for estimating the conformational entropy of unfolded peptides.

These shortcomings have recently been addressed with increasingly more complex studies that combine various techniques, like NMR and vibrational spectroscopy, as well as theoretical methods to yield realistic conformational distributions in the Ramachandran space. The first step in this direction was made by Graf et al. by combining experimentally derived sets of seven different NMR J coupling constants,

each of which relate differently to ϕ and ψ backbone angles, along with distributions derived from constrained all atom Molecular Dynamics (MD) simulations to determine conformational ensembles of tri- to hepta-alanines (74). The authors determined that these peptides predominantly sample pPII (up to 90% for trialanine), with some minor admixtures of extended β structure. In agreement with the earlier alanine studies of Shi et al. (49) and Eker et al. (73), the population of right-handed helical conformations was found to be negligible.

Following the study of Graf et al. (74), Schweitzer-Stenner et al. subsequently used the published coupling constants as constraints in a newly developed algorithm which, for the first time, described the conformational ensembles of residues as a superposition of two-dimensional Gaussian functions (90). Using this model, the authors were able to simultaneously reproduce the J-coupling constants reported by Graf et al. (74) for trialanine, as well as the amide I' band profiles of Raman, IR and VCD reported by Eker et al. (73), with a conformational ensemble containing 84% pPII. The total 3D-distribution function (i.e. conformational ensemble) obtained by Schweitzer-Stenner et al. (90) for alanine in AAA is shown below in Figure 1.10. Thus, they confirmed the high sampling of pPII for this peptide as obtained by Graf et al (74). In this study, the combined use of the structurally very sensitive VCD signal of amide I' and a large number of φ-dependent coupling constants allowed a rather precise differentiation between pPII and β- strand sub-distributions, whereas the ψ-dependent coupling

constants were useful for assessing the relative population of states associated with the upper and lower left quadrant of the Ramachandran plot.

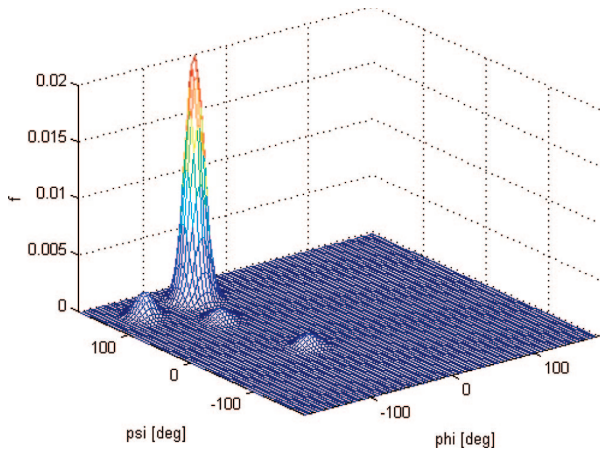

Figure 1.10: 3-D plot of the distribution function used for the optimized simulation of the amide I′ profiles and J coupling constants of trialanine as obtained from ref (90).

1.2.4.5 Intrinsic Propensity of Alanine

While the above referenced studies on short oligoalanines are useful in ascertaining the high preference of alanine for the pPII conformation, they also reflect the influence of neighboring alanine residues, and hence, one critical issue has not been explicitly dealt with, namely, the "intrinsic" conformational propensity of alanine. Kallenbach and associates addressed this issue by exploring the pPII population of alanine in glycine based host-guest system $AcGGAGGNH_2$ (50). Utilizing an approach reminiscent of their previous XAO study they combined NOE data and ^3J coupling

constants derived from NMR as well as UVCD to show that alanine in $AcGGAGGNH_2$ predominantly samples pPII, but to a lesser extent than the seven alanine containing XAO peptide (49) In a subsequent study, Hagarman et al. (54) measured and analyzed the J coupling constants earlier utilized by Graf et al. (74) for polyalanines and amide I‘ profiles of the unblocked tripeptide GAG. They found a pPII fraction of 79% for alanine, which is slightly lower than the value obtained for AAA (84%), thus indicating that alanine neighbor indeed stabilizes the pPII conformations slightly (54).

1.2.5 Theoretical Studies on Alanine

Over the last fifty years the alanine dipeptide (AdP) has been the classical model system for studying the conformational sampling of amino acid residues in the unfolded state. Numerous MD studies still use this peptide to explore the conformational propensity of alanine in water (56, 91-100). Most of the results of these and similar studies on other alanine based peptides still predict statistical-coil like distributions for short alanine peptides (88, 101, 102) and hence are at odds with the most of the aforementioned experimental results. This does not concern only the above described propensity studies but also simulations of e.g. helix-coil transitions, for which many MD force fields just overestimate the nucleation parameter σ owing to their intrinsic oversampling of helical conformations (101). This tendency is kept alive even in some more recent MD simulations. For instance, Beck et al reported ‘intrinsic’ propensities of amino acids using *in lucem* MD simulations in the host-guest motif GGXGG and

reported a significantly lower pPII propensity (16%) and an α-helical propensity above 50% for alanine (103). Generally, the results of MD simulations with regard to unfolded peptides depend heavily on the choice of the force field, as demonstrated with systematic comparisons by Zagrovic et al. (88) and Kwac et al. (96).

Among several attempts to move the results of MD simulations closer to experimental constraints, some success is noteworthy. In a computational study by Mu et al. in which the GROMOS package was utilized with Amber 09, it was also found that extended pPII and β conformations dominate for trialanine (65% and 12% respectively) (104). Gnankaran and Garcia found that the Amber 96 forcefield could be forced to yield good agreement with experiment (i.e., high pPII content for polyalanines) only by elimination of a backbone dihedral potential (105). However, the physical rationale for these changes remains obscure. Moreover, Garcia found pPII to be particularly stable for tetra-alanine, observing that four residues are necessary for an optimal backbone solvation in pPII (105). In line with Garica's results, Kentsis et al. (106) utilized ergodic sampling algorithms to show that in explicit water pPII is the predominant state of polyalanines. Contrary, Sosnick and coworkers used many different force fields to calculate the MD populations for the central residue in trialanine and found a significantly reduced pPII propensity for most force fields, except for the OPLS-AA-97 force field, for which they reported >80% pPII propensity (107).

More recently, Best and Hummer tried to remedy the confusing situation by using Garcia's insights but avoiding their drastic change of AMBER force fields (102). They modified rather than eliminated the dihedral backbone force constants to reproduce experimental data on the fraction of helix measured in short peptides. In addition to re-paramaterizing the force fields based on quantum chemical calculations, each was optimized to account for helical content of short peptides by modifying the torsional dihedral angel functions. However, though these authors were able to mostly reproduce the J-coupling constants reported for polyalanines, their distribution still yielded a much lower pPII content than that reported by Graf et al. (74) for the respective polyalanine. To resolve this contradiction, Verbaro et al. (108) used the distributions of Best and Hummer (102) to simulate the amide I' band profile of the VCD and IR spectrum of the A5W peptide and found them to be incapable of reproducing the strong amide I' VCD couplet. Moreover, these distributions led to an overestimation of the end to end distance, which Verbaro et al. (108) assessed by fluorescence resonance energy transfer (FRET) measurements. A subsequent modification of distributions led to a higher pPII fraction and the reproduction of all experimental data, including the end-to-end distance.

Although much of the discrepancy with MD simulations of unfolded peptides likely resides in force field parameterization, some studies have indicated that the underlying reason might be more complex. In particular, the choice of water model was found to be important. Realizing the above shortcomings of MD simulations to reproduce many experimental observations for short peptides, Duan et al. used quantum mechanical

methods with continuum solvent models and an effective dielectric constant (ε=4) to account for the polarizability in the system (109). The authors were thus able to determine that the pPII region is the most favorable. In a more recent approach by Lanza et al. (110), N-terminal blocked alanine peptides Ac-Ala_n-NH2 (N=2-4) were studied using MP2, CCSD(T) and DFT *ab initio* methods with implicit hydration. These authors found that nearly all major conformations (α- helical, pPII, β-strand, and turn-like conformations) as well as a large number of mixed structures are energetically accessible, more in agreement with statistical coil models. This result is in agreement with results from Han et al. (76), who showed on a lower level of theory that obtaining a preference for pPII requires the explicit consideration of water.

Many of the above referenced studies point to the role of hydration for conformational stabilization and, with respect to theoretical calculations, the choice of water representation (i.e., implicit vs. explicit solvation) and water model. Indeed, the water model seems particularly relevant for simulating unfolded states and intrinsically disordered peptides, for which energetic differences between conformations are generally on the order of RT, and interactions with water are enhanced due to larger solvent exposure. The most commonly used water models for MD studies are the so called 'three site' water models, namely, TIP3P, SPC, and SPCE, and the four site water model TIP4P (111-115). Although these models are still widely used in the protein community, numerous studies have shown that the resulting conformational ensembles obtained by using these models are far from being in agreement with each other and with

experimental findings. Hence, many attempts have been conducted to re-parameterize these water models. Notable in this regard is a study conducted by the Head-Gordon group in which they re-parameterized the standard TIP4P water model (TIP4P-ew) with inclusion of both Coulomb and Lennard-Jones long-range interactions (114). The authors found that the TIP4P-Ew water model yielded slightly better agreement with experimentally measured scalar couplings for trialanine (114). However, they still obtained a conformational ensemble with under-estimated (relative to many experimental results) pPII fractions (pPII 52%, 40% β). Using this water model Wickstrom et al. (116) demonstrated that ensembles generated with TIP4P-Ew predicted the NMR scalar couplings for trilanine reported by Graf et al. (74) more accurately than ensembles generated with TIP3P model.

1.2.6 Physical Basis for pPII Stabilization

The physical reason for the preferred sampling of the pPII trough by alanine is still not well understood and is still widely debated in the literature. The most prominent theories involve solvation (72, 77, 117-121), side-chain backbone interactions (122), and electrostatic and steric effects (107, 117, 123), as key driving forces for inducing conformational biases towards extended regions of the Ramachandran plot. Most theories, however, and a limited number of experiments emphasize the role of solvation for the stabilization of pPII (76, 93, 106, 124), while they still disagree on the underlying mechanism. Qualitatively, one can imagine that water interactions with the backbone and

side chains will deviate between different extended conformations (i.e., pPII and β-strand) by simply examining the backbone orientation of these structures (Figure 1.11).

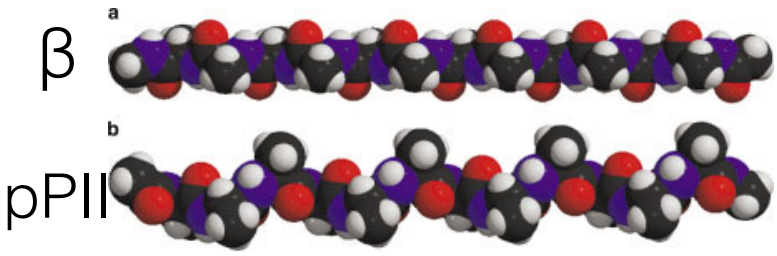

Figure 1.11: Backbone structure of polyalanine in β-strand and pPII conformation. Taken from ref (119) and modified.

Results from a computational density functional theory (DFT) based study on the alanine dipeptide in explicit water (i.e., the peptide complexed with two and four water molecules), suggest that water stabilizes pPII by forming a bridge between adjacent carbonyl and amide groups by hydrogen bonding (76). For illustration, Figure 1.12 shows the optimized pPII structure coordinated to water obtained by Han et al. Similarly, MD simulations of various polyalanines in explicit water with a modified Amber force field led Garcia to the conclusion that the pPII conformation is optimizing the packing of water molecules in the hydration shell of the peptide. The authors found that in the pPII

conformation here exists a high water density chain where water is delocalized and not pinned around polar atoms (Figure 1.12).

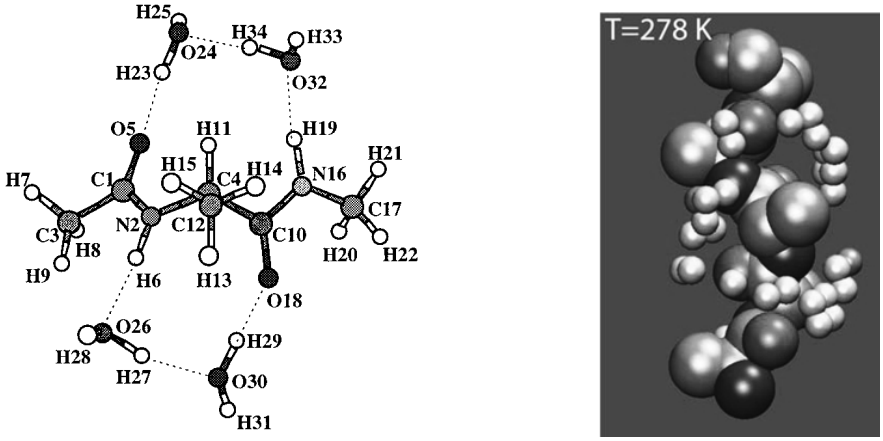

Figure 1.12: (Left) Optimized pPII structure of AdP obtained when coordinated to water. Taken from ref (76) and modified. (Right) Sites of high water density around pPII conformations at 278K. Taken from ref (119) and modified.

A somewhat similar rationale has been suggested by Kentsis et al. based on MD simulations of a series of GGXGG peptides (106). They found that only alanine avoids a disruption of the backbone hydration shell, which explains why it seems to be the only residue with a high pPII propensity. However, a different explanation was given by Drozdov et al. (93). They performed MD and Monte Carlo simulations for the alanine dipeptide in explicit water and found that peptide-solvent interactions stabilize compact

conformations that do not include pPII. However, water plays a role by attenuating attractive electrostatic interactions between the peptide atoms. This leaves steric interactions as the decisive force for determining conformational preferences. pPII minimizes these interactions and thus becomes the most prominent conformation. This notion was recently confirmed by Law and Dagett (125), who found that the occurrence of pPII conformations in proteins does not correlate with the existence of water bridges. In a study by Mezei et al. (119), it was found using Monte Carlo type simulations that the solvation free energy of pPII is favored over both β-strand conformers by 0.7 kcal/mol/ residue, and over α-helix by 2.7 kcal/mol/residue. The only parameter that was found to correlate with this solvation free energy was the first hydration shell solvation energy.

All these different theories have in common that they emphasize the role of water and solvation for the stabilization of pPII. This notion seems to be in good agreement with experimental results. Eker et al., for instance, showed that the pPII content of the blocked AcAAOH peptide is practically eliminated if water is replaced by DMSO as solvent (72). Simillarly, Liu et al. used CD spectroscopy to investigate the solvent dependence of the conformational distributions of $AcGGAGGNH_2$ (126). These authors interpreted their data as indicating that the pPII content of this peptide decreased in the order water > methanol > ethanol > 2-propanol, which reflects a linear correlation between pPII content and solvent polarity. However, no correlation between the dielectric

constant of the solvent and pPII content was found, which is somewhat at odds with the model proposed by Drozdov et al. (93).

1.3 Conformational Studies of Non-Alanine Residues

1.3.1 Non-Glycine Based Systems

While experimental results have generally converged in measuring a high pPII propensity for alanine, conformational propensity studies on residues other than alanine are more limited in number and vary in terms of the choice of short peptide model system. One of the first studies in this regard was conducted by Creamer and colleagues where they explored the residue level bias for pPII by studying short polyproline based peptides, PPP-X-PPPGY where x=A, G, V, L, I, N, Q, and M, with UVCD spectroscopy (118, 122). They found that proline has the highest pPII content followed by glutamine and alanine. The lowest pPII propensities were found for branched amino acid residues (I, V).

Eker et al. chose an alanine-based host-guest system, AXA, where x= G, V, M, H, S, P, L, K, Y, and F, to experimentally investigate the conformation of guest residues in aqueous solution (73). They combined Fourier transform IR, polarized Raman spectroscopy, and vibrational CD measurements of the amide I′ band profile of alanyl-X-alanine tripeptides in D_2O to obtain the dihedral angles of a representative conformation of their central amino acid residue. The obtained results led them to sort the investigated peptides into three classes. Valine, phenylalanine, tryptophan, histidine, and serine predominantly adopt an extended β-strand conformation.

Cationic lysine and proline prefer a pPII-like structure. They found alanine, methionine, glycine, and leucine populate these two conformations with comparable probabilities.

More recently, Grdadolnik et al. performed a comprehensive experimental study and reported propensity scales for 19 non-proline residues in blocked dipeptides (i.e., XdP) based on an analysis of the amide III' region of their Raman and IR spectra as well as the φ-dependent J coupling constant (127). By measuring the amide III' region with both attenuated total reflection IR spectroscopy and Raman spectroscopy, they assigned three resolvable sub-bands to pPII, β, or right-handed helical conformers. By considering these three main conformations they were able to obtain pPII fractions of 0.68 and 0.53 and β fractions of 0.17 and 0.43 for alanine and valine, respectively.

1.3.2 Glycine Based Host Guest Studies

In the attempt to determine the mole fraction of pPII and other backbone conformations reflecting the intrinsic propensity of amino acid residues, many researchers have turned to glycine based model systems. Glycine-based host-guest systems had been frequently used as models beforehand to obtain $^{3}J(H^{N}H^{\alpha})$ and chemical shift values for what researchers considered as "random coil" conformations of amino acid residues (128-131). Glycine as neighbor is considered as an ideal reference residue as it provides minimal nearest neighbor interaction due to its hydrogen based side chain,

thus allowing for an accurate intrinsic propensity scale. Moreover, since glycine is non-chiral, CD spectra can be expected to reflect solely the conformational distribution of the X-residue.

One of the first conformational propensity studies was conducted by Shi et al., who chose Ac-GGxGG-NMe in water, where "x" represents the 19 natural amino acids except glycine, to investigate the guest residues conformational preference in the unfolded state (51). Shi et al. measured the UVCD spectra for 18 residues (glycine was excluded, alanine had been measured previously) at three different temperatures. For most of the investigated peptides the spectra exhibit isodichroic points, indicating a two-state model, which was assigned to a pPII/β equilibrium. To obtain quantitative information on the amount of pPII/β present, the authors utilized the conformational sensitive $^{3}J(H^{N}H^{\alpha})$ coupling constants derived from H NMR experiments. The $^{3}J(H^{N}H^{\alpha})$ coupling constant varies dramatically in the ϕ region between pPII and β strand (-65° and 125° , respectively) and thus provides a rather sensitive measure of the angle ϕ. (*see Ch. 2 of this thesis)*. The authors reported their results as indicating high proportions of pPII conformations at low temperature for nearly all investigated residues, shifting to an increasing β-strand population at high temperature for all peptides. The pPII content thus obtained ranged from 40 to 80% depending on the residue, with alanine exhibiting the highest pPII propensity of 83%. However, the data analysis of Shi et al. (51) has been challenged on various accounts. The interpretation of the experimental ^{3}J coupling constants reported are questionable, because (1) the $^{3}J(T)$ coupling constants should

represent Boltzmann-averaged conformational distributions rather than just two isolated conformations, (2) experimental ^3J coupling constants are generally three-fold degenerate with respect to the angle ϕ and hence could be reproduced with various conformational ensembles and (3) the reference ^3J coupling constants used to represent pure pPII and β in these studies were obtained from coil libraries of Avbelj and Baldwin which may not actually reflect real distributions of amino acid residues in unfolded peptides (132).

1.3.2.1 GxG Peptides

A detailed conformational analysis aimed at obtaining realistic conformational distributions of amino acid residues in glycine based host systems was performed by Hagarman et al. (54). They used unblocked GxG peptides as the host guest system. Individual conformational distributions of the guest residue were described by constructing variable two-dimensional Gaussian functions in the (ϕ,ψ)-space. This way, the authors were able to avoid using less realistic representative structures. Using these functions in a global analysis of amide I' band profiles and various J-coupling constants (see *Ch 2 of this thesis),* they were able to obtain Ramachandran plots for 15 amino acid residues. A representative set is shown in Figure 1.13. Their results confirm the high pPII propensity of alanine (79%), as well as the variable conformational preferences of all other residues. However, alanine is the only residue with such a high pPII fraction. The pPII scale derived by Hagarman reads as follows: A, M> L, E, R, K> >I, V, S> D,N,T,C, with most residues exhibiting a lesser pPII content than the corresponding values reported

by Shi et al. (51). Figure 1.14 shows a comparison of the pPII and β-strand content obtained for various amino acid residues in the aforementioned studies.

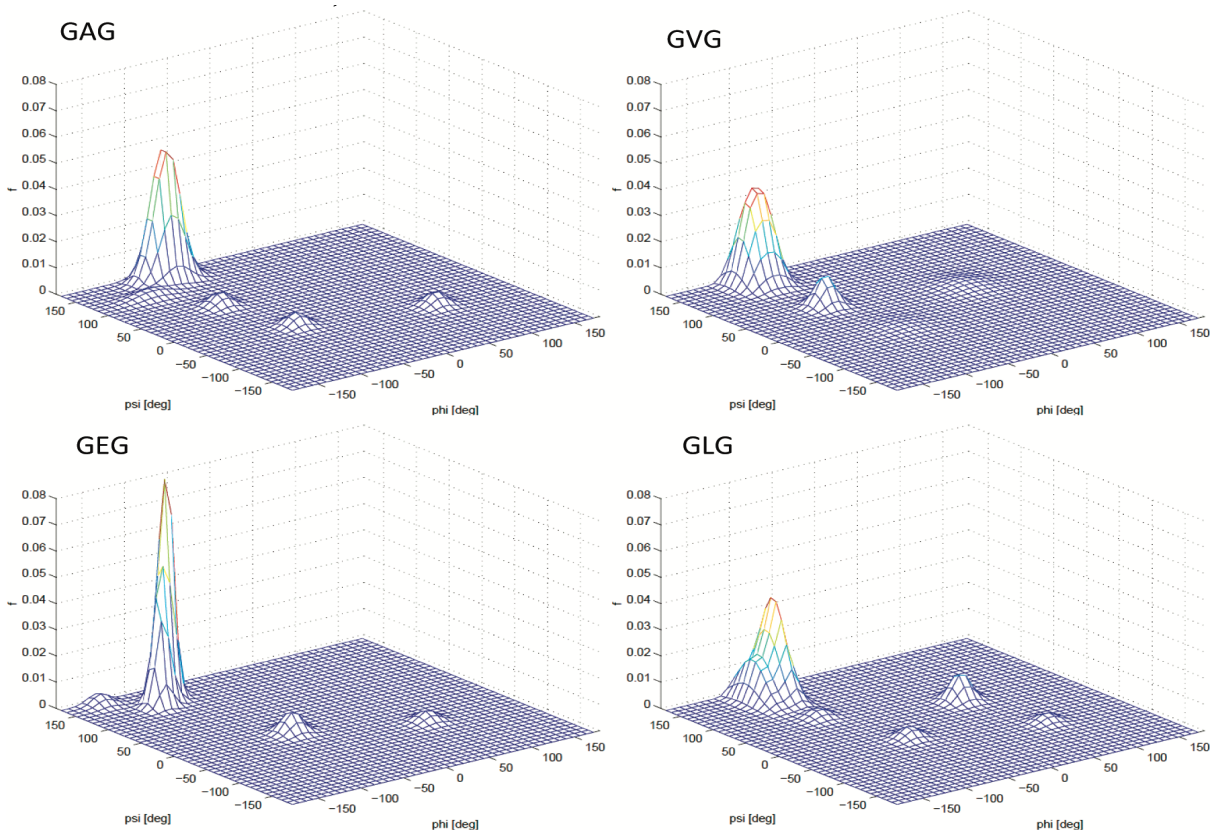

Figure 1.13. Conformational distributions of the central residue in GAG, GVG, GLG, and GEG obtained from analysis of amide I' band profiles and J coupling constants, illustrating the 2D distribution approach used by Hagarman et al. Taken from ref (54).

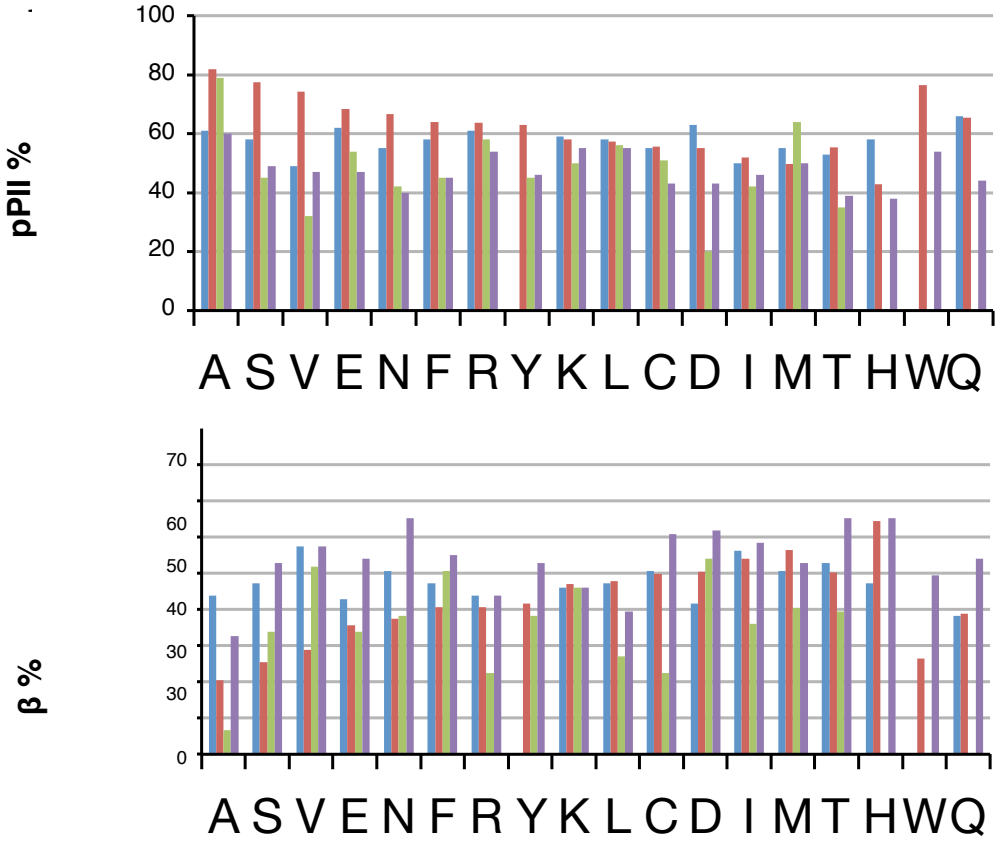

Figure 1.14: Comparison of pPII and β-strand populations for guest residues as derived by Rucker et al. for PxP (blue) (118), Shi et al. for GGxGG (red) (51), Hagarman et al. for GxG (green) (54, 55), and Grdadolnik et al. (133) for XdP (purple).

1.3.2.1 Intrinsic Turn-like Populations

Reproduced in part, from:

I. Hagarman, A., Mathieu, D., Toal, S., Schwalbe, H., Schweitzer-Stenner, R. "Amino Acids with Hydrogen-Bonding Side Chains have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution." Chem. Eur. J. 2011, 17, 6789 – 6797

II. Rybka, K., Toal, S., Verbaro, D., Mathieu, D., Schwalbe, H., Schweitzer-Stenner, R. "Disorder and order in unfolded and disordered peptides and proteins: A view derived from tripeptide conformational analysis. II. Tripeptides with short side chains populating asx and b-type like turn conformations." Proteins, 2013, 81, 968–983.

In addition to distinct biases for pPII and β-strand populations, in a subsequent study, it was found by us that a sub-set of amino acid residues (D, N, T, S) investigated in GxG peptides have an unusually high preference for so-called "turn-like conformations" (55). The term 'turn-like' herein means that the respective dihedral angles adopted by the investigated residue are assignable to the β-, ɣ- and/or asx turn regions in the Ramachandran plot, which are formed by turn motifs in proteins. In the case of ɣ- and asx-turns this could lead to intra-peptide hydrogen bonding and thus to the formation of real turns, so that the use of the terms 'turn' or 'turn-like' is justified.

Turn motifs are ubiquitous in folded proteins, in which they are key determinants, particularly for protein tertiary structure. Turn sequences, for the most part, reside on protein surfaces and as such are comprised of mainly hydrophilic residues (134). Turn motifs are increasingly recognized as relevant not only for protein folding but for formation of local order in IDPs. Wright, Dyson and coworkers have provided NMR-based

evidence for a role of turns in the initial phase of protein folding (135). Some 'native-like topologies' have been detected in the denatured state of proteins. Mukrasch et al., for instance, showed that an otherwise disordered tau protein contains short aspartic acid-rich segments with a preference for type I β-turn conformations (31).

Turn types in proteins can be broadly classified according to the separation between the two end residues with β-turns defined as i→i±3 type and ɣ-turns defined as i→i±4 type being the most common. These turns in proteins may or may not be stabilized by intra-peptide backbone-backbone H-bonding between end residues (3). Turns can be further classified into sub-types designated I-VIII depending on the preferred dihedral angles within the turn (Table 1.1). Each turn type additionally has a backbone mirror-image conformation designated I'-VIII'. The two most common types of turns are (1) β-type I with canonical backbone dihedral angles for the i+1 and 1+2 positions of (-60°, -30°) and (-90°, 0°), and (2) β-type II with canonical backbone dihedral angles for the i+1 and 1+2 positions of (-60°, 120°) and (80°, 0°) (3). For visualization, Figure 1.15 displays two types of β-turns.

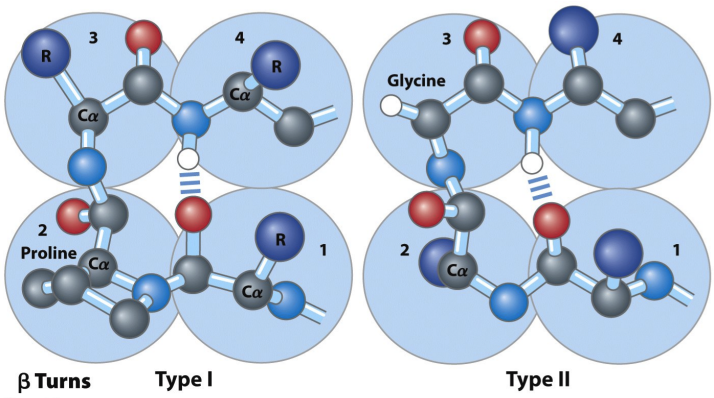

Figure 1.15. Type I and II β-turns shown for a tetrapeptide. Taken from ref. (3) and modified

In addition so-called "asx" turns are a special type of i, i+2 turn motif found in proteins in which the first residue (typically aspartic acid or asparagine) has a side chain that may H-bond with the i+2 residue (136, 137). Similar motifs occur with serene and theronine, which are called ST-turns (138). Detailed analyses of turns in a database of proteins with different folds revealed that residues with short, polar side-chains, such as S, T, D, and N, also have a high tendency to be located in the i, i+1, i+2 and i+3 positions in β-turns. Duddy et al. showed for a database of 500 proteins that 10% of their N, D, S and T residues are involved in these asx- and ST- type turns (139).

Table 1.1: Common turn types and associated dihedral angles at the i+1 and i+2 position.

Turn Type	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
I’	60	30	90	0
II	-60	120	80	0
II’	60	-120	-80	0
VII	-60	-30	-120	120
IV	-60	10	-53	17

It is generally thought that the existence of turn conformations requires formation of non-local interactions. For instance, NMR evidence suggests that short peptide segments with proline at the i+1 position (i.e., YPxyV) can form turn structures in solution if x and/or y are occupied by residues that rank high on the Chou–Fasman scale of turn propensities (12). Whether formation of turn structures is also possible for short peptides that do not contain proline is unclear. It was previously thought that the formation of asx- or ST-turns in aqueous solution requires the presence of either P or the unnatural amino acid aminoisobutyric acid (Aib) at the ith position or an insertion of a d-amino acid at the i + 2 position (140). To address this, Hagarman et al. used the aforementioned analysis of both amide I’ band profiles and J-coupling constants to extract the conformational ensemble of these residues (55). It was found that D, N, T, and

S residues in model GxG peptides indeed sample dihedral angles similar to those found in the corner residues of various turns, namely, type I/I', II/II', and IV b-turns in the absence of non-local interactions (55). The side chains of the central residues of these three peptides have in common that they are short and exhibit hydrogen bonding capability. The total of turn populations for these residues are abnormally high in comparison to all other residues and range between 20% for asparagine and 32% for aspartic acid. Aspartic acid was found to predominantly sample regions at the upper border of the upper-right quadrant of the Ramachandran plot, which bear resemblance to asx-turns observed in proteins. This asx conformation enables intra-peptide hydrogen bonding between the side-chain carboxylate of aspartic acid and the C-terminal amide group. Figure 1.16 shows the total conformational ensemble obtained for aspartic acid for reference. These finding show that the high propensity for D, N T, S to be located in turn motifs reflects an intrinsic property and supports the role of these residues as initiation sites for hierarchical folding processes that can lead to compact structures in the unfolded state.

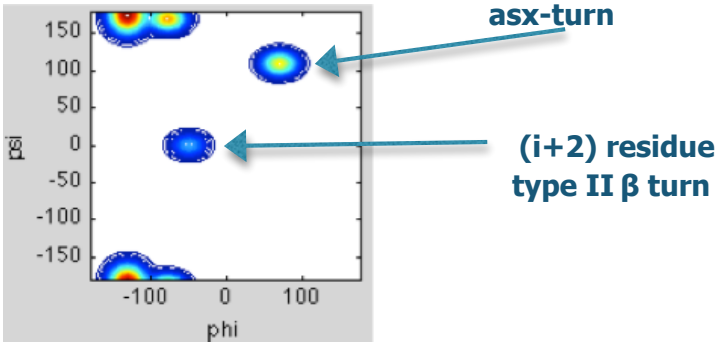

Figure 1.16: Conformation distribution of the central residue in unblocked GDG illustrating the high turn propensities of aspartic acid. Taken from (55) and modified.

In a separate analysis, the side-chain involvement in turn formation was analyzed by determining the homo-nuclear $^{3}J(H^{\alpha},H^{\beta})$ coupling constant (141). As demonstrated for short model peptides by West and Smith (142), and for unfolded proteins by Hennig et al. (143), Hahnke et al. (144), and Sziegat et al. (145), this coupling constant yields information about the population of side chain rotamers with respect to the angle χ_1 ($\angle$ N-Cα-Cβ-Cɣ). These coupling constants were used to stereospecifically assign the two diastereotopic Hβ protons of the central residues D, N, and C within GxG peptides and determine the respective rotamer populations by use of a Pachler model (146). This model assumes three different rotamers of the side chains, which are associated with χ_1 angles of 60°, 180°, and -60°. By analyzing the coupling constants as a function of temperature in terms of a Pachler analyses, the mole fractions of each rotator could be

obtained. These mole fractions of turn conformations as a function of temperature are displayed in Figure 1.17 for the fully protonated and fully ionized state of GDG and GNG (141). Within their limit of accuracy, the data do not depict any discernible temperature dependence. This is a surprising result since one would expect that turn-like conformations are entropically disfavored and therefore, less likely to be populated at higher temperatures. However, a recent computational study aimed at exploring the behavior of the YKGQ sequence in a hairpin-forming peptide found that it can form turns at low (room) as well as at high temperature because enthalpy and entropy are both stabilizing (147).

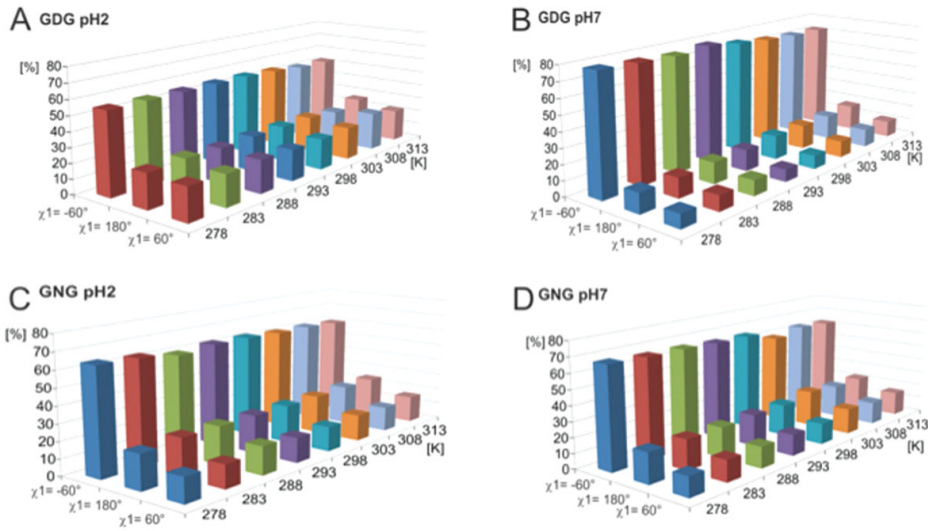

Figure 1.76: Visual representation of the mole fractions of the side chain rotamers of the indicated peptides plotted as a function of temperature. Taken from ref (141).

1.2.7 Coil Libraries

An alternative strategy for determining the conformational preferences of amino acids in the unfolded state is the analysis of conformational distributions found in coil libraries. In general, there are two main types of coil libraries constructed for structure prediction. “Unrestricted” coil libraries simply consider large sets of proteins from the Protein Data Bank (PDB) without any resections as to the secondary structure elements of selected proteins. This strategy was based on the argument that the contextual influence would be eliminated if one averages over many protein environments (36, 148). Support for this assumption is based on the linear correlations between average $^{3}J(HNH\alpha)$ coupling constants derived from unrestricted libraries and corresponding values obtained for $AcGGxGGNH_2$ peptides in water. However, distributions obtained from these types of libraries are still noticeably biased towards right-hand helical conformations and do not agree with experimentally derived distributions of amino acids in solution (148, 149). In contrast to the unrestricted library, only a subset of the database can be chosen (i.e. a restricted coil library) in which certain types of secondary structure sequences are purposefully omitted. This way, it is assumed that any possible effect of the secondary structure on the amino acid conformation is effectively eliminated, hence mimicking the unfolded state.

Swindells et al. constructed a restricted “coil library” along this line by considering only residues in coil regions and omitting residues lying within alpha helical or beta strand structures within a dataset of 85 proteins from the PDB (150). By using this

library they determined that correlations between intrinsic conformational propensities and observed secondary structure propensities for helices are modest and strongest for β strand propensities/structures. Serrano took this notion a step further by constructing a coil library in which all regular secondary structures were omitted, including amino acids in β-turns, which could also have associated nonlocal interactions (148). As displayed in Figure 1.17, the author was able to show that the distribution for alanine dramatically changes from mostly right-handed helical to mostly pPII when removing all stable secondary structure conformations, in qualitative agreement with most experimental studies on alanine based peptides. However, as noted by Jha et al. (149), these restricted coil libraries may have a systematic bias towards the pPII conformation due to their inclusion of residues at the ends of structured regions which would inherently disfavor sheets and helices (and hence favor pPII conformations).

Jha et al. (150) constructed a much more restricted coil library in which they tried to eliminate this bias by omitted residues flanking regular secondary structures (including prolines) and considering only residues that lie in "coil" stretches of four or more. Their resulting data still indicated a moderate preference for the pPII conformation in the most restricted coil library (library without helices, sheets, turns, and terminal, pre-proline, and most exposed residues) for most residues, with the highest pPII preference seen for alanine (49%). High pPII levels were found even for buried residues indicating that preferential solvation may not be the only contributor to pPII preference in the unfolded state.

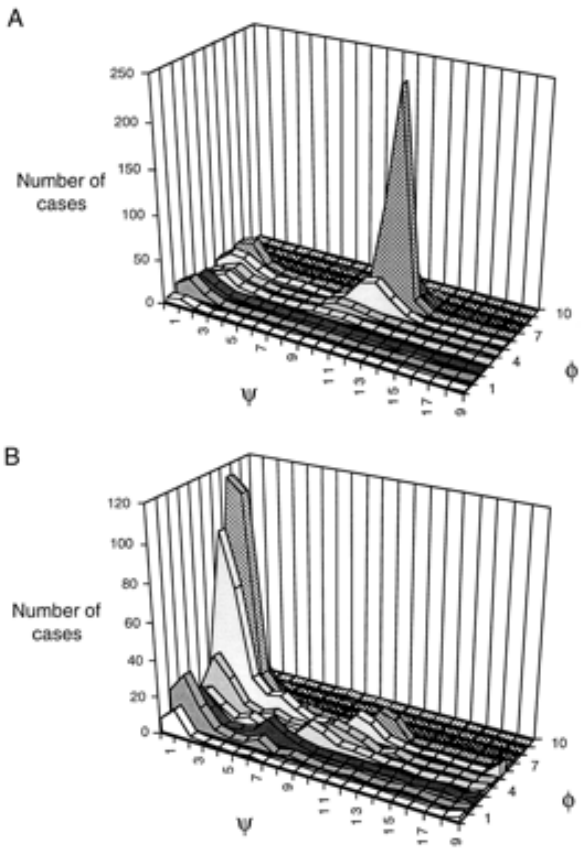

Figure 1.17: Distribution obtained in the upper left quadrant of the Ramachandran plot for alanine using (A) all secondary structure conformations in the protein database (i.e. an unrestricted library) or (B) only those alanine residues in a coil conformations (i.e., helices, sheets, turns omitted). Each unit reprints 18°. Taken from ref. (150) and modified.

1.2.8 Nearest Neighbor Influences

One of the pillars of Flory's random coil model for the unfolded state of proteins is the isolated pair hypothesis (IPH). It stipulates that the conformational distribution of a distinct residue in the polypeptide chain is independent from the nature and the adopted conformation of adjacent residues (22). As a consequence, a random coil chain carries a

significant amount of conformational and combinatorial entropy, which led Levinthal to propose his paradoxon (8). The total conformational entropy and energy of a chain could then be written as a sum of residue enthalpies and entropies. Since the conformational distribution of different amino acid residues were expected to be comparable (with the exception of proline and glycine), enthalpy and entropy would linearly increase with the number of residues.

Over the last ten years the validity of the IPH has been questioned based on experimental, computational, and bio-analytical results. One of the first pieces of evidence for deviations from IPH came from the analysis of $^{3}J(H^{N}H\alpha)$ constants of a 130-residue fragment of the unfolded fibronectin-binding protein from Staphylococcus aureus, a protein with three homologous and a terminal segment (151). In this study, Penkett et al. showed that the average $^{3}J(H^{N}H^{\alpha})$ constants of individual amino acid residues in this protein correlate well with expectation values derived from ϕ distributions of coil libraries. However, a more detailed analysis revealed that individual $^{3}J(H^{N}H^{\alpha})$ coupling constants depend on the nature of the respective neighbors. With respect to the strength of this nearest neighbor interaction, Penkett et al. divided the investigated residues into two categories: category L contained aromatic and branched aliphatic side chains (V, I, F, W, Y) and category S, which contain all the other residues with the exception of glycine. Category L residues were found to increase the $^{3}J(H^{N}H^{\alpha})$ coupling constant of their downstream neighbors (0.4 Hz on average), whereas S-residues had a negligible influence. The authors interpreted the influence of L-residues as an equilibrium

shift from α-helical to β-strand conformation. They invoked steric effects as a reason: extended β-strand conformations exhibit less steric strain than right handed helices.

Avbelj and Baldwin supported the interpretation of Penkett et al. (151) by a computational thermodynamic analysis (152). The result of their study suggest that nearest neighbor interactions are dictated by solvent mediated processes rather than by direct steric interactions. Contrary to Penkett et al. (151), they took pPII as a distinct residue conformation into consideration. Three contributions to the ϕ-dependence of the overall Gibbs energy were discussed: (a) the torsional potential V(ϕ), (b) electrostatic interactions and (c) protein/peptide solvation and its modification by side chains. V(ϕ) exhibits a maximum (i.e. its most destabilizing effect) at 120°, thus, it does not favor β-strand. On the contrary, electrostatic interactions favor β-strand like conformations, but solvation screens Coulomb interactions thus allowing a dominance of V(ϕ), which in turn favors pPII. If side chains (like valine's β-branching side chain) perturb the hydration shell, they reduce the pPII propensity. Avbelj and Baldwin further showed that the influence of side chains on propensities is not limited to their own residue (152). They are capable of affecting the solvation of nearest neighbors as demonstrated in Figure 1.19, which shows how the substitution of the fifth residue of a heptaalanine peptide by the β-branched valine increases the solvation Gibbs energy both in the pPII and the β-strand conformation. An increase in energy here means less solvation, since the solvation Gibbs energy is negative. The effect is much more pronounced for pPII than for β, thus causing a preferential destabilization of the former. These changes clearly affect the solvation of

the neighbors, where it also favors the β-strand conformation. By performing similar calculations for other amino acids, Avbelj and Baldwin (152) found that branched residues and residues with aromatic side chains exhibit much stronger nearest neighbor effect.

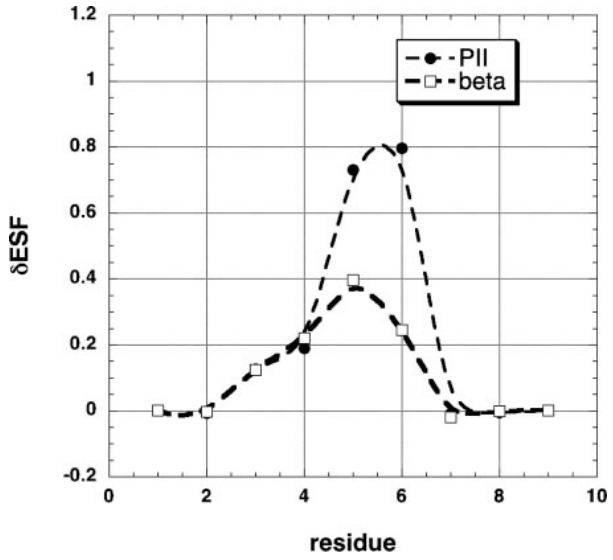

Figure 1.19: Representation of the change of the electrostatic solvation free energy induced by substituting the fifth alanine residue of a heptaalanine peptide by valine. Changes are plotted for pPII and β-strand conformations, as indicated. Taken from ref. (152) and modified.

Thus far the most detailed information about nearest neighbor interactions came from coil library studies of Sosnick and coworkers (107, 149, 153, 154). For instance, Jha et al. (149) observed (a) that the contributions of nearest neighbors to apparent propensities

of an amino acid residue can be significant, (b) that these interactions are both side chain and conformation dependent and (c) that they affect the propensities as well as the position of local maxima of distributions. In order to simplify the representation of nearest neighbor effects, they classified neighbors according to the properties of their side chains: β-branched (and aliphatic), aromatic and alanine like. G and P were not considered (149). The influence of these different types of amino acid residues on the pPII, β-strand, and helical content of alanine obtained by Sosnick et al. (107) are visualized in Figure 1.19. The authors found that if the side chain is β-branched (V, I), the pPII and β-fractions are actually slightly above the average value in coil libraries, but once the neighbor adopts a helical conformation, pPII drops and β-strand increases. Aromatic neighbors have very limited influence on the distribution, while alanine like neighbors behave like valine and isoleucine. The situation is different for the C-terminal neighbor. An aromatic neighbor in pPII increases the β-strand population and decreases the helical content compared with β-strand and right handed helical. The pPII fraction of alanine is less affected (149), it is highest for a helical neighbor and lowest if the neighbor adopts β-strand. The influence of C-terminal β-branched neighbors is modest, but qualitatively similar to that of the corresponding N-terminal neighbor. Alanine-like N-terminal neighbors stabilize pPII and destabilize the helical conformation.

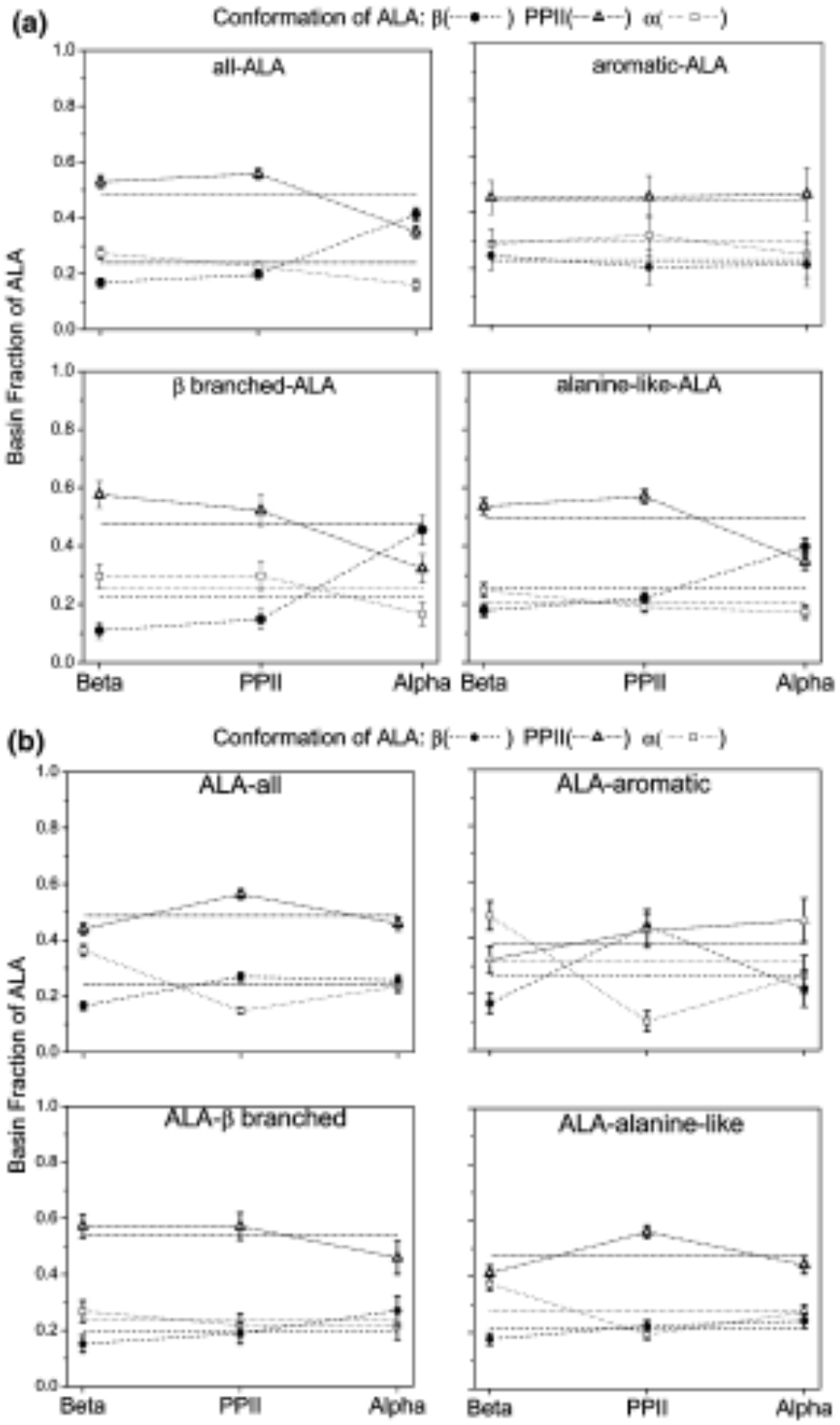

Figure 1.20: Changes in pPII, β-strand, and helical conformations in alanine induced by (a) N-terminal neighboring residues and (c) C-terminal neighboring residues. Taken from ref. (149) and modified.

The influence of the residue conformation on nearest neighbor interactions was also investigated by Pappu et al. (155). They used a simple hard-sphere model to explore the conformational space of a polyalanine peptide. The repulsive potential utilized in this model was augmented by hydrogen bonding formation for certain ϕ,ψ conformations. The Ramachandran space was subdivided into 49 quadratic meso-states and their population was calculated with Monte-Carlo methods. The authors thus found that steric clashes involving conformations in the right handed helical region of the Ramachandran plot (which violates the IPH), while such clashes were obtained when neighbors adopted conformations in the upper left quadrant of the Ramachandran plot. Hence, this work questions the validity of the IPH, but only for a limited set of conformations. This is a somewhat less complex picture than that arising from the coil library studies of Sosnick and colleagues, even though the latter also suggest a strong nearest neighbor effect involving residues in helical conformations (149).

A similar message emerges from other studies of the Sosnick laboratory by which they tried to reproduce residual dipole coupling (RDC) values of unfolded proteins observed with NMR experiments. To this end, Sosnick and coworkers used coil library distributions of amino acid residues with and without nearest neighbor influences (149). Considering the latter led to a much better reproduction of experimentally obtained coupling values, thus emphasizing the relevance of nearest neighbor interactions for the structural ensemble of unfolded proteins and peptides.

CHAPTER 2. THEORY BEHIND SPECTROSCOPIC METHODS

X-ray crystallography, and in some cases, NMR spectroscopy, are the most widely used methods for determining protein structure in native or thermodynamically stable states. However, studying unfolded or disordered states of proteins and peptides presents a challenge in this regard due to the inherent inhomogeneity and dynamics of the system. To study conformation in the unfolded state, the combined use of multiple, complimentary, spectroscopic techniques has hence proved beneficial, as one can extract different layers of information, which can then be combined to yield clear pictures of conformational distributions.

All of the studies performed herein combine the use of three main spectroscopic techniques, namely, ultraviolet circular dichroism (UVCD), Nuclear Magnetic Resonance (NMR), and various vibrational spectroscopies (infrared, Raman, and vibrational circular dichroism). The basic theoretical background needed to understand and analyze each spectroscopy is described in turn below. Specific materials and methods used within each study are described separately in *Chapter 3 of this thesis*.

2.1 ULTRA-VIOLET (ELECTRONIC) CIRCULAR DICHROISM

Circular dichroism (CD) is increasingly recognized as a valuable tool for evaluation of protein and peptide structure and conformation (61, 156). CD is a spectroscopic technique that takes advantage of the inherent chirality of protein

environments by measuring the differential absorption between left- handed and right-handed circularly polarized light:

$$\Delta A = A_L - A_R \qquad \text{(eq. 2.1)}$$

where A_L and A_R are left-handed and right-handed absorbance, respectively. If the circularly polarized light is absorbed to different extents by a chromophore, the resulting radiation possess elliptical polarization. A CD signal will be observed when a chromophore is chiral which, in the case of proteins and polypeptides, occurs primarily due to the asymmetric 3D environment (156). Generally, CD instruments report ellipticities (θ), which can simply be converted to absorption via:

$$\Delta A = \frac{\theta}{32980} \qquad \text{(eq. 2.2)}$$

Typically, units of Δε ($M^{-1}cm^{-1}$) are used for spectral comparisons which are calculated via Beer-Lambert's Law:

$$\Delta\varepsilon = \frac{\Delta A}{c \cdot b} \qquad \text{(eq. 2.3)}$$

where c is the concentration of chromophore (peptide) in units of molarity and b is the path length in centimeters. The parameter Δε ($M^{-1}cm^{-1}$) is herein referred to as "dichroism."

Ultraviolet circular dichroism (UVCD) is the application of the CD technique in the ultraviolet region. Typically, far UVCD (<250nm) is used to probe the chirality about the amide group. The amide group has four π electrons, three π orbitals (π_b, π_{nb} and π^*)

and two lone pairs in the non-bonding orbitals (*n* and *n'*) of the valence shell. Two ππ* transitions, $\pi_{nb} \rightarrow \pi^*$ (denoted NV_1, approximately 190 nm, electric dipole transition moment $|\mu| \approx 3.1$D) and $\pi_b \rightarrow \pi^*$ (NV_2, 139 nm, $|\mu| \approx 1.8$ D), and one $n\pi^*$ ($n \rightarrow \pi^*$, 220 nm) transition in the amide chromophore have been identified (157). Figure 2.1. shows a simple MO diagram depicting these transitions.

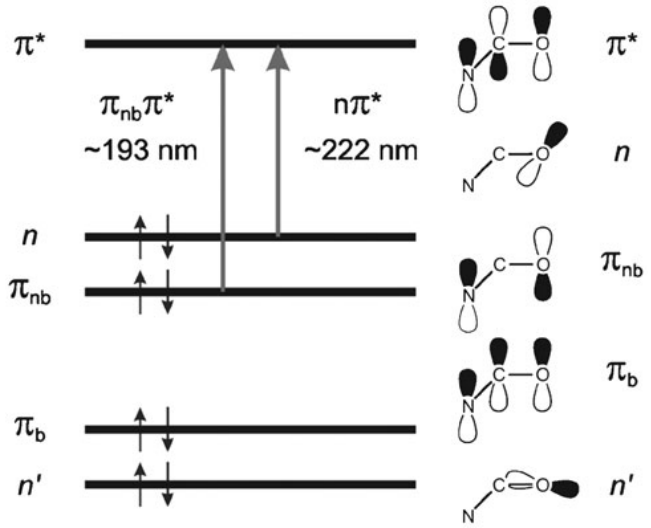

Figure 2.1: Electronic transitions of the amide group in the far-UV region. The molecular orbitals shown are the bonding, nonbonding and antibonding π orbitals (π_b, π_{nb} and π^*) and two lone pairs on the oxygen atom. Taken from ref (157) and modified.

The spatial arrangement of the amide groups along the polypeptide backbone determines the far−UV CD spectrum. Therefore, different types of secondary structures of proteins give rise to characteristic UVCD profiles in this region, as displayed in Figure 2.2.

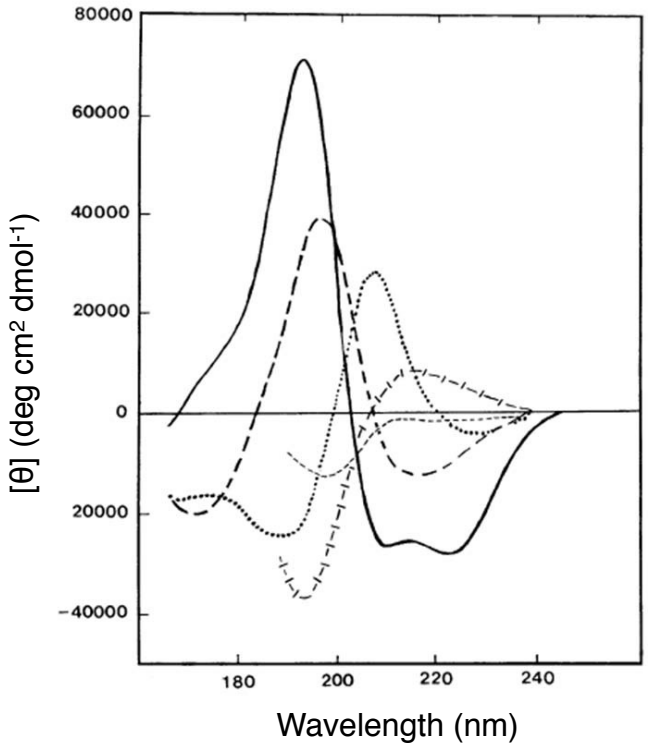

Figure 2.2. Far UVCD spectra of various secondary structures. Solid line: a-helix; long dashed line: anti-parallel β-sheet; dotted line: type I β-turn; cross dashed line: pPII helix; short dashed line, irregular structure. Taken from ref. (156).

2.1.1 Origin of CD Activity in Polypeptides

The integrated intensity of a CD band for a given transition gives a measure of the strength of CD, called the rotational strength. Rotational strength is defined theoretically, as the imaginary part of the scalar product of the electric (μ) and magnetic (m) dipole transition moments of an electronic transition:

$$R = c\int \frac{\Delta\varepsilon}{\lambda} d\lambda = \mathrm{Im}\left[\vec{\mu} \bullet \vec{m}\right] \qquad \text{(eq. 2.4)}$$

where c is a constant, which in cgs units is approximately 2.295 X 10^{-39} erg cm^3 (64). $\vec{\mu}$ is the electric transition dipole moment, which can be considered a measure of the linear charge displacement upon transition. $\vec{m}$ is the magnetic dipole moment transition, which can be envisioned as a measure of the circular motion of charge associated with the transition. Therefore, to give rise to a CD signal, a transition must involve both linear and circular charge movement resulting in helical charge displacement (64). Mathematically, this means that the angle between μ and m ≠ 90 and μ ≠ 0 and m ≠ 0.

For certain symmetries, electrically allowed transitions may be either magnetically allowed or forbidden depending on the particular geometry of the molecule. The electronic transitions in an achiral molecule have either the electric (e.g., $n\pi^*$ transitions) or the magnetic (e.g., $\pi\pi^*$ transitions) dipole transition moment equal to zero, or the two kinds of transition moments are perpendicular to each other (e.g., $n\sigma^*$ transitions), which results in zero rotational strength. For instance, if a molecule has a center of symmetry, the given electronic transition cannot be both electrically and

magnetically allowed. In a molecule with a plane of symmetry, a given electronic transition may be both electrically and magnetically allowed, however this would result in orthogonal relationship between the two, and hence a vanishing rotational strength. Therefore, a CD signal will not be observed unless a molecule is chiral.

For protein and peptide conformational analysis in the far UV, the amide group is the chromophore of interest. The planarity of the NHC group results in a plane of symmetry within the chromophore, and hence a vanishing rotational strength. Therefore the chromophore itself does not exhibit an intrinsic CD activity (156). In order to generate rotational strength, the environment must be perturbed by a chiral field. The interaction between protein chromophores in the (3D) chiral field of the protein introduces these perturbations leading to optical activity. This occurs through the coupling of electrically and magnetically allowed transitions through Coulomb interactions. Moffitt laid the foundation for describing the coupling in the excited state, which results in the exciton splitting of electronic transitions (158). There are two basic interactions of this type that lead to the observed CD in polypeptides. The primary method is the called coupled-oscillator or μ-μ* mechanism in which one member of a pair of oscillators induces a magnetic dipole moment in the second oscillator resulting in a composite transition with non-vanishing electric and magnetic moments. The second mechanism for induced rotational strength is the so called μ−*m* mechanism (63), wherein the electrically allowed transition on one chromophore mixes with the magnetically allowed transition on another. It is in this way that interaction between transition

moments leads to CD activity and the geometrical relation between transition moments in different secondary structures that determines the characteristic secondary structure UVCD spectra shown in Figure 2.2 (63).

A basic assumption for analyzing UVCD spectra of proteins is that the measured spectrum can be expressed as a linear combination of basis spectra that reflect the secondary structure content within the molecule. In the unfolded state, conformational ensembles are largely composed of high levels of extended conformations such as pPII and β-strand. The UVCD spectrum of pPII-type structures exhibits a weak positive band at approximately 215nm and a negative maximum at 195nm (Figure 2.2). The long-wavelength band (~215 nm) is generally assigned to the $n\pi^*$ transition and the short-wavelength band (~195 nm) to the $\pi\pi^*$ transition (64). Strong mixing with higher energy transitions are thought to be responsible for the non-conservative CD spectrum. A typical UVCD spectrum of displaying the pPII signal at low temperatures is shown below in Figure 2.3 for reference. With increasing temperature this signal generally decays due to increased sampling of β-strand conformations which displays the opposite CD signature (i.e. β-strand conformations exhibit a positive $\pi\pi^*$ band at 195 nm followed by a negative $n\pi^*$ band 215 nm., see Figure 2.2)

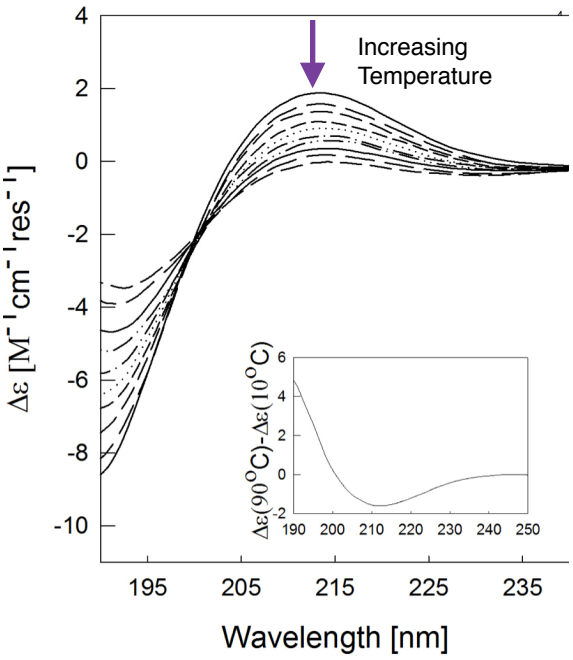

Figure 2.3: Far UVCD spectrum of trialanine as a function of temperature from 10-90°C. The inset depicts the difference spectrum.

2.2. NMR Spectroscopy

2.2.1 J-coupling Constants

NMR has long been used as a quantitative technique for structure determination of proteins and peptides. When conducting NMR on unfolded or disordered states, conformational fluctuations that take place on time scales faster than the time resolution of NMR (i.e. milliseconds) must be taken into account. For the highly dynamic unfolded state, resulting NMR-derived parameters are ensemble averages and reflect a mixture of

coexisting conformations interconverting on faster time scales (159). Within *this thesis*, the determination and analysis2of so called J-coupling constants plays a key role in conformational analysis. J-coupling, also referred to as spin-spin coupling, results when two nuclei separated by 1-3 bonds perturb the respective valence electrons. The alignment of electron and nuclear spins at one atom is communicated through the chemical bond to neighboring atoms which results in a splitting observed in the respective NMR signal. Since this interaction depends on the overlap of electron and nuclear wave functions, the resultant splitting is orientationally dependent, i.e. it depends on the intervening torsional angles.

2.2.1.1 Homo-Nuclear ^3J Coupling Constants

The importance of three-bond J-couplings (^3J) for determining three-dimensional molecular conformation by NMR methods is well-established (43, 159-162). For peptides and proteins, a wide variety of different J-couplings are commonly being used in the structure determination process. Each coupling constant has an associated Karplus equation (160), which relates the J-coupling constant to the intervening torsion angles. The respective coefficients for these equations have been empirically parametrized on the basis of known protein structures (162-164). The most widely used coupling constant in this regard is the so called $^3J(H^NH^\alpha)$ coupling constant which reflects the spin coupling between amino and the alpha carbon hydrogen, and hence is sensitive to the intervening ϕ-angle. The respective Karplus equation is given by (160):

$$^{3}J(H^{N}H^{\alpha}) = A\cos^{2}(\phi - 60^{\circ}) + B\cos(\phi - 60^{\circ}) + C \qquad \text{(eq. 2.5)}$$

The coefficients A, B, and C are empirical parameters that are generally derived through correlation of homo- and hetero- nuclear NMR data with crystallographic information for large sets of proteins of known secondary structure. Several sets of Karplus coefficients have been reported in the literature and are listed in Table 2.1. In line with experiments by Graf et al. on alanine based peptides and Schwalbe et al., we choose to utilize the Karplus coefficients reported by Hu and Bax (A = 7.09, B = -1.42 , C = 1.55). These coefficients have proven the best in our global analysis of NMR and vibrational data.

Table 2.1: Karplus coefficients obtained from various studies.

	A	B	C
Vuister and Bax	6.51	-1.75	1.6
Hu and Bax	7.09	-1.42	1.55
Schmidt et al.	7.9	-1.05	0.65

[a]Data from Vuister and Bax taken from ref (163). [b]Data from Hu and Box taken from ref (164). [c]Data from Schmidt et al. taken from ref (162).

An example of the amide region in a typical H-NMR spectrum for a short peptide, trialanine, is shown in Figure 2.4. Amide proton signals generally show up between 6-9 ppm. Since trialanine has two labile NH groups which are exchangeable with solvent, we

observe two amide signals that are split into doublets due to interaction with neighboring CαH protons through the aforementioned spin-spin coupling mechanism.

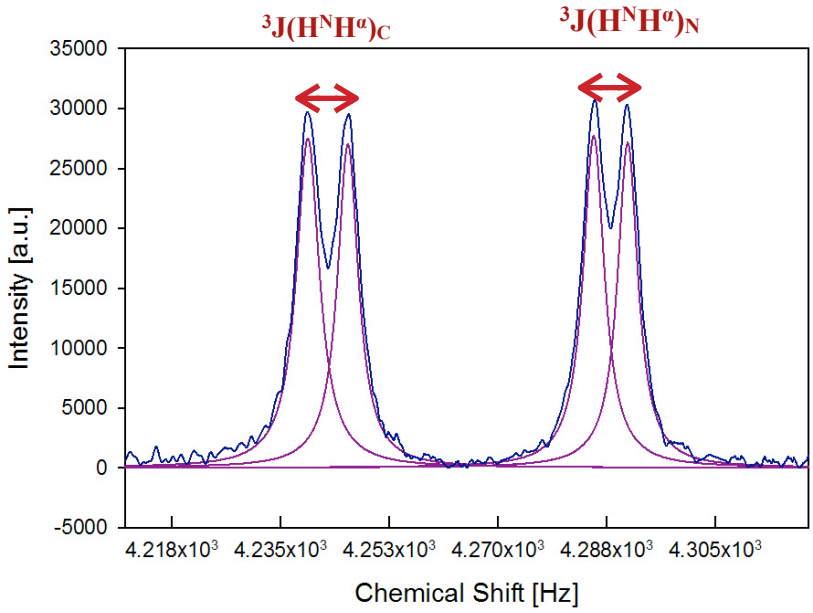

Figure 2.4: Experimentally obtained H-NMR spectrum for AAA in H_20 at 35 ^oC (blue line) along with the resulting fit (purple lines) using Voigtian profiles. The two $^3J(H^NH^\alpha)$ coupling constants are outlined with red arrows where we use the subscript "C" to denote the C terminal amide signal and "N" to denote the N-terminal (central residue) amide signal

The associated Karplus curve for $^3J(H^NH^\alpha)$ coupling constants is given below in Figure 2.5. The structural sensitivity of this parameter can be immediately seen when comparing the two predominate conformations in the unfolded state, namely β-strand and

pPII conformations. Canonical β-strand conformations have a ϕ angle of approximately -125^{o} , compared to pPII which has a canonical ϕ angle of -70^{o}. Therefore, the ^{3}J coupling constant in this region varies from 10-4.5Hz.

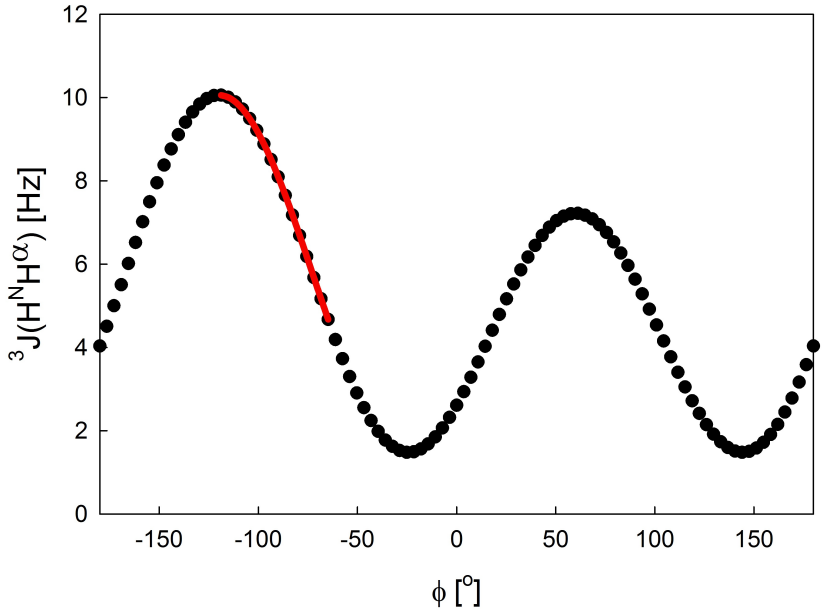

Figure 2.5: Graphical representation $^{3}J(H^{N}H^{\alpha})$ Karplus relationships that depends on the backbone dihedral angle ϕ. The red shaded area corresponds to the phi region between pPII and β.

2.2.1.2 Hetero-nuclear Coupling Constants

The information provided by the $^{3}J(H^{N}H^{\alpha})$ coupling constant is most beneficial when one has *a priori* information regarding the conformational ensemble of the target amino acid. For instance, as seen in Figure 2.5 the $^{3}J(H^{N}H^{\alpha})$ coupling constant has up to

three ϕ values that satisfy a given coupling magnitude. For this reason, in order to extract quantitative information on the conformation of an amino acid for which there is no *a priori* information, one should use a combination of several J-coupling constants. Figure 2.6 shows the six conformationally sensitive J-coupling constants as a function of their respective dependent backbone angle (ϕ or ψ). Each coupling constants depends differently on the backbone angle and can be obtained by appropriate hetero-nuclear 2D NMR experiments. Four of the coupling constants: $^{3}J(H^{\alpha},C')$, $^{3}J(H^{N},C')$, $^{3}J(H^{N},C^{\alpha})$, $J(H^{N},C^{\beta})$ depend solely on the angle ϕ, while the two remaining ones, $J(NC^{\alpha})$ and $^{2}J(NC^{\alpha})$ depend on ψ.

When using these constants to obtain information on short peptide or unfolded proteins, it becomes particularly important to consider the ensemble character of the conformational distribution (159). In a seminal study by Graf et al. (74) it was shown that the combined use of all NMR J-coupling constants along with distributions derived from constrained all atom MD simulations was sufficient to accurately obtain the conformational ensemble of various alanine based peptides. Work by the Schwalbe group has shown that various J coupling constants along with relaxation rates and SAXS data can be used to deduce residual structure due to non-local interactions in denatured proteins and IDPs (36, 45, 47, 48, 165, 166). Hagarman et al. recently showed that the use of these coupling constants in combination with vibrational spectroscopy was sufficient to quantify conformational ensembles of amino acid residues (54).

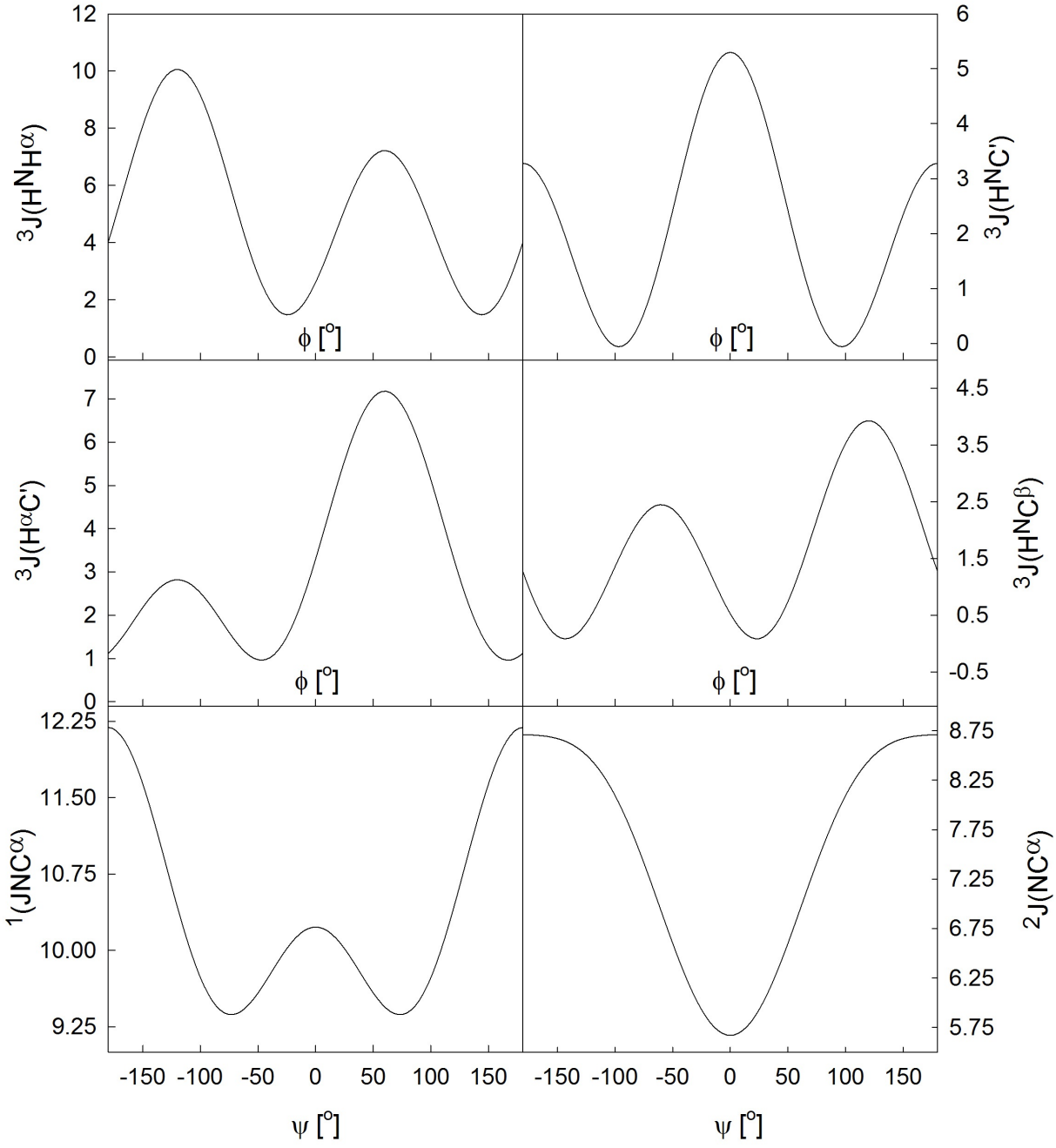

Figure 2.6: Graphical representation of six Karplus relationships (160), which depend on ϕ (upper and middle panels) and the angle ψ (lower panels).

2.3 VIBRATIONAL SPECTROSCOPIES

Proteins have vibrational spectra that contain important details on their structure. Bands are observed in infrared (IR) or Raman spectra that correspond to the normal modes of vibration of that particular structure. The following section will introduce amide I as the primary normal mode used herein for the conformational analysis of tri- and tetra-peptides, as well as the excitonic coupling model, which accounts for the vibrational mixing between different amide I mode in a polypeptide chains. The strength and sign of this mixing is orientationally dependent and thus enables spectroscopists to simulate the band profiles of amide I modes in infrared (IR), polarized Raman, and vibrational circular dichroism (VCD) spectra in terms of conformational distributions of amino acid residues. The mathematical formalism derived in the following section was employed via Matlab in order to simulate and fit experimental amide I band profiles and extract conformational distributions of amino acid residues in various GxG peptides, dipeptides, and GxyG peptides throughout *this thesis*.

2.3.1 Amide I/I' Mode

The amide I band is generally the most intense band in the IR spectra of proteins and peptides. According to normal mode decomposition, the amide I mode is primarily a CO stretching vibration together with an out-of-phase CN stretch component and small CCN deformation and NH in-plane bending (ipb) (167, 168). The transition dipole moment, therefore, has a direction that forms an angle between 15-25° with the CO bond direction in the plane of the peptide group (167) .

In water, the amide I mode mixes with HOH bending modes within the hydration shell, which complicates the use of this band for structural determination (76, 169, 170). However, this effect, as well as the admixture of NH in-plane-bending to amide I, can be eliminated by using heavy water (D_2O) as solvent. This ensures the replacement of NH by ND (which causes a slight downshift to the amide I frequency) and decouples it from any vibrations of molecules in the hydration shell (DOD bending modes are approximately 300 cm^{-1} lower than amide I). The amide I mode in D_2O is referred to as amide I' to reflect this difference. Depending on experimental conditions this band is located in the wavenumber region 1610-1700cm^{-1}. In the absence of vibrational coupling between amide I modes (i.e. the intrinsic amide I), the structural sensitivity of this mode arises from the conformationally dependent strength of hydrogen bonding between protein CO and NH group with solvent (168, 171-173). In addition, the frequency position and the oscillator strength of the amide I' mode is slightly side chain dependent, as shown by Measey et al.(174).

2.3.2 Excitonic Coupling

In polypeptides and proteins (i.e., more than one peptide group), the structural sensitivity of the amide I mode is primarily due to vibrational coupling between local amide I modes on adjacent carbonyl groups (167, 175). This coupling is primarily electrostatic in nature and can in most cases be approximated by the so-called "transition

dipole coupling" (TDC) formalism, which invokes the interaction potential between local amide I transition dipole moments. A transition dipole moment associated with a vibrational mode is in first order describable as the first derivative of the permanent electronic dipole moment with regard to the modes normal coordinate. The corresponding potential describing the TDC interaction between pairs of local amide I oscillators is given by:

$$V_{ij} = \frac{1}{\varepsilon}\left[\left|\left(\frac{\partial\vec{\mu}}{\partial Q_i}\right)Q_i\right|\left|\left(\frac{\partial\vec{\mu}}{\partial Q_j}\right)Q_j\right|\right]\left[\frac{\hat{n}_i\cdot\hat{n}_j - 3\left(\hat{n}_i\cdot\hat{R}_{ij}\right)\left(\hat{n}_j\cdot\hat{R}_{ij}\right)}{R_{ij}^3}\right] \qquad \text{(eq. 2.6)}$$

where $\partial\vec{\mu}/\partial Q_i$ is the derivative of the electronic ground state dipole moment (i.e., the transition dipole moment) with respect to the *i*th normal coordinate (Q_i). $\hat{n}_i$ are the unit vectors in the direction of the respective dipole moments, R_{ij} is the distance between the *i*th and *j*th dipole, and $\hat{R}_{ij}$ is the unit vector along this line.

Quantum mechanically, this coupling occurs between the excited vibrational states of the local amide I oscillators due to through bond and through space (TDC). As a consequence of this coupling, new delocalized eigenstates are generated, termed excitonic eigenstates. These excitonic eigenstates can be written as a linear combination of the excited vibrational states of local amide I' oscillators (176):

$$|\chi_i\rangle = \sum_i^n a_{ij}|\chi_j\rangle \qquad \text{(eq. 2.7)}$$

Where χ_i are the wavefunctions of the *i*th excitonic state, χ_j are the vibrational eigenfunctions of the local (unperturbed) amide I mode of the *j*th oscillator, and a_{ij} are the mixing coefficients (eigenvectors) which describe the amplitude of the *i*th excited state wavefunction at the *j*th position.

The coefficients a_{ij} can be obtained by solving the time-independent Schrödinger equation, which for a polypeptide of *n* oscillators reads:

$$\left(\hat{H}_0 + \hat{H}_{exc}\right)\left|\chi_i\right\rangle = E\left|\chi_i\right\rangle \qquad \text{(eq. 2.8)}$$

where $\hat{H}_0$ and $\hat{H}_{exc}$ represent the Hamiltonians for the unperturbed and coupled excitonic system respectively, E is the set of *n* eigenenergies, and $\left|\chi_i\right\rangle$ is a state vector representing the *n* excitonic states. The excitonic Hamiltonian contains information on the strength of vibrational coupling of the system, and hence is a function of the dihedral angles and distances between coupled oscillators. For a tetrapeptide, the total Hamiltonian ($H_T = H_0 + H_{exc}$) can be written in matrix notation as (177):

$$H_T = \begin{bmatrix} v_1 & \Delta_{12} & \delta_{13} \\ \Delta_{21} & v_2 & \Delta_{23} \\ \delta_{31} & \Delta_{32} & v_3 \end{bmatrix} \qquad \text{(eq. 2.9)}$$

where the diagonal elements, v_j are the unperturbed local amide I frequencies, $\Delta_{j,j\pm1}$ are the nearest-neighbor excitonic coupling energies, and $\delta_{j,j\pm2}$ are non-nearest neighbor coupling.

In order to obtain the excitonic eigenenergies through application of the above Hamiltonian, one requires information on the strength of vibrational coupling. In a seminal study by Torii and Tasumi, the nearest and second-nearest-neighbor coupling constants as a function of dihedral angles of a glycine dipeptide were determined using *ab initio* based normal mode calculations (176). The authors found that while the second nearest-neighbor coupling constants could be well approximated by TDC, the nearest neighbor coupling constants required additional factors. This likely reflects both the increased importance of through-bond interactions and a breakdown of the TDC approximation at short distances. The coupling map as a function of dihedral angles for nearest-neighbor coupling contains a negative trough in the center of the Ramachandran plot as well as two maxima centered at (0°, 180°) and (180°, 0°). DFT calculations performed by Gorbunov et al. (178) confirmed the conformational dependence of coupling strength obtained by Torii and Tasumi. In addition, Gorbunov et al. found that the vibrational coupling was transferable to dipeptides with hydrophobic and hydrophilic side chains as well as tripeptides with charged end groups (178).

In a recent study, Schweitzer-Stenner was able to heuristically reproduce the coupling map obtained by Torii and Tasumi in order to utilize the *ab initio* information in calculating amide I band profiles of conformational distributions (90). The coupling constant strength as a function of dihedral angels produced by Schweitzer-Stenner is displayed in Figure 2.7. This plot very closely mimics those obtained by Torrii and Tasumi, as well by Gorbunov et al. (178) Most importantly for use in calculating band

profiles, Schweitzer-Stenner was able to quantitatively reproduce the maximum at (0°, 180°) that causes a large gradient in the coupling strength along the pPII region (90). Implementation of this nearest-neighbor coupling plot allows for diagonolization of the total Hamiltonian (eq 2.9) to yield new eigenenergies v'_1, v'_2, v'_3, and mixing coefficients a_{ij} for each excitonic state.

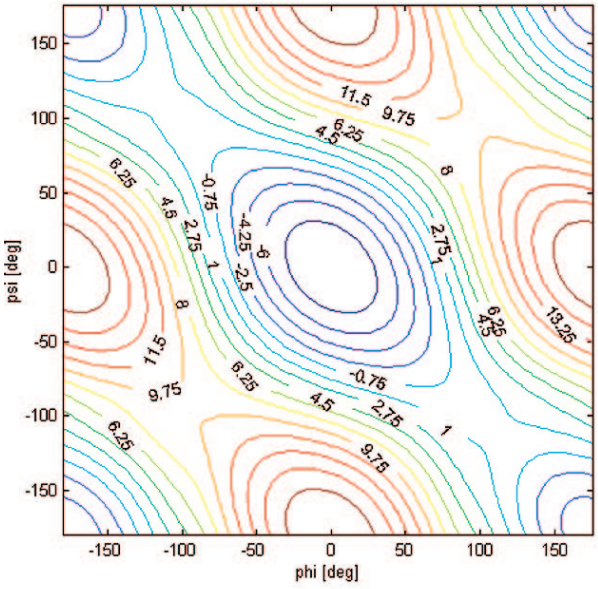

Figure 2.7. Contour plot of the nearest-neighbor coupling constant as function of ϕ and ψ calculated with a heuristic model by Schweitzer-Stenner (90) in reproduction of *ab initio* results of Torii and Tasumi (176).

2.3.3 Calculation of Band Profiles

In *this thesis*, the calculation of IR, isotropic Raman, anisotropic Raman and VCD amide I' band profiles for short peptides is conducted in the framework of the above described excitonic coupling model. In this section we first describe calculation specifics for Raman band profiles, followed by IR and VCD. The mathematical formalism derived here was implemented via Matlab for simulations and fitting of amide I' band profiles in order to obtain conformational distributions for amino acid residues throughout *this thesis.*

2.3.3.1 Calculation of Raman Profiles

Excitonic coupling causes a re-distribution of intensities between sub-bands underlying the total amide I band profile which is different in respective isotropic, anisotropic Raman and IR band profiles (177). Utilizing the excitonic coupling model along with the coupling strengths, it is therefore possible to describe this effect in terms of backbone dihedral angles. Diagonalization of the total Hamiltonian (eq 2.8) yields the mixing parameters a_{ij} which can be used to calculate corresponding Raman tensors (and dipole moments) of the *ith* excitonic states. Similar to equation (2.7), we can write the Raman tensor as a linear combination of the local amide I tensors and dipole moments, respectively, as follows:

$$\left|\hat{\alpha}'_i\right\rangle = \sum_i^n a_{ij}\left|\hat{\alpha}_j\right\rangle \qquad \text{(eq. 2.10)}$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ are the amide I Raman tensors of the three unperturbed (local) oscillators. The Raman tensor of an amide I mode in the above coordinate system can be written as:

$$a_j = \begin{pmatrix} a & b & 0 \\ b & 1 & 0 \\ 0 & 0 & c \end{pmatrix} \qquad \text{(eq. 2.11)}$$

The elements a b and c are here expressed in units α_{yy} since this is the dominant contribution to the tensor and we are only interested in relative units. The tensor elements of the local amide I' oscillators were recently determined by Measey et al. a using AX and XA dipeptides where X signifies one of the 20 naturally occurring amino acid residues (52). Transformation into a coordinate system coinciding with the "major axis of the Raman tensor" (mart) yields (83):

$$b = \frac{1-a}{9.3} \qquad \text{(eq. 2.12)}$$

In order to use eq. 2.10-2.12 to calculate the tensors of the excitonic states, the tensors of the local amide I modes have to be transformed into a common coordinate system. Choosing the C-terminal peptide group as the reference coordinate system and the corresponding NCα bond as the x axis, each tensor can be sequentially rotated from its initial coordinate system into the C-terminal coordinate system via appropriate

rotational matrix operations (83). Figure 2.8 shows the tetrapeptide GSAG (for an example) with assigned coordinate systems S1-3 used to express the local Raman tensors and dipole moments for reference.

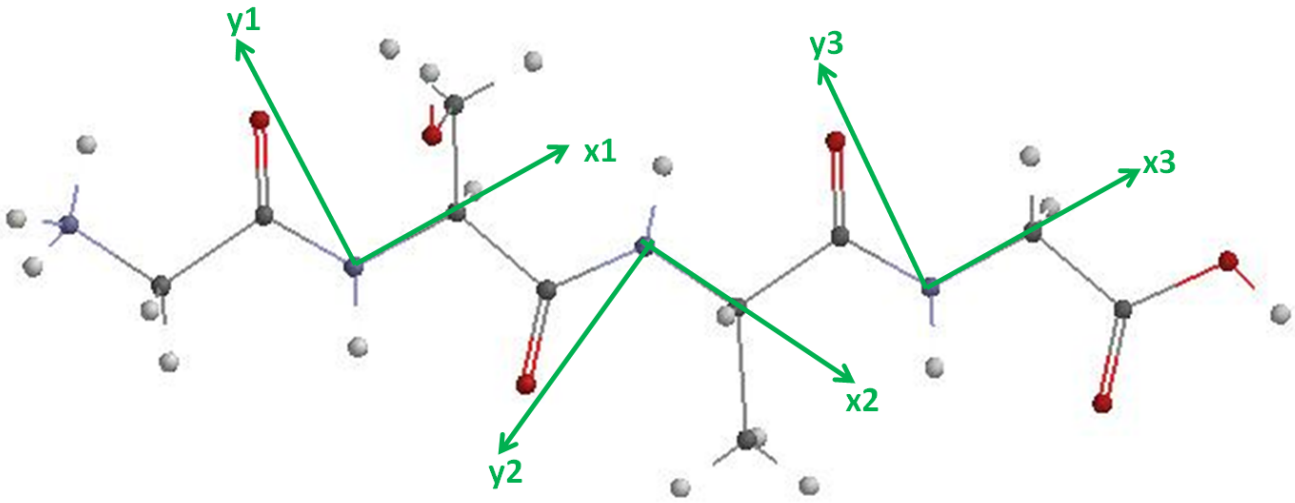

Figure 2.8: Planar structure of the tetra-peptides GSAG with coordinates fixed to completely extended (ϕ,ψ =180, 180) for demonstration purposes. The coordinate system S1(x1, y1, z1) was used to express the Raman tensors of the individual uncoupled amide I mode and their transition dipole moments. The z- component has been omitted for clarity. The structure was obtained by using the program TITAN from Wavefunctions.

For instance, the $\hat{\alpha}_1$ tensor, which belongs to the N-terminal peptide group was rotated from coordinate system S_1 into the reference coordinate system S_3 (C-terminal) via standard rotational matrices dependent on peptide geometry:

$$\hat{\alpha}_1(S_2) = R^T(\omega_{12})R^T(\psi_{12})R^T(\xi_{12})\left(R^T(\phi'_{12})\hat{\alpha}_1(S_1)R^T(\phi'_{12})\right)R(\xi_{12})R(\psi_{12})R(\omega_{12}) \qquad \text{(eq 2.13)}$$

Equation 2.12 can be understood using Figure 2.8 as follows. First the coordinate system S_1 has to be rotated by $\phi'_{12} = \phi_{12} - \pi$. Next, rotation about the angle ξ, which is the angle between NCα and CαC bond, brings the y axis colinear with the CαC bond. Next the system is rotated about the dihedral angle ψ_{12}. Subsequent rotation about an angle ω, which is the angle formed by the CαC bond of S_2 and the y_2 axis, brings the x axis of the coordinate system coplanar with the NCα bond of S_2. The angles ω and ξ are known from peptide geometry to be approximately 96° and 20°, respectively. This series of rotations brings S_1 into S_2. A similar set of transformations then brings the system into the final reference coordinate system S_1 at the C-terminal:

$$\hat{\alpha}_1(S_3) = R^T(\omega_{23})R^T(\psi_{23})R^T(\xi_{23})\left(R^T(\phi'_{23})\hat{\alpha}_1(S_2)R^T(\phi'_{23})\right)R(\xi_{23})R(\psi_{23})R(\omega_{23}) \qquad \text{(eq. 2.14)}$$

Now, with the local tensors in a common coordinate system, the isotropic and anisotropic scattering for the ith excitonic state can be calculated as follows (170):

$$\beta'^2_{s,i} = \frac{1}{9}\left(Tr\,\hat{\alpha}'_i\right)^2 \qquad \text{(eq. 2.15)}$$

$$\gamma'^2_{aniso,i} = \frac{1}{2}\left[\left(\hat{\alpha}'_{xx,i} - \hat{\alpha}'_{yy,i}\right)^2 + \left(\hat{\alpha}'_{yy,i} - \hat{\alpha}'_{zz,i}\right)^2 + \left(\hat{\alpha}'_{zz,i} - \hat{\alpha}'_{xx,i}\right)^2\right] + \frac{3}{4}\left[\left(\hat{\alpha}'_{xy,i} - \hat{\alpha}'_{yx,i}\right)^2 + \left(\hat{\alpha}'_{yz,i} - \hat{\alpha}'_{zy,i}\right)^2 + \left(\hat{\alpha}'_{zx,i} - \hat{\alpha}'_{xz,i}\right)^2\right]$$

(eq. 2.16)

Once these tensor invariants are obtained, the Raman band profiles can be calculated as a superposition of Gaussian profiles representing the respective excitonic modes (83, 170):

$$I_{iso}(\Omega) = \sum_{i=1}^{n-1}\left[45\beta_{s,i}^2 f_i(\Omega)\right] \quad \text{(eq. 2.17)}$$

$$I_{aniso}(\Omega) = \sum_{i=1}^{n-1}\left[7\gamma_{s,i}^2 f_i(\Omega)\right] \quad \text{(eq. 2.18)}$$

where

$$f_i = \frac{1}{\sigma_i\sqrt{2\pi}} e^{\frac{-(\Omega-\Omega_i)^2}{2\sigma_i^2}} \quad \text{(eq. 2.19)}$$

Ω_i is the eigenenergy of the *ith* excitonic state in cm^{-1} and σ_i is the corresponding halfwidth of the band profile.

In our analyses we exploit the additional structural sensitivity offered by measuring polarized Raman (83). By measuring the light polarized parallel and perpendicular to the scattering plane, one can obtain the isotropic and anisotropic intensity profiles via:

$$I_{iso} = I_x - \frac{4}{3} I_y \quad \text{(eq. 2.20)}$$

$$I_{aniso} = I_y \quad \text{(eq. 2.21)}$$

2.3.3.2 Calculation of IR profiles

The band profile for IR can be similarly determined. IR absorption intensity is proportional to the square of the transition dipole moment associated with the respective vibrational transition. The dipole moments of the excitonic states can be written generally as:

$$\left|\bar{\mu}'_i\right\rangle = \sum_i^n a_{ij}\left|\bar{\mu}_j\right\rangle \qquad \text{(eq. 2.22)}$$

Each transition dipole moment can be described by:

$$\bar{\mu}_i = \begin{pmatrix} \mu_{0i}\cdot\cos\vartheta \\ \mu_{0i}\cdot\sin\vartheta \\ 0 \end{pmatrix} \qquad \text{(eq. 2.23)}$$

where μ_{0i} is the local amide I transition dipole moment of the *ith* peptide group and ϑ is the angle between the dipole vector and the x-axis of the coordinate system.

To utilize eq. (2.22-2.23) for determination of transition dipole moments in excitonic states, the local dipole moments along the peptide backbone must first be transferred into a common coordinate system. Similar to the procedure described above for transformation of Raman tensors, we choose the C-terminal as our basis coordinate system. The local coordinate systems are the same as those shown in Figure 2.8 in the example of tetrapeptide GSAG. The respective rotational transformations to move the

local transition dipole moments of residue 1 and 2 in to basis coordinate system S_3 are as follows (83):

$$\bar{\mu}_1(S_3) = R(\omega_{23})R(\psi_{23})R(\xi_{23})R(\phi'_{23})R(\omega_{12})R(\psi_{12})R(\xi_{12})R(\phi'_{12})\bar{\mu}_1(S_1) \qquad \text{(eq. 2.24)}$$

$$\bar{\mu}_2(S_3) = R(\omega_{23})R(\psi_{23})R(\xi_{23})R(\phi'_{23})\bar{\mu}_2(S_2) \qquad \text{(eq. 2.25)}$$

Once the dipole moments are in the basis system S_3, the dipole moments of the excitonic states can be calculated via eq 2.22. The IR intensity can then be written as a superposition of Gaussian bands representing the respective bands of the excitonic modes:

$$I_{IR}(\Omega) = \sum_{i=1}^{n-1}\left[I_{s,i}^{IR} f_i\right] \qquad \text{(eq. 2.26)}$$

where

$$I_i^{IR} = C\tilde{\nu}_0\left(\bar{\mu}_i \cdot \bar{\mu}_i\right) \qquad \text{(eq. 2.27)}$$

where C is a constant, which in electrostatic units is 9.2×10^{-39} esu.

2.3.3.3 VCD Profiles

Vibrational circular dichroism (VCD) is a powerful tool for structural determination in short peptides and proteins. VCD intensity is proportional to the rotational strength of the *ith* excitonic state (179). Rotational strength was previously defined in the context of UVCD, however in the context of VCD we need to re-write it in terms of the derivative of electric and magnetic dipole moments:

$$R_i = \mathrm{Im}\left[\vec{\mu}_i' \bullet \vec{m}_i'\right] \qquad \text{(eq. 2.28)}$$

As one can see from eq 2.28, rotational strength depends on the scalar product between the transition dipole and transition magnetic moments of the respective excitonic state. Holzwarth and Chabay used a coupled oscillator model to derive the rotational strength for isoenergetic oscillators (179). Eker at al. expanded this model to consider two non-isoenergetic oscillators (70). We follow Eker et al. and write the rotational strength for a peptide with N residues as:

$$R = \mathrm{Im}\left[\sum_{j=1}^{N}\alpha_{ij}\vec{\mu}_j^{'}\sum_{j=1}^{N}\alpha_{ij}\vec{m}_j^{'} - \frac{i\pi}{2}\left(\sum_{l=1}^{N-1}\sum_{m=2}^{N}v_{lm}T_{lm}\times\left(\alpha_{il}\vec{\mu}_l' - \alpha_{im}\vec{\mu}_m'\right)\right)\right] \qquad \text{(eq. 2.29)}$$

where $\vec{\mu}_j'$ and $\vec{m}_j'$ are the derivatives of electric and magnetic dipole moments of the local amide I' oscillators. $\vec{T}_{lm}$ is the distance vector between oscillator l and m. The

coefficients a_{ij} are the afore-described mixing coefficients, which reflect the contribution of the *jth* residue to the *ith* excitonic state. The first term describes the rotational strength induced by excitonic coupling of excited vibrational states. The second term describes the chirality generated by the magnetic moment at residue i induced by the electronic transition dipole moments of neighbors. The VCD profile of the amide I can be calculated as a superposition of Gaussian bands using:

$$\Delta\varepsilon(\Omega) = \frac{\tilde{\nu}_0}{2.3 * 10^{-38}} \sum_{i=1}^{n-1} \left[R_i f_i \right] \qquad \text{(eq. 2.30)}$$

where v_0 is the first moment of the amide I' band profile and f_i is the afore-defined Gaussian band profile (eq 2.19).

2.3.3.4 Calculation of Distribution Functions

For the simulation of all amide I' profiles and *J* coupling constants of short peptides, we generally assumed a statistical conformational ensemble for the corresponding investigated amino acid residue consisting of normalized two dimensional Gaussian distributions. The central coordinates and halfwidths of these distributions are used as variable parameters for our simulations. This type of ensemble takes into account that each residue has multiple co-existing conformations that make up the conformational ensemble and that these conformations should be approximated by distribution functions

that mimic the breadth of the sub-space. The total ensemble distribution function is given by (90):

$$f(\phi,\psi) = \sum_j \frac{\chi_j}{2\pi\sqrt{\left|\hat{V}_j\right|}} e^{-0.5\left(\vec{p}-\vec{p}_0^0\right)^T \hat{V}_j^{-1}\left(\vec{p}-\vec{p}_0^0\right)} \qquad \text{(eq. 2.31)}$$

where

$$\vec{p} = \begin{pmatrix} \phi \\ \psi \end{pmatrix} \qquad \text{(eq. 2.32)}$$

and

$$\hat{V}_j = \begin{pmatrix} \sigma_{\phi,j} & \sigma_{\phi\psi,j} \\ \sigma_{\phi\psi,j} & \sigma_{\psi,j} \end{pmatrix} \qquad \text{(eq. 2.33)}$$

is the covariance matrix that contains the halfwidths along ϕ and ψ as diagonal elements. The diagonal elements of the matrix V_j are the half half-widths of the *j*th distribution along the coordinates ϕ and ψ, and the corresponding off-diagonal elements reflect correlations between variations along the two coordinates. The factor χ_j is the mole fraction of the *jth* sub-distribution.

If multiple conformations are considered, each then gives rise to its own band profile, which must be weighted accordingly (90). The expectation value of any observable, x, that depends on the dihedral angles of the conformational ensemble (i.e. IR, Raman intensities, rotational strengths and J coupling constants) can then be written as:

$$\langle x \rangle = \frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} x f(\phi,\psi) d\phi d\psi}{Z} \qquad \text{(eq. 2.34)}$$

where Z is the partition function of the ensemble that can be calculated by:

$$Z = \sum_{j=1} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} x f_j(\phi,\psi) d\phi d\psi \qquad \text{(eq. 2.35)}$$

2.3.4 Conformationally Dependent Band Profile Differences

Excitonic coupling between the local amide I' modes increases the frequency splitting between them and re-distributes IR and Raman intensities (83). If only two coupled oscillators are considered, two different excitonic energy levels are observed depending on whether the oscillators are in phase or out of phase. The result is a splitting of the amide I band, which can be as large as ~70 cm^{-1} for β-sheet structures (167). Figure 2.9 shows the band splitting for a tetrapeptide, namely tetraalanine, for example (83). Schweitzer-Stenner et al. (83) showed that the band profile can be decomposed into three

sub-bands (1643 cm^{-1} (AI'_1), 1656 cm^{-1} (AI'_2), and 1674 cm^{-1} (AI'_3)) that are a consequence of the three delocalized excitonic states formed from coupling of the three local amide I modes in tetraalanine. As seen in Figure 2.9 the sub-bands noticeably differ in terms of there respective depolarization ratios and IR intensity ratios. This leads to the observed "non-coincidence" between peak positions with maximum intensity for the total amide I' band profiles, which indicates a substantial strength of excitonic coupling. The strongly negative-positive couplet in the VCD spectrum is indicative of an extended pPII structure.

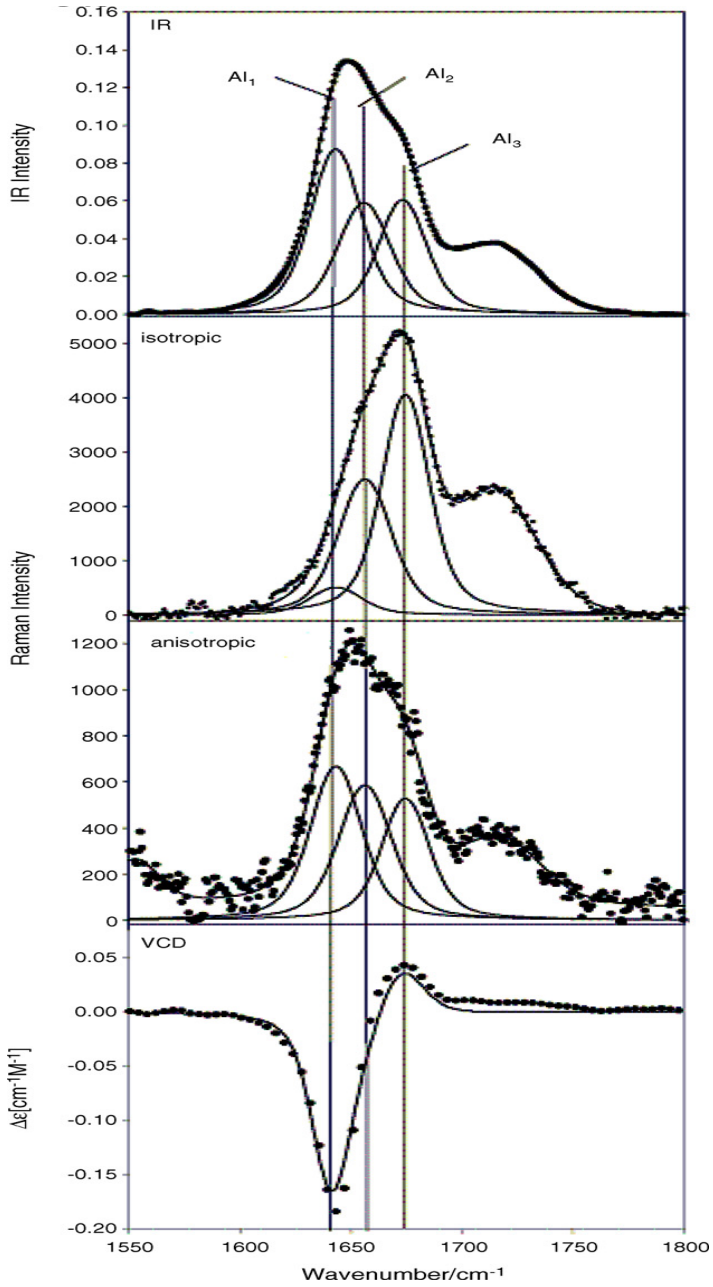

Figure 2.9. Amide I′ region of the IR, isotropic Raman, anisotropic Raman, and VCD spectra of tetraalanine in D_2O measured at pD 1. The line profiles in the spectra result from a global spectral decomposition based on excitonic coupling. Taken from ref (83).

To illustrate how the amide I' band profile reflects various conformational distributions, Figure 2.10 depicts the band profiles calculated for (a) 100% right-handed helical, (b) 100% β-strand, (c) 100% pPII, and (d) a mixture of 50% pPII and 50% β-strand. The IR and Raman spectra of both extended pPII and β-strand conformations exhibit a large non-coincidence. This is due to the large strength of excitonic coupling in this region of the Ramachandran plot as shown in Figure 2.7. The isotropic Raman and IR profiles for the pure pPII and β-strand conformations are qualitatively similar, however, these two conformations can be discerned when examining the anisotropic Raman profile, which, for pPII gives rise to a higher intensity at lower wavenumbers. β-strand conformations give rise to a comparatively more symmetric band profile in the anisotropic spectrum. In addition, the pPII conformation gives rise to a particularly large negative-positive couplet in the VCD amide I profile which is clearly discernible for β-strand conformations. In contrast, the IR and Raman spectra of right handed helical conformations do not exhibit a marked non-coincidence, and additionally show a very weak positive-negative couplet in the VCD. These differences illustrate that the combined use of IR, polarized Raman, and VCD comprises a powerful tool for investigating secondary structure content in the unfolded state. In particular, VCD has proven very useful in this regard as the couplet pattern and strength of the amide I couplet is very sensitive to backbone dihedral angles. As seen in Figure 2.10, one can clearly discern between mixtures of even extended conformations (90). In addition, Schweitzer-Stenner et al. has recently shown that VCD signal can even be used to discriminate

conformational content for longer disordered peptides, as well as parallel and anti parallel β-sheets (67, 180).

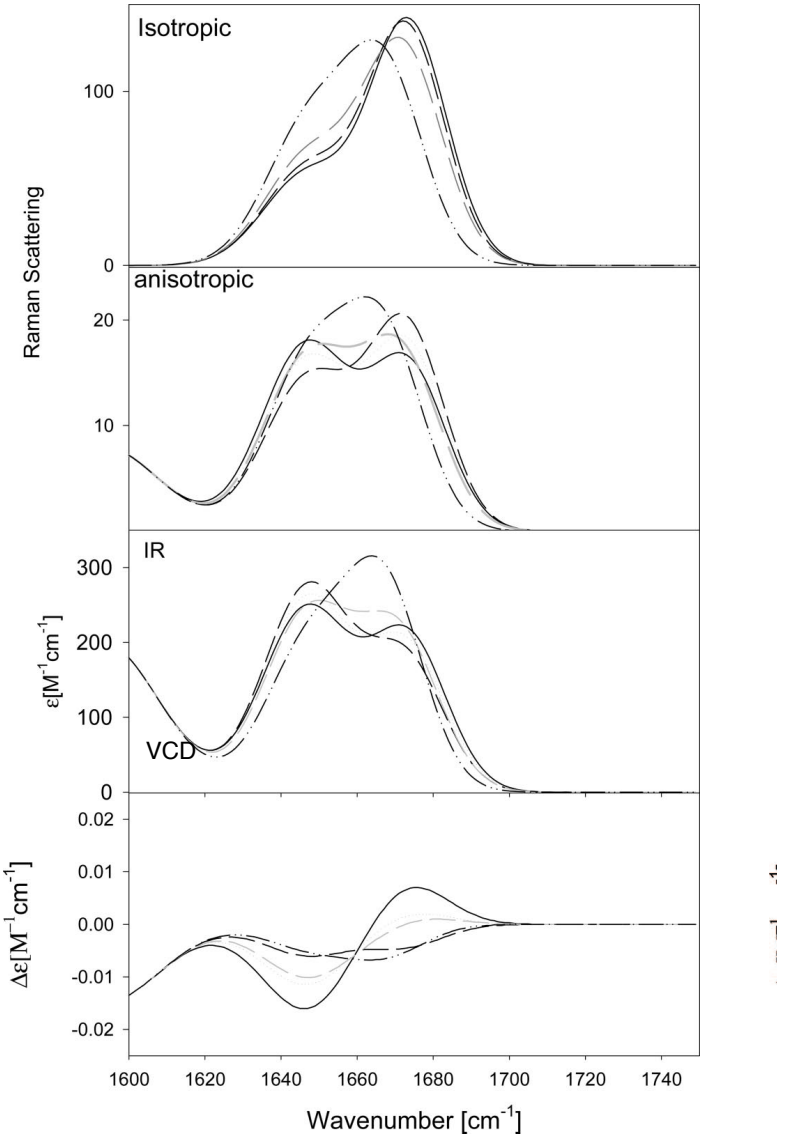

Figure 2.10. Isotropic Raman, anisotropic Raman, IR, and VCD amide I profiles of tripeptides simulated for different conformational ensembles, that is, 100% PPII (solid line), 100% β-strand (dashed line), 50:50 mixture of pPII and β-strand (dashed gray), and 100% right- handed helical (dashed-dot-dot). Taken from ref (90).

CHAPTER 3. MATERIALS AND METHODS

3.1 MATERIALS

Short alanine based peptides, L-alanyl-L-alanyl-L-alanyl (AAA) and the alanine dipeptide (H_3−CONH−$CHCH_3$−CONH−CH) were purchased from Bachem Bioscience with >98% purity. These peptides were used without further purification. Glycine capped peptides ("GxG") used for temperature dependent studies: L-glycyl-L-alanyl-L-glycine (GAG), L-glycyl-L-glutamic acid-L-glycine (GEG), L-glycyl-L-lysyl- L-glycine (GKG), L-glycyl-L-phenylalanyl-L-glycine (GFG), L-glycyl-L-methionyl-L-glycine (GMG), glycyl-L-tyrosyl-L-glycine (GYG), and L-glycyl-L-leucyl-L-glycine (GLG) were purchased from Bachem Biosciences Inc. (King of Prussia, PA) and used without further purification. L-glycyl-L-valyl-L-glycine (GVG), L-glycyl-L-aspartic acid-L-glycine (GDG), L-glycyl-L-seryl-L-glycine (GSG), L-glycyl-L-Isoleucyl-L-glycine (GIG), L-glycyl-L-threonyl-L- glycine (GTG), glycyl-L-asparagyl-L-glycine (GNG), and L-glycyl-L-cystyl-L-glycine (GCG), L-glycyl-L-aspartic acid-L-glycine (GDG), L-glycyl-L-threonyl-L-glycine (GTG) were custom synthesized by Genscript Corp. (Piscataway, NJ) at >98% purity. These peptides were subsequently purified by dialysis against aqueous HCl in 100MWCO Spectra/Por CE Float-A-Lyzer bags (Spectrum Laboratories Inc., Ranch Dominquez, CA) and lyophilized to remove trace amounts of trifluoroacetic acid (TFA).

Isotopically labelled tetrapeptides ("GxyG") for heteronuclear NMR experiments: L-glycyl-L-aspartic acid-L-alanyl-L-glycine (GDAG), L-glycyl-L-aspartic acid-L-lysyl-

L-glycine (GDKG), L-glycyl-L-aspartic acid-L-leucyl-L-glycine (GDLG), L-glycyl-L-aspartic acid-L-valyl-L-glycine (GDVG), L-glycyl-L-seryl-L-alanyl-L-glycine (GSAG), L-glycyl-L-seryl-L-lysyl-L-glycine (GSKG), L-glycyl-L-seryl-L-leucyl-L-glycine (GSLG), L-glycyl-L-seryl-L-valyl-L-glycine (GSVG), L-glycyl-L-alanyl-L-valyl-L-glycine (GAVG), L-glycyl-L-lysyl-L-valyl-L-glycine (GKVG), L-glycyl-L-leucyl-L-valyl-L-glycine (GLVG), L-glycyl-L-valyl-L-valyl-L-glycine (GVVG), L-glycyl-L-alanyl-L-leucyl-L-glycine (GALG), L-glycyl-L-lysyl-L-leucyl-L-glycine (GKLG), L-glycyl-L-leucyl-L-leucyl-L-glycine (GLLG), L-glycyl-L-valyl-L-leucyl-L-glycine (GVLG) were ^{13}C carbonyl labelled at residue 1, uniformly ^{13}C and ^{15}N labelled at residue 2 and 3 and ^{15}N labelled at residue 4. The C-terminal residue was manually attached to a chlorotrityl resin. The synthesis was manually carried out by standard Fmoc chemistry. Peptides were purified by reversed- phase HPLC. Products were characterized using electrospray ionization mass spectrometry and analytical HPLC. All isotopically labelled Fmoc-protected amino acids were purchased from Cambridge Isotope Laboratories (Andover, MA).

3.2 METHODS

3.2.1 UVCD Measurements

Temperature-dependent UVCD spectra for all peptides were obtained using a Jasco J-810 spectropolarimeter in the wavelength range of 180-300 nm. Peptide samples were placed in a 100 μm ICL cell (International Crystal Laboratories) and into a nitrogen-purged system. Each sample was allowed to equilibrate to the desired temperature with a Peltier heating system (model PTC-423S, accuracy 1°C). Spectra were generally recorded from 10°C to 90°C in increments of 5°C with a 100s delay time and ten accumulations at each temperature. The spectra were collected as ellipticity as a function of wavelength and converted to $\Delta\varepsilon$ [M^{-1} cm^{-1} res^{-1}] via standard calculations. For all UV-CD measurements, peptides were dissolved in the desired solvent at a concentration of 5mM peptide. For studies in pure aqueous environment H_2O was used as the solvent. The omission of the 10% D_2O solvent, which was used for NMR experiments, has a negligible effect (~2.9%) on the UVCD signal magnitude and thermodynamic parameters derived from our experiments based on a comparison of UVCD spectra of AAA in 100% H_2O and D_2O reported by Eker et al. (70). For co-solvent studies, the peptide was dissolved in the desired binary mixture of % H_2O / % alcohol (v/v). The alcohol solvents chosen were either 5-60% ethanol or 5-60% glycerol. The pH for all samples was adjusted to approximately 2.0 with HCl. The choice of pH is critical to ensure on one side that nearly all peptides are in the cationic state, which requires a pH below 2.5, and on the other side to avoid a high Cl^- concentration (pH > 1.5), which would affect UVCD

absorbance below 200 nm. We have to confine ourselves to the cationic state to avoid the contributions from n->π* and π->π* charge transfer transitions from the carboxylate to the peptide group.

3.2.2 NMR Measurements

For all aqueous NMR measurements (homo- and hetero-NMR), the desired peptide was dissolved in a solution of 90% H_2O/10% D_2O at a final peptide concentration of 0.05 M. The D_2O (Sigma-Aldrich) contained 0.1% trimethylsilane (TMS), which was used as an internal standard. For co-solvent studies, the same general procedure was used, however, adding the desired volume % co-solvent to create a stock binary mixture, which was then used to dissolve the peptide. The pH for all samples was adjusted to approximately 2.0 with HCl. The choice of this pH is necessary to ensure that the peptide is in the cationic state in order to match UVCD experiments. In addition, using acidic conditions also facilitates the determination of $^3J(H^NH^\alpha)$ of the labile amide protons.

3.2.2.1 ^{1}H-NMR Temperature Dependent Studies:

The ^{1}H-NMR spectra as a function of temperature for all GxyG, GxG, as well as short alanine based peptides, were recorded on a Varian 500 MHz FT-NMR with a 5 mm HCN triple resonance probe (Drexel University). Varian's VNMR software (v. 6.1) was used to process all spectra, and the presaturation mode (presat) was used to suppress solvent signals. The sample was allowed to equilibrate for 100s at each temperature, and

4-32 transients were collected for each, depending on the temperature (with the highest temperatures generally requiring 32 scans). The raw FIDs were analyzed using Mestrec software, which was used to Fourier transform and phase the data. To obtain accurate values for $^{3}J(H^{N}H^{\alpha})$ coupling constants at all temperatures and solvation conditions, the following procedure was carried out. Amide proton signals were decomposed using Voigtian profiles with flexible half-widths. An example of a typical H-NMR spectrum in the amide proton region of trialanine is shown below for reference. The frequency positions of the Voigtian profiles (in Hz) were then plotted as a function of temperature as shown in Figure 3.1, and a linear regression was used to fit the data. The linear equations generated from these fits were then subtracted in order to generate consistent $^{3}J(H^{N}H^{\alpha})$ coupling constants as a function of temperature for each peptide.

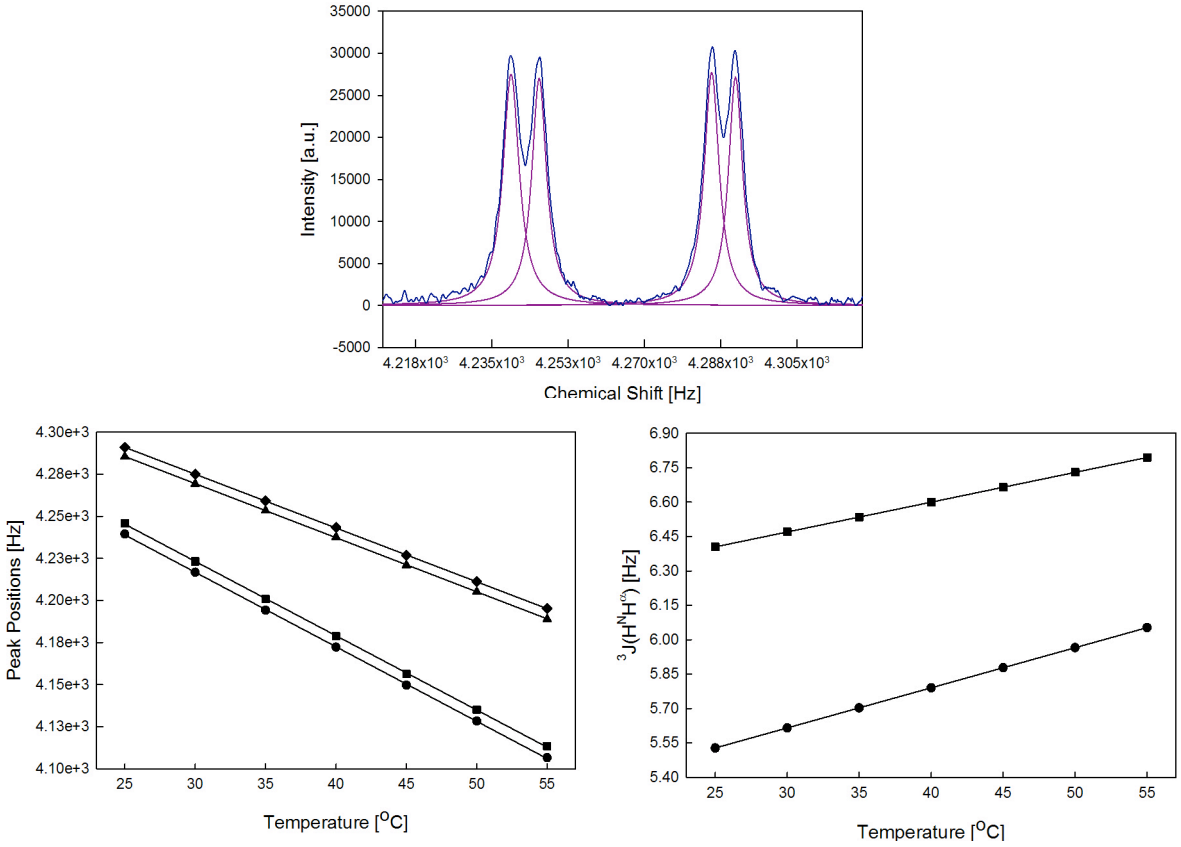

Figure 3.1: (Top) H-NMR spectrum of amide protons obtained for AAA at 35°C. The pink lines represent the optimized Voigtian sub-bands used to fit the doublets. (Bottom, Left) Peak positions [Hz] of the four component peaks of the two amide doublets as a function of temperature. (Bottom, Right) Resulting coupling constants as a function of temperature obtained by procedure described in text.

3.2.2.2 Homo- and Hetero-Nuclear NMR:

Amide proton $^{3}J(H^{N}H^{\alpha})$ coupling constants were determined using ^{1}H NMR. The spectra for GxyG peptides were recorded on a Bruker 600 MHz DRXmeter with a 5 mm HCN triple resonance probe with xyz gradients (Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe University, Schwalbe Research Group). All measurements were performed at 298 K. Spectra were acquired and analyzed using the program TopSpin V 2.1. Proton assignment for all GxyG peptides could be obtained from 1D spectra and ^{1}H, ^{1}H-TOCSY spectra using a DIPSI-2* mixing sequence (181). $^{3}J(H^{N}H^{\alpha})$-coupling constants were obtained from a ^{13}C decoupled H-NMR spectrum using standard presaturation to suppress water solvent. The exact coupling constants were determined by fitting the proton signals with Voigtian profiles with flexible halfwidths. The remaining coupling constants used to garner information on the angle ϕ were determined through the appropriate E.COSY-type measurements: $^{3}J(H^{\alpha},C')$ was determined through (CO- coupled (H)NCAHA experiment (182); $^{3}J(H^{N},C')$ was determined through soft HNCA-COSY(165, 183); the $^{3}J(H^{N},C^{\alpha})$ coupling constant was determined through HNCO(C^{α})-E.COSY (165, 184); and the $^{3}J(H^{N},C^{\beta})$ coupling constant was determined through HNHB(C^{β})-E.COSY (183). Exact peak positions for calculation of coupling constants were extracted by deconvolution of the respective 1D trace. Information of the angle ψ is obtained through determination of $^{1}J(NC^{\alpha})$- and $^{2}J(NC^{\alpha})$-coupling constants. These coupling constants were determined through the J-modulated

$^{1}H,^{15}N$-HSQC method (166). In order to extract these coupling constants, the peak intensities for different mixing times were fitted to:

$$I_{\exp} = A\cos(\pi^{1}J\tau)\cos(\pi^{2}J\tau)e^{-\tau/T_2^*} \qquad \text{(eq. 3.1)}$$

where I_{exp} is the experimental peak height intensity, $^{1}J = {}^{1}J(NC^{\alpha})$, $^{2}J = {}^{2}J(NC^{\alpha-1})$, A is a fitted amplitude factor, and T^*_2 is the transverse relaxation rate for the $\tau/2 - \pi - \tau/2$ echo sequence. Parameters used for 2D experiments were implemented based off of previous peptide studies of the same kind, reported by Hagarman et al. (54).

3.2.3 Vibrational Spectroscopy

All FT-IR and VCD spectra were recorded with a Chiral IR™ Fourier Transform VCD spectrometer (BioTools, Jupiter, Fl). A CaF_2 demountable cell (BioTools) with a path length of 100 μm was used for all experiments. The spectral resolution was 8 cm^{-1} for all measurements and a total of 720 min acquisition time (648 min for VCD, 72 min for IR spectra). Backgrounds of each solvent (i.e., D_2O, H_2O, and all binary alcohol mixtures) were obtained separately and appropriately subtracted out of the sample FTIR spectra using the program Multifit.

All polarized Raman spectra were obtained with either a 514 nm (800 mW) laser line from Spectra Physics Argon-Krypton laser (Santa Clara, CA), or a 442 nm (32 mW) excitation from a HeCd laser (Model IK 4601R-E, Kimmon Electric US). The Raman

scattered light was collected in a backscattering geometry. Peptide samples (0.2M in appropriate solvent) were placed in a 1.0 mm Q Silica depression well slide with a thin glass cover slip. The x50 objective of the microscope was focused past the coverslip and into the sample. Spectra were measured at room temperature. The laser beam was directed into a RM 100 Renishaw confocal Raman microscope with a back thinned CCD camera, from Renishaw Inc. (Hoffman Estates, IL). Scattered light was filtered with an appropriate notch filter, dispersed by a single-grating 2400 l/mm grating, and polarized with a λ/2 plate. The light was dispersed by the spectrometer and then imaged onto a CCD (Wright Instruments). Both parallel x-polarized and antiparallel y-polarized (with respect to the polarization of the excitation laser beam) spectra were obtained with 10 accumulations each. All spectra were then averaged to sufficiently eliminate background noise. The corresponding solvent spectra were then subtracted from the sample spectra using the program Multifit (185).

CHAPTER 4. SPECIFIC THEORETICAL PROTOCOL DERIVED FOR THE THERMODYNAMIC ANALYSES OF SHORT PEPTIDES

4.1 TWO-STATE THERMODYNAMIC MODEL FOR UVCD ANALYSES

To obtain the enthalpic and entropic differences between pPII and β-strand, we employed a global fitting procedure to analyze the temperature dependence of the conformationally sensitive maximum dichroism $\Delta\varepsilon$ (T) and the $^3J(H^NH^\alpha)$(T) constants. In the case of alanine based peptides (AAA), we know from our own vibrational work and others that the conformational ensemble is dominated by pPII and β (56, 70, 74), and hence we assume that the small contribution of turns ≤ 8% is negligible in their analysis. Therefore, the fitting of these peptides is a "two-state" thermodynamic analysis. With regard to GxG peptides, the turn population may constitute up to 32% of the conformational ensemble, and so for these peptides we used a "pseudo two-state thermodynamic analysis," which includes the addition of temperature independent turns. The temperature independence of turn states is justified within GxG peptide background information provided in *Ch 1.3 of this thesis*.

In our thermodynamic analysis, the experimentally measured $^3J(H^NH^\alpha)$ and $\Delta\varepsilon$ values can be expressed in terms of mole-fraction weighted contributions from each conformation. It is important to note that UVCD spectra provide information on the net conformational populations of pPII and β-strand, whereas the $^3J(H^NH^\alpha)$ values obtained from ^{1}H-NMR provide site-specific information regarding the average ϕ-values of the central and C-terminal residue, according to the specific empirical Karplus relationship

(160). The analysis of CD data therefore was based on the assumption that the behavior of the entire peptide is describable by a two state model, for which each residue may exhibit either pPII or β-strand conformation. This assumption is justified by the clear isodichroic points we and others obtain for alanine based peptides as well as GxG model peptides (75, 186, 187). This is most likely due to the fact that both residues provide nearly identical contributions to the observed dichroic value if they adopt similar conformations, in line with recent theoretical results by Woody (64). The applicability of a two state model for explaining the CD spectra of trialanine has recently been confirmed by Oh et al. (186). This implies that peptides with mixed residue conformations (e.g. pPII for the central residue and β-strand for the C-terminal) are treated as a 50:50 mixture of pure pPII and β residues. In addition, in all that follows we assign the ground state as pPII. Therefore, we can express Δε generally as:

$$\Delta\varepsilon = \Delta\varepsilon_{pPII} \cdot \left(2\chi_{pPII-pPII} + \chi_{pPII-\beta} + \chi_{\beta-pPII}\right) + \Delta\varepsilon_{\beta} \cdot \left(2\chi_{\beta-\beta} + \chi_{\beta=pPII} + \chi_{pPII-\beta}\right) \qquad \text{(eq. 4.1)}$$

where χ_{i-j} (i,j = pPII, β) are the mole fractions of the four different net peptide conformations that can contribute to the UVCD signal for a tripeptide, and $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$ are the intrinsic dichroism values of a residue in pPII and β-strand, respectively, in units of $M^{-1}cm^{-1}$. The factor of 2 for $\chi_{pPII-pPII}$ and $\chi_{\beta-\beta}$ is necessary to account for the case where both residues adopt the same conformation and hence contribute twice to the

dichroic value. In this formalism we assume that there is no significant electronic coupling between residues for the considered conformations, in accordance with Woody (63). In addition, the N-terminal residue has been shown not to contribute to the far UVCD of peptides (64). Therefore, trialanine has two UVCD active residues, and the possible peptide level conformations are: $\chi_{pPII-pPII}$, $\chi_{pPII-\beta}$, $\chi_{\beta-pPII}$ and $\chi_{\beta-\beta}$.

However, for a dipeptide (e.g. the AdP), although there are two peptide bonds, there is only one residue with Δε values that contribute to the UVCD spectra. Similarly for GxG model peptides, since the glycine residues are achiral and therefore CD inactive, there is only one CD contributing residue. Therefore, in the case of AdP and GxG mixed mole fraction terms are set to zero and the equation can be simplified to:

$$\Delta\varepsilon = \Delta\varepsilon_{pPII} \cdot \chi_{pPII} + \Delta\varepsilon_{\beta} \cdot \chi_{\beta} \qquad \text{(eq. 4.2)}$$

The mole fractions in eq.4.2 can then be expressed as a function of temperature using Boltzmann factors. For the Δε (T) of a tripeptide like trialanine, this yields:

$$\Delta\varepsilon(T) = \frac{\left[\Delta\varepsilon_{pPII}\left(2 + e^{-\Delta G_1/RT} + e^{-\Delta G_2/RT}\right) + \Delta\varepsilon_{\beta}\left(e^{-\Delta G_1/RT} + e^{-\Delta G_2/RT} + e^{-(\Delta G_1+\Delta G_2)/RT}\right)\right]}{1 + e^{-\Delta G_1/RT} + e^{-\Delta G_2/RT} + e^{-(\Delta G_1+\Delta G_2)/RT}}$$

(eq. 4.3)

where $\Delta G_i = G_{\beta,i} - G_{pPII,i}$ denotes the Gibbs energy difference between pPII and the β-strand conformations of the i^{th} CD-contributing residue, with i=1 denoting the central and i=2 the C-terminal residue. For AdP and GxG peptides, eq. 4.3 can be reduced to:

$$\Delta\varepsilon(T) = \frac{\Delta\varepsilon_{pPII} + \Delta\varepsilon e^{-\Delta G/RT}}{1 + e^{-\Delta G/RT}} \qquad \text{(eq. 4.4)}$$

4.2 Two-State and Pseudo-Two State Model for HNMR Analysis

When analyzing ^{1}H-NMR data with a two state pPII↔β model, mixed terms of the type pPII-β are entirely unnecessary as the $^3J(H^NH^\alpha)$ coupling constant is site specific for the *ith* amide proton, where we denote i=1 for the amide associated with the central residue and i=2 for the C-terminal amide. Therefore the ^{3}J constant can be written as:

$$^3J_i = \chi_{i,pPII}\,{}^3J_{i,pPII} + \chi_{i,\beta}\,{}^3J_{i,\beta} + \sum \chi_{i,T}\,{}^3J_{i,T} \qquad \text{(eq. 4.5)}$$

Where we include here the mole factions ($\chi_{i,T}$) and intrinsic J-coupling constants ($^3J_{i,T}$) associated with turn-like conformations in anticipation of dealing with GxG peptides, which have been shown by us to have much larger populations of turn-like conformation (11-32%) (55). This high population of turn-like conformations is non-negligible when calculating ensemble averages. However, their populations have also recently been shown by NMR analysis to be temperature independent (141). Therefore, we construct a

pseudo-two state thermodynamic model in which re-distributions among pPII and β are temperature dependent with temperature independent turn conformations. The corresponding equation describing the temperature dependence of $^3J(H^NH^\alpha)$ can be then written as:

$$^3J_i(T) = \frac{^3J_{i,pPII} + ^3J_{i,\beta}e^{-\Delta G_i/RT}}{1+e^{-\Delta G_i/RT}}\left(\chi_{pPII}+\chi_\beta\right)+\sum \chi_{i,T}\cdot {}^3J_{i,T} \qquad \text{(eq. 4.6)}$$

Where the second term vanishes when dealing with alanine based peptides (AAA) for which the turn population ($\sum\chi_{i,T} < 8\%$) approaches zero. To obtain reference values for $^3J_{i,pPII,}$ $^3J_{i,\beta}$ and $^3J_{i,T}$ to be used in equation 4.5 and 4.6, we use the unique pPII, β-strand, and turn sub-distributions obtained from vibrational analysis for each residue. These distribution functions can be subsequently used to calculate statistically meaningful sub-ensemble average $<^3J_{PPII}>$, $<^3J_\beta>$, $<^3J_T>$ expectation values via the newest version of the Karplus equation (163) for use in equation 4.6. This is an improvement over methods where reference coupling constants are obtained by use of residue specific coil library data which may not resemble conformational distributions of amino acid residues in water (51).

These reference coupling constants can then be used in eq. 4.6 along with the average Gibbs free energy difference between pPII and β-strand sub-states:

$$\Delta G_i = -RT\cdot\ln\left(\frac{\chi_{\beta,i}}{\chi_{pPII,i}}\right) \qquad \text{(eq. 4.7)}$$

This can be used to relate ΔH_i and ΔS_i through:

$$\Delta S_i = \frac{\Delta H_i - \Delta G_i(T_R)}{T_R} \qquad \text{(eq. 4.8)}$$

so that

$$\Delta G_i = \Delta H_i \cdot \left(1 - \frac{T}{T_R}\right) + \Delta G_i(T_R) \qquad \text{(eq. 4.9)}$$

was obtained as the equation to be finally inserted into eq. (4.6) to fit $^3J(H^N,H^\alpha)$ (T), thus using ΔH_i as the only free parameter.

CHAPTER 5. THE CHOICE OF PEPTIDE MODEL SYSTEM AND THE PH INDEPENDENCE OF AAA

Reproduced in part, from:

I. Toal, S., Meral, D., Verbaro, D Urbanc, B., Schweitzer-Stenner, R. "pH-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides: A combined Vibrational, NMR, UVCD, and Moleculer Dynamics Study. J. Phys. Chem. B, 2013, 117, 3689-3706.

II. Rybka, K., Toal, S., Verbaro, D, Mathieu, D., Schwalbe, H., Schweitzer-Stenner, R. "Disorder and order in unfolded and disordered peptides and proteins: A view derived from tripeptide conformational analysis. II. Tripeptides with short side chains populating asx and β-type like turn conformations" Proteins, 2013, 117, 3689-3706.

5.1 INTRODUCTION

Investigation of local order in the unfolded state is complicated by experimental limitations and the inherent dynamics of the system, which has in some cases yielded inconsistent results from different types of experiments. One method of studying these systems is the use of short model peptides. Multiple experimental and theoretical evidence utilizing various types of short model peptides has accumulated supporting the notion that the conformational space of most amino acid residues is much more restricted than suggested by the random coil model (51, 54, 55, 68, 188, 189). Moreover, it has become clear that each amino acid residue has a unique and restricted conformational ensemble. In this context, the aforementioned polyproline II (pPII) conformation has emerged as the dominant conformation for many amino acid residues, specifically alanine (54). The absolute number for the pPII preference of alanine has been a matter of debate,

with percentages of pPII ranging from 40-90% (50, 54, 127). However, the most recent NMR and vibrational studies on polyalanines and short alanine based peptides analyzed using realistic conformational distributions confirm the high (>70%) pPII propensity of alanine (74).

While experimental results have qualitatively converged in reporting a high intrinsic pPII propensity for alanine, some issues are still unresolved. One of these is the choice of model peptides for conformational quantification. Blocked dipeptides are often considered as an ideal choice. With respect to alanine, the alanine dipeptide, Ac-Ala-NHMe "AdP", has been the classical model system particularly for computational studies of alanine conformations (76, 93, 95-97). Several experimental studies (IR, Raman, NMR) on this peptide have been carried out as well (79, 190). As outlined in *Chapter 1 of this thesis*, many recent experimental investigations have chosen to use various blocked dipeptides or blocked tripeptides (56, 132, 189). In contrast, other researchers use unblocked peptides, such as trialanine (AAA) and GxG (x: different guest amino acid residues), for conformational studies, in part because these types of peptides allow more comprehensive NMR studies and provide a better spectral resolution in the amide I window of vibrational spectra (54, 69, 70, 74, 75, 191).

Recently, Kallenbach and coworkers launched a major criticism of the use of tripeptides for conformational studies (192). The authors cite the fact that four guest residues in GxG, AcGxGNH2, and AcGGxGGNH2, and the respective dipeptides, show slightly different $^{3}J(H^{N}H^{\alpha})$ coupling constants at different pH as an argument for the

influence of terminal groups biasing results. Using a two-state analysis of ^{3}J coupling data along with reference J_{pPII} and J_{β} values obtained from pPII/β maxima in coil libraries, the authors obtained an increase in pPII content along the series (GxG)<(AcGxGNH$_2$)<(AcGGxGGNH$_2$). This analysis led He et al. to conclude that the free terminal groups of, e.g., GxG cause a 15% reduction of pPII propensities of the central residue and that blocked dipeptides or even blocked glycine-based host-guest systems would be more appropriate model systems (192). However, caution has to be taken when solely analyzing $^{3}J(H^{N}H^{\alpha})$ constants, as the observed differences could well arise from small shifts of conformational distributions in the Ramachandran space rather than variations in percent pPII.

To address this issue, we explored the influence of terminal groups on central amino acid residues in short alanine peptides by performing a combined analysis of NMR coupling constants and amide I' band profiles of all three protonation states of AAA as well as of the alanine dipeptide (AdP) (56). Thus, we are addressing two questions: (1) to what extent does the protonation state of the terminal groups affect the intrinsic conformational propensity of central amino acid residues in tripeptides with unblocked termini, and (2) how does termini blocking (i.e., "capping") affect this conformational propensity? In addition to determining the influence of free termini on central alanine residue's conformational distribution at room temperature, we also explore the thermodynamics governing the pPII preference for AdP and AAA in all protonation states

by analyzing the temperature dependence of conformationally sensitive CD and NMR parameters.

In the second, computational part of this investigation we utilized molecular dynamic simulations (MD) (in collaboration with Dr. Urbanc, and Derya Meral, Drexel University). As indicated above, the assumed suitability of AdP as the simplest model system for studying peptide conformations has led to a flood of MD studies on this peptide *in vacuo* and in aqueous solution (76, 95-100, 193) (see *Chapter 1 of this thesis*). One of the reasons for this multitude of studies is that MD simulations of unfolded peptides heavily depend on the choice of the force field (88, 109). While earlier simulations with CHARMM and AMBER force fields led to an overemphasis of right-handed helical conformations (95, 105), more recent modified CHARMM and AMBER as well as OPLS force fields yielded a dominant population of the pPII/β conformations in the upper left quadrant of the Ramachandran plot (88, 94, 101). In an effort by theorist to match experiment, some advances have recently been made by combining MD simulation for the solvent with DFT calculations for the peptide (74, 193). However, most of these simulations still predict pPII populations in amounts well below the recently reported experimental values. Exceptions to the rule are strongly modified AMBER and CHARMM force fields, which yielded a rather high fraction of pPII (0.8 and 0.99) for AAA and AdP, respectively. However, the physical rationale for these changes (i.e. eliminating the torsional potential for φ and ψ in AMBER and a highly polarizable version of CHARMM) remain somewhat obscure (96, 194). Kwac et al. performed MD

simulations of AdP with several normal and polarizable force fields and different water models and found that only the combination of a polarizable AMBER ff02 force field with a polarizable water model yielded pPII fraction slightly higher than 0.5 (96). Here, our experimental studies of AAA and AdP are complimented by MD simulations that combine two common force fields with several commonly used water models. The relative validity of the resulting conformational manifolds was evaluated by comparison to experimentally derived distributions.

5.2 Results and Discussion

5.2.1 Trialanine conformations derived from Amide I' simulation are pH-independent.

In this section it is shown that the conformational distribution of the central amino acid residue of AAA in aqueous solution is practically independent of the protonation state of the terminal groups. To this end we first measured and analyzed the IR, polarized Raman, and VCD profiles of cationic AAA utilizing the four ^{3}J-coupling constants dependent on ϕ and the two $^{2(1)}$J coupling constants dependent on ψ reported by Graf et. al. as simulation restraints (74). The total set of amide I′ profiles of all three protonation states of AAA is shown in Figure 5.1.

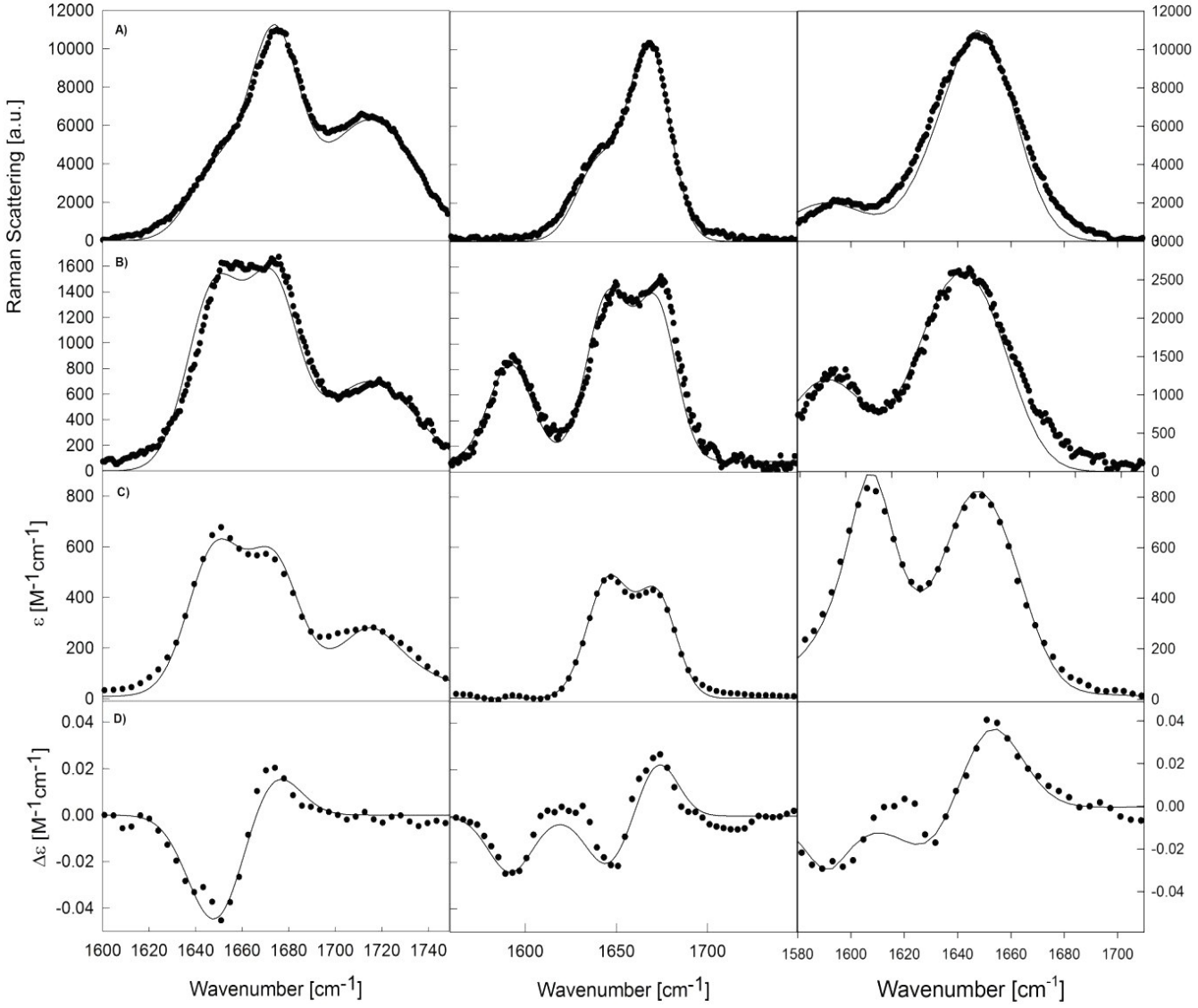

Figure 5.1: Isotropic Raman (A), anisotropic Raman (B), IR (C), and VCD (D), band profiles of the amide I' mode of cationic AAA (left column), zwitterionic (middle column) and anionic (right column) in D_2O. The solid lines result from the optimized simulation.

Although the respective profiles look different, this is primarily due to (a) the overlap with bands outside of the amide I region (CO stretch above 1700 cm^{-1} and COO antisymmetric stretch below 1600 cm^{-1} in the spectrum of cationic and zwitterionic AAA, respectively) and (b) the electrostatic influence of the protonated N-terminal group on the N-terminal amide I modes. In the absence of the N-terminal proton the amide I shifts down by 40 cm^{-1}. This leads to a much stronger overlap with the amide I band predominantly assignable to the C-terminal peptide group.

The experimental amide I' band profiles and J-coupling constants were fit according to the theoretical protocol based on excitonic coupling. The extent of spectral changes depends on the strength of excitonic coupling and hence on the dihedral angles of the central amino acid residue. This brings about the conformational sensitivity of amide I band profiles (167). The underlying theory of excitonic coupling as well as our formalism used for the simulation of amide I band profiles are outlined in detail in *Chapter 2 of this thesis*. Here, it is sufficient to mention that the (ϕ, ψ) dependence of amide I and J-coupling constants are accounted for by mathematically describing the mixing of excited vibrational states via excitonic coupling(167, 177) and by Karplus relations for J-coupling constants (160, 163), respectively. The variable parameters for simulation are the 2D Gaussian distributions reflecting the conformational ensemble.

The results of the optimized simulation are depicted by the solid lines in Figure 5.1 and the calculated J-coupling constants in Table 5.1. The final simulation of all amide I' profiles is in very good agreement with experiment. The mole fractions, ϕ and ψ coordinates, and half-halfwidths of the sub-distributions used to attain these final simulations are listed in Table 5.2. In agreement with the results of Graf et al. (74) and Schweitzer-Stenner et al. (90) the analysis reveals a dominant pPII fraction of 0.84 for the central residue of cationic AAA, the remaining conformations are β-strand, type II β-turn, right-handed helix and γ-turn-like. The addition of minor turn fractions were necessary in order to reproduce all J-coupling constants without deteriorating the simulation of amide

I' profiles. The respective mole fractions of these sub-conformations carry an uncertainty of up to 5%.

Table 5.1: Comparison of experimental and calculated J-coupling constants for cationic AAA.

COUPLING CONSTANT	EXPERIMENTAL [HZ]	CALCULATED [HZ]
$^{3}J(H^{N}H^{\alpha})$	5.68 ± 0.07	5.63
$^{3}J(H^{N}C')$	1.13 ± 0.1	1.09
$^{3}J(H^{\alpha}C')$	1.84 ± 0.3	1.57
$^{3}J(H^{N}C^{\beta})$	2.39 ± 0.25	2.10
$^{1}J(NC^{\alpha})$	11.34 ± 0.2	11.19

Table 5.2: Center (ϕ,ψ)-coordinates denoted in parenthesis, and respective mole fractions of the two dimensional Gaussian sub-distributions used for simulation of Vibrational Spectra and J-coupling constants for Cationic AAA (AAA+), Zwitterionic AAA (AAA+-), Anionic AAA(AAA-), Alanine dipeptide (AdP), and cationic GAG (GAG+).

Conformation	AAA+	AAA+-	AAA-	AdP	GAG+
pPII	0.84 (-69,145)	0.84 (-69,145)	0.84 (-69,130)	0.74 (-69,160)	0.72 (-69,155)
β-strand	0.08 (-125,160)	0.08 (-125,160)	0.08 (-125,150)	0.16 (-115,160)	0.18 (-115,155)
right-hand helical	0.04 (-60,-30)	0.04 (-60,-30)	0.04 (-60,-30)	0.04 (-60,-30)	0.03 (-60,-30)
inverse γ-turn	0.04 (-85,78)	0.04 (-85,78)	0.04 (-85,78)		
type II β-turn				0.03 (-60,120)	0.03 (-60,-120)
type I' β-turn				0.03 (20,40)	
inverse γ-turn					0.04 (20,-60)

We used the conformational ensemble obtained for cationic AAA to simulate the amide I' profiles of zwitterionic and anionic AAA (Figure 5.1). For the former, we used the $^3J(H^NH^\alpha)$ of the N-terminal amide proton to constrain our simulation. The $^3J(H^NH^\alpha)$ of the N-terminal amide proton for anionic AAA is unattainable due to loss of NMR resolution in basic conditions. For anionic AAA, it was necessary to use different intrinsic wavenumbers for the individual local amide I modes, since the deprotonation of the N-terminal is known to shift the respective amide I' mode wavenumber from 1672 to 1635 cm^{-1} (195). This causes a much larger overlap with the amide I' band of the C-terminal peptide group (1649 cm^{-1}). Otherwise, we achieved the best fit of the amide I' band profile of both protonation states with only minor variations of the distribution function obtained for the cationic state. The parameters of the conformational distributions for zwitterionic AAA and anionic AAA are listed in Table 5.2. Any significant changes made to either the occupation or breadth of sub-distributions defining the conformational ensemble resulted in less accurate simulations of amide I' profiles and J coupling constants for both protonation states. The $^3J(H^NH^\alpha)$ = 5.74 Hz coupling constant observed for the zwitterionic state was exactly reproduced. Most importantly, the mole fractions obtained for each conformation remain essentially unaltered among the three different protonation states of AAA. The central alanine in all three protonation states was found to have a high pPII content of 84% followed by β-strand conformation (8%) and minor fractions of turns totaling 8%. The positions of the corresponding sub-distributions for all three protonation states of AAA show only slightly different ϕ and ψ

values. Upon deprotonation of the carboxyl group of cationic AAA there is no discernible conformational difference. The only remarkable change is that the pPII distribution shifts to lower ψ-coordinates upon deprotonation of the N-terminal in forming anionic AAA (Table 5.2). The small difference between the experimental $^3J(H^NH^\alpha)$ coupling constants of cationic ($^3J(H^NH^\alpha)$=5.68Hz) and zwitterionic AAA ($^3J(H^NH^\alpha)$=5.74Hz) are accounted for by a very small shift of the ϕ-coordinate of the pPII sub-distribution. Taken together, the data show no substantial decrease of the pPII population upon the deprotonation of either termini, in contrast to what He et al. reported for GxG peptides using a two state analysis. Our results also show that differences between $^3J(H^NH^\alpha)$ coupling constants can reflect small changes of coordinates of sub-distribution rather than variations of their statistical weight. This issue is often overlooked in studies determining conformation in peptides and proteins (50, 133, 189, 192, 196). Since local residue conformation may significantly differ from canonical values, assuming static distributions with variant mole fractions may be an over-simplification. The combined analysis of amide I profiles along with J coupling constants, and particularly the sensitivity of the VCD signal strength, is useful here for discriminating between population and coordinate changes (54).

5.2.2 Alanine Dipeptide has a conformational ensemble similar to corresponding GAG

In addition to the various protonation states of trialanine, the amide I' band profiles of AdP were also measured and analyzed in order to check if termini blocking may effect conformational ensembles. For reference, the structures of AAA, GAG, and AdP are shown in Figure 5.2. While GAG consists of an alanine residue flanked by unblocked glycine based "end groups," the alanine dipeptide contains an alanine residue capped by acetyl group (N terminal) and methyl amide (C-terminal). The IR and Raman profiles for AdP are very reminiscent of what was obtained for anionic AAA, owing to the absence of the charge on the N-terminal group, but the VCD is negatively biased due to the intrinsic magnetic moment of the C-terminal (Figure 5.3) (197).

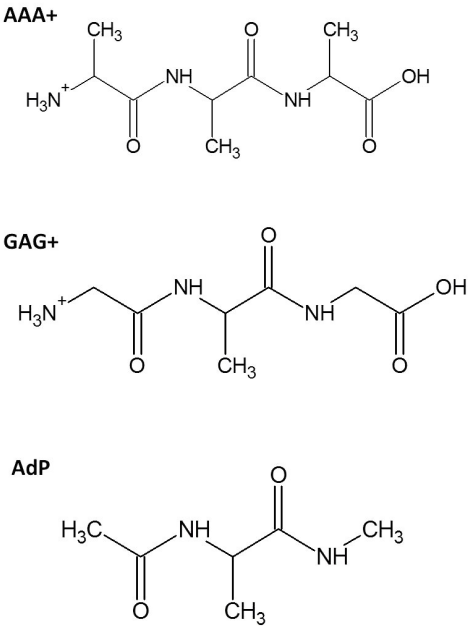

Figure 5.2: Chemical Structure of AAA, GAG, and AdP peptides.

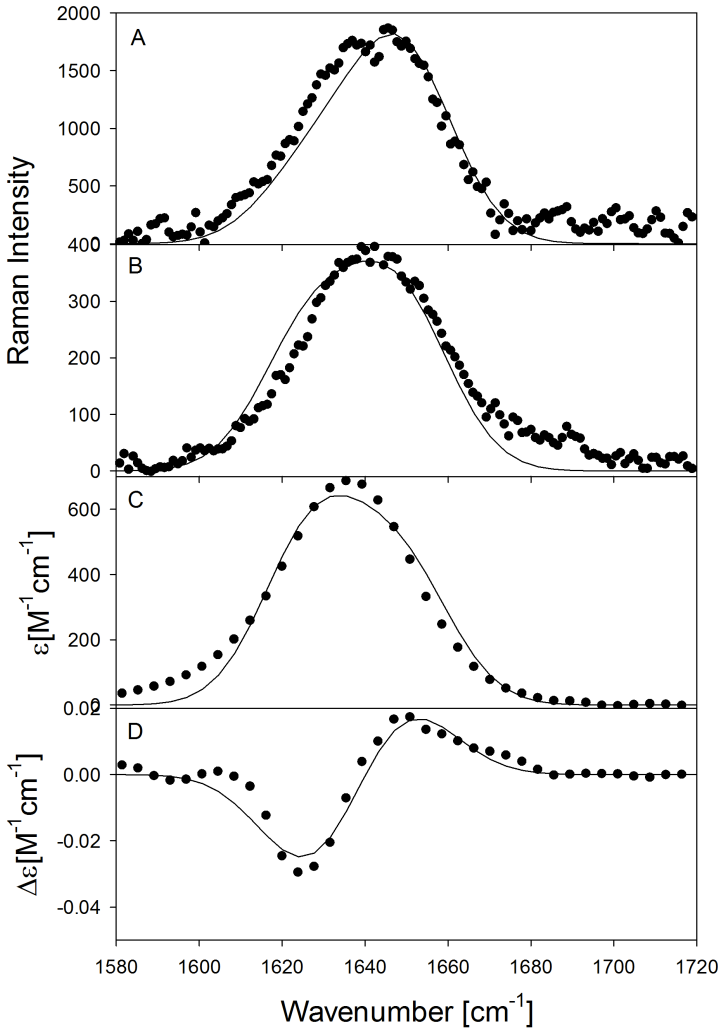

Figure 5.3: (A) Isotropic Raman, (B) anisotropic Raman, (C) IR, and (D) VCD band profiles of the amide I' mode of AdP in D_2O. The solid lines result from the simulation described in the text.

As seen in Figure 5.3, the VCD signal was fully reproduced by the optimized simulation. In addition, the experimental $^{3}J(H^{N}H^{\alpha})$ at room temperature (5.88Hz) was also fully reproduced by the final simulation ensemble. The resulting sub-states and respective statistical weights of the final simulation ensemble for AdP are listed in Table 5.2. The pPII fraction of the alanine residue in the dipeptide analog is slightly lower (74%) than the value observed for all protonation states of AAA (84%). The respective ψ-values for the pPII conformation in AdP are also slightly downshifted relative to trialanine. Interestingly, the final conformational ensemble obtained for AdP (Table X) is very similar to what Hagarman et al. previously reported for the unblocked GAG peptide (54). Altogether, the distributions of AdP and GAG (Table 5.2) agree quite well. Actually, this is what one might expect in view of the fact that in both GAG and AdP peptides, the two peptide bonds surrounding the central alanine residue are directly flanked by methylene and methyl groups respectively. Therefore, the blocked terminal CH_3-groups of AdP are somewhat more reminiscent of glycine's methylene Cα component rather than the CH-CH_3 group in alanine since glycine lacks the additional β-carbon. The conformational similarity of alanine within AdP and GAG suggests that the interaction between the end groups of a dipeptide with the central residue is analagous to the (most likely weak) interaction between terminal glycine and the central residue in GxG. This indicates that the strength of neighbor interactions is practically absent for any atoms beyond nearest-neighboring Cα side-chains. Said another way, the terminal side chains block any detectable influence of charged end groups on the conformation of the

middle “guest” residue under investigation. The only remaining difference between GAG and AdP are the free termini of glycine which are absent in AdP, but present in GAG. Since we find the central alanine residue in these two peptides have nearly identical conformational ensembles, our results demonstrate a very limited influence of terminal charges on non-ionized central residues of tripeptides.

To check the generality of the above results for non-alanine residues, we examined the unblocked fully protonated Gly-Val-Gly (GVG) peptide and the valine dipeptide (VdP). Figure 5.4 shows the experimental polarized Raman, IR, and VCD, amide I’ profiles for GVG and VdP. Following the same theoretical protocol above, and as previously described (see *Ch 2 of this thesis*), we simultaneously simulated all amide I’ profiles for GVG and VdP utilizing the six conformationally sensitive J-coupling constants as restraints. The final fit to experimental data is plotted as the solid lines in Figure 5.4. The $^{3}J(H^{N}H^{\alpha})$ coupling constants for both valine peptides are very well reproduced by our simulation procedure. The thus obtained conformational ensemble for GVG and VdP (Table 5.3) are both similar to those recently reported for the GVG peptide (198). In contrast to the alanine peptides, GVG has a much decreased pPII content (χ_{pPII}=0.32) in preference for an increased sampling of β-strand-like conformation (χ_{β} =0.46) in agreement with literature. The ϕ and ψ coordinates of these sub-distributions are also shifted to lower and higher values, respectively, as compared to those obtained for the alanine-based peptides.

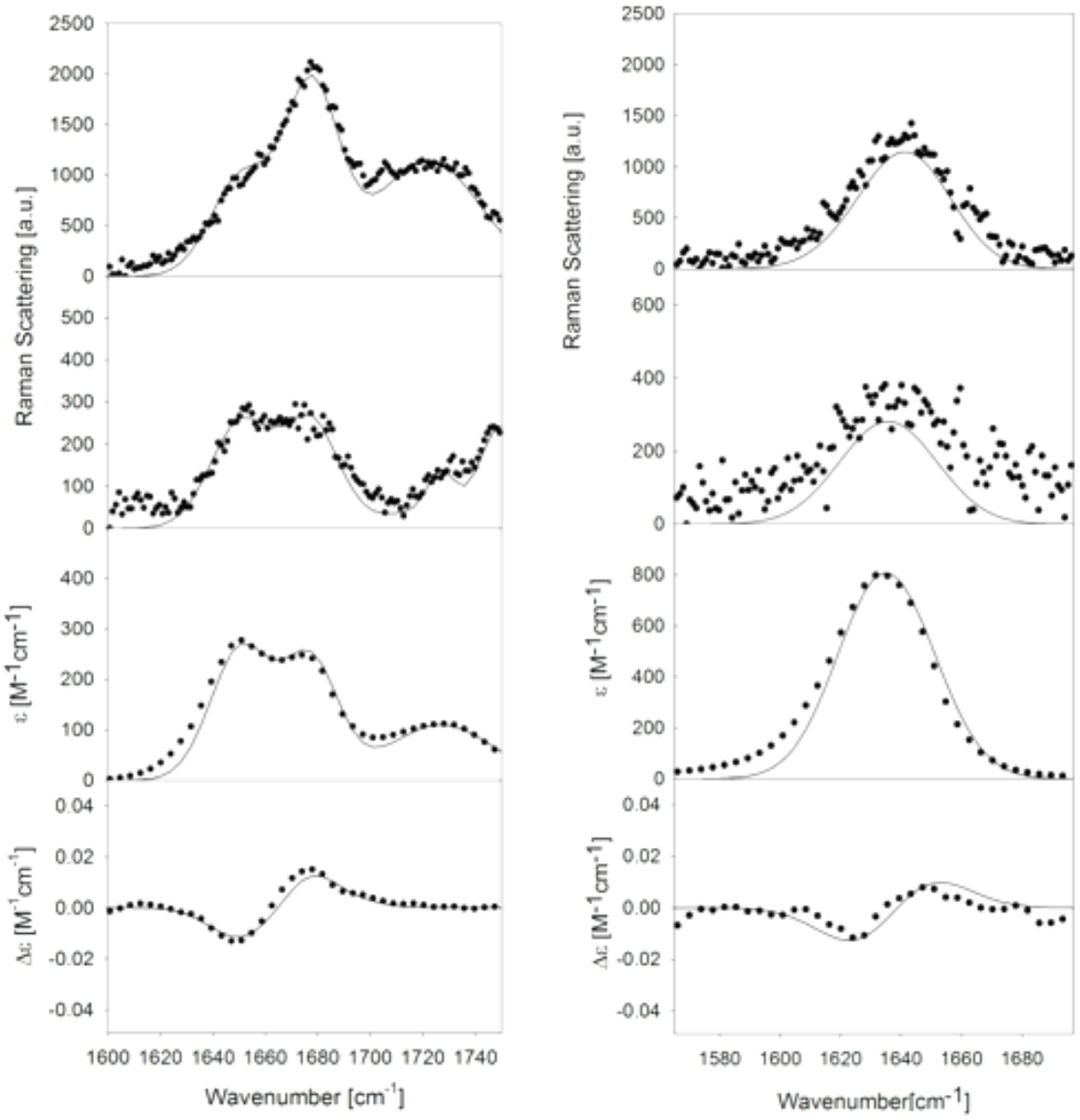

Figure 5.4: Isotropic Raman, anisotropic Raman, IR and VCD band profiles of the amide I' mode of GVG (left panel) and VdP (right panel) in D_2O. Experimental data was taken for GVG was taken and modified from Hagarman et al.(54) The solid lines result from the simulation described in the text.

Table 5.3: Center (ϕ,ψ)-coordinates and mole fractions of the two-dimensional Gaussian sub-distributions obtained for cationic GVG and the valine dipeptide (VdP)

Conformation	GVG+	VdP
pPII	0.32 (-80,170)	0.32 (-78,150)
β-strand	0.46 (-120,170)	0.46 (-120,150)
right-hand helical	0.04 (-60,-30)	0.04 (-60,-30)
inverse γ-turn	0.11 (-60,60)	0.11 (-60,60)
(i+1)residue type I' β-turn	0.07 (60,30)	0.07 (60,30)

Similar to the case of alanine peptides, the experimental data for the VdP could be reproduced with nearly the same conformational distribution and statistical weights obtained for GVG. This result again demonstrates that there is limited conformational influence of terminal groups on central residues in tripeptides, and moreover, that the similarity of uncapped glycine termini to methyl-blocked termini holds true even for peptides with non pPII-preferring central residues.

While these results indeed show negligible end-group effects on conformations of aliphatic residues in tripeptides, one might still expect a different situation for polar and/

or ioniziable side chains. In order to check whether end effects are influential in these types of residues we chose to analyze the capping effect on aspartic acid. As described in detail in *Chapter 1.3.2.1 of this thesis*, in previous GxG conformational studies we found aspartic acid to be a member of a special sub-class of amino acid residues (D, N, T, S) that all have above-average (>30%) preference for conformations generally found in turn-supporting structures (55). Structurally, these residues have relatively short flexible side chains with some degree of H-bonding capability. One might hence wonder whether the propensities for turns and turn-like conformations could be induced by the influence of the partially charged terminal group of unblocked peptides. To test this we measured and analyzed the amide I' profiles and $^{3}J(H^{N}H^{\alpha})$ coupling constants of the aspartic acid dipeptide (DdP) at the same conditions (acidic pH) as the afore-studied GDG peptide. The amide I' profiles for DdP are shown in Figure 5.5. The solid lines in this figure result from a simulation for which we used exactly the same conformational distribution as obtained for GDG. As one can infer from Figure 5.5, the resulting simulation for DdP agrees well with experiment. The conformational ensembles for both DdP and GDG are dominated by 48% β-strand like conformations followed by turn-like conformations (30%) and pPII (20%). In addition, the simulated $^{3}J(H^{N}H^{\alpha})$ coupling constant (7.55Hz) is close to the experimental value obtained for DdP of 7.47 Hz. Any attempt to reduce the turn population or increase either extended sub-populations (i.e., pPII or β strand) resulted in deterioration in both amide I' profiles as well as $^{3}J(H^{N}H^{\alpha})$ coupling. The fact that aspartate samples the same conformational manifold in a free glycine environment

(GDG) and in the blocked dipeptide (DdP) indicates that the obtained asx-turn population for aspartic acid in the GxG peptide model is not caused by the terminal groups but rather an intrinsic property of the aspartate residue. These results, along with those on AdP and VdP underscore the notion that GxG and blocked dipeptides (XdP) are both suitable model systems for investigating the intrinsic conformational preferences of amino acid residues.

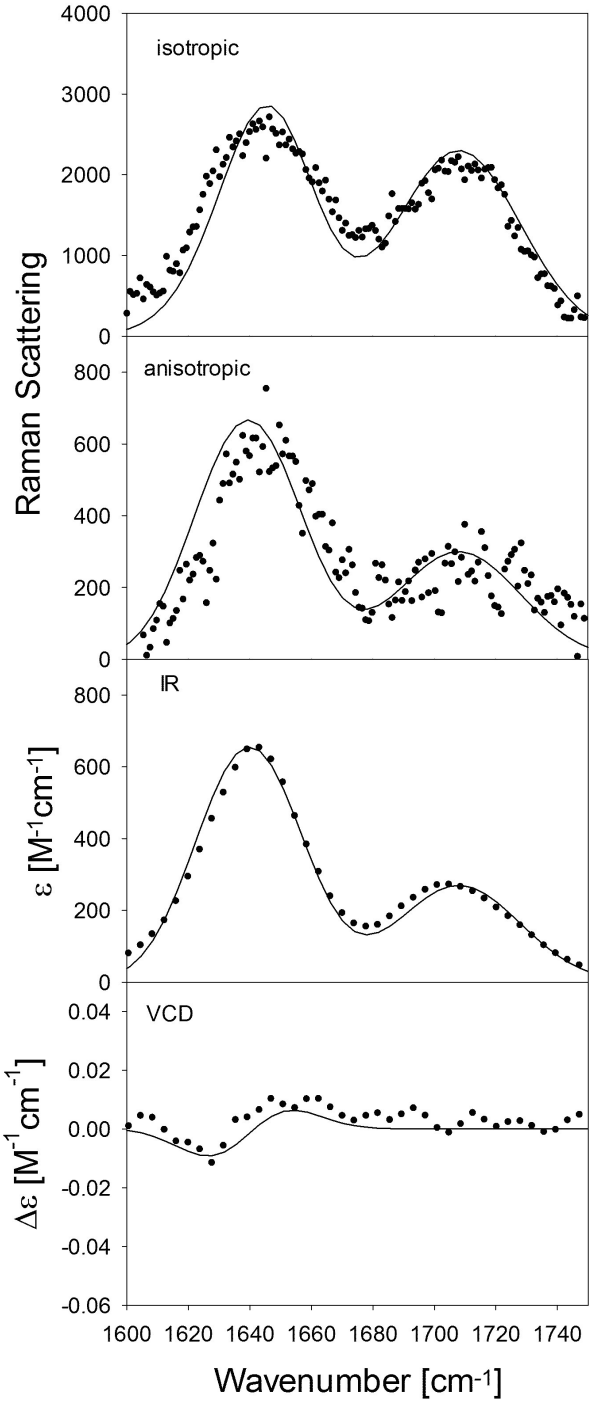

Figure 5.5: Amide I' profiles of the isotropic Raman, anisotropic Raman, IR and VCD spectra of DdP at pH 2. The solid lines result from the simulations described in the text.

5.2.3 The Gibbs energy landscape of alanine residues in unblocked tri- and blocked dipeptides is not influenced by end-effects.

To further explore the factors stabilizing the conformational distributions of the three alanine based peptides (cationic AAA, zwitterionic AAA, and AdP), we characterized their ensembles in thermodynamic terms. While the above studies revealed very limited differences between the protonation states of AAA and AdP, it is possible that differences emerge at e.g. higher temperatures because of different enthalpic and entropic contributions between coexisting conformations. Indeed, an analysis of UVCD spectra of cationic and zwitterionic AAA has led Oh et al. to the conclusion that the thermodynamic parameters of the two protonation states are different (186). In a first step, we measured the far UVCD spectra of zwitterionic AAA and Adp as a function of temperature between 5^{o}C and 85^{o}C, which are shown in Figure 5.6. A comparison of these spectra with UVCD spectra for cationic AAA measured between 0^{o}C and 90^{o}C is also shown in Figure 5.6. To facilitate comparison of the three peptides they are all plotted on the same scale in units of $\Delta\varepsilon$ [$M^{-1}cm^{-1}residues^{-1}$], where the number of residues contributing to the UVCD signal for AAA and AdP are 2 and 1, respectively. At low temperature, all three of these alanine based peptides exhibit UVCD signals characteristic of a dominant sampling of pPII conformation, in agreement with what is now well established in literature for alanine (see *Chapter 1.2.4 of this thesis*). Cationic AAA is most prominent in this regard, with a positive maximum at approximately 215 nm and a pronounced negative maximum at 190 nm. The insets in Figure 6 depict the difference spectra calculated by subtracting the lowest temperature spectra from the highest

temperature spectra, and are all indicative of a population re-distribution from pPII to more β-like conformations.

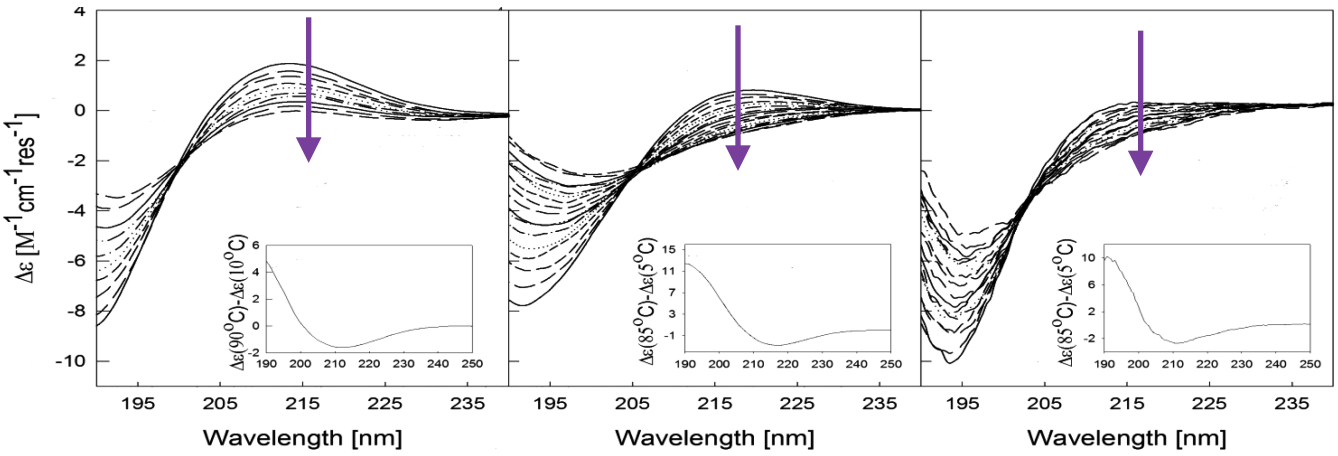

Figure 5.6. UVCD spectra of (left panel) cationic AAA, (middle panel) zwitterionic AAA, and (right panel) the AdP as a function of temperature. Cationic AAA spectra range from 0 to 90 °C with ΔT = 10 °C. Zwitterionic AAA and the alanine dipeptide range from 5 to 85 °C with ΔT = 5 °C. The arrows indicate increasing temperatures. The inserts show the difference spectrum ΔΔε between highest and lowest temperatures.

A word of caution deserves to be mentioned here regarding the use of UVCD to characterize peptide conformation. Although UVCD spectra can provide powerful qualitative information, the sole use of this technique to define conformational populations in peptides is problematic and may yield ambiguous results. However, the

ability of UVCD to track spectral *changes* reflecting population re-distributions with e.g., changing temperature can indeed provide useful information regarding the energetics of the system, especially when backed up by *a priori* knowledge of conformational sub-space.

Although the temperature dependence of the UVCD spectra for all three alanine based peptides is qualitatively similar, a direct comparison of cationic AAA with zwitterionic AAA and AdP reveals distinct differences in the spectral line shape at all temperatures. As reported earlier (186, 192), the spectra for zwitterionic AAA is noticeably red-shifted as well as lower in intensity at both the positive and negative maxima compared to that of cationic AAA. Although we find that the UVCD spectral line shape at both low and high temperatures for cationic and zwitterionic AAA differ, it is not likely that this difference is due to structural changes, as this would be reflected in a significant change in the $^{3}J(H^{N}H^{\alpha})$ constants for each peptide. For the amide hydrogen associated with the central residue of cationic and zwitterionic AAA we obtain $^{3}J(H^{N}H^{\alpha})$ values at room temperature of 5.68 and 5.74 Hz, respectively. In addition, *a priori* knowledge from the above-described vibrational analysis dictates that the conformational ensemble of both protonation states of trialanine are nearly identical. Therefore, it is more likely that this pH-dependent spectral change in the UVCD is due to interference of the charge transfer (CT) transition between the C-terminal carboxylate and the peptide group of zwitterionic AAA. This transition has been previously reported by Pajcini et al.(199) for glycylglycine and by Dragomir et al. (52) for AX and XA peptides, and is assignable

to a n_{coo}->π* transition. Indeed, it was found by Dragomir et al. (52) that the frequency position of this CT band correlates well with the positive dichroic maxima of pPII in the respective UVCD spectrum.

A comparison of the UVCD spectra of cationic AAA with AdP reveals differences in line shape at both low and high temperatures. Because AdP is blocked at the C-terminal carboxylate, these spectral changes cannot be a result of the CT transition. The positive maximum at 210 nm, diagnostic of pPII conformation, is noticeably decreased for AdP relative to cationic AAA, indicating less sampling of pPII-like conformation in favor of more extended conformations. This is in agreement with the results from our above vibrational analysis in which we obtain a slightly lower pPII fraction for AdP and an increased β-content relative to both cationic and zwitterionic AAA.

The temperature dependence of the UVCD for each peptide displays an isodichroic point (Figure 5.6), indicating that all three peptides predominantly sample two conformational states within the temperature region (i.e pPII- and β-like). This two-state behavior is typical of short alanine-based peptides (200), and is again in line with the conformational ensembles we obtained for these peptides through the simulation of the amide I' vibrational profiles (Table 5.2).

In order to investigate the free energy landscape of each alanine peptide, we employed a global fitting procedure to analyze the temperature dependence of the conformationally sensitive maximum dichroism $\Delta\varepsilon$ (T) and the $^3J(H^NH^\alpha)$(T) values with a two-state pPII-β model (see Ch 4). To be consistent with the conformational ensembles of

each peptide derived above, we began the fitting process by using the statistical average $<^3J_{pPII}>$ and $<^3J_{\beta}>$ and the Gibbs energy difference between the peptide-specific pPII and β distributions derived from our vibrational analysis. However, this process originally led to a poor fit to the experimental $^3J(H^NH^{\alpha})(T)$ data. This is likely due to the presence of weakly populated conformations in the conformational ensembles of the investigated peptides. For both ionization states of AAA, vibrational analysis revealed that 8% of the conformational ensemble is not of pPII/β type. For AdP this number is 11%. To compensate for this slight deviation from two-state behavior we lowered the average ϕ_{pPII}-value, representing the center of the pPII sub-distribution, relative to that obtained from our vibrational analysis. The best fit to the thermodynamic data was achieved by lowering ϕ_{pPII} by 0.25^o and 0.36^o per 1% population of non-pPII/β conformations for AAA and AdP, respectively.

The thus modified distribution was subsequently used to calculate statistical average $<^3J_{PPII}>$ and $<^3J_{\beta}>$ expectation values via the newest version of the Karplus equation (163). The final values of $<^3J_{PPII}>$ and $<^3J_{\beta}>$ obtained from this procedure are 5.02 Hz and 9.18 Hz, respectively, for cationic AAA, 5.09 Hz and 9.18 Hz for zwitterionic AAA, and 4.69 Hz and 9.17 Hz for AdP. We used these 'effective' reference coupling constants and the respective experimental $^3J(H^NH^{\alpha})$ values to calculate the mole fractions of pPII and β-strand conformations for the residues in each alanine peptide. This procedure results in pPII mole fractions for the central residues, $\chi_{i=1}$(pPII), of 0.86, 0.84, and 0.74 for cationic AAA, zwitterionic AAA, and AdP, respectively, which exactly

match the mole fractions we derived from our vibrational analysis of amide I' modes (Table 5.2). This shows that our forced reduction to a two-state model for the thermodynamic analysis indeed preserved the Gibbs energy difference between the pPII and β-strand conformations. This observation indicates that the population of turn conformations might not be very temperature dependent, in agreement with recent theoretical predictions and experimental results (141, 201). For the C-terminal residue, we obtained pPII fractions of 0.67 and 0.60, for cationic and zwitterionic AAA, respectively.

Using the calculated reference $<^3J>$ values obtained, we could then employ equation 4.7 to fit the experimental $^3J(T)$ data and extract thermodynamic information regarding the pPII/β-strand equilibrium for all peptides. The resulting fits for all three peptides are shown as solid lines in Figure 5.7. The thermodynamic parameters obtained from this procedure are shown in Table 5.4. For the central residue of cationic AAA we obtain ΔG_1= -4.44 kJ/mol at room temperature, with an enthalpic difference between pPII and β-strand of ΔH_1 = -20.7 kJ/mol, whereas the entropy difference is ΔS_1= -54.4 J/ mol*K. These values reflect the fact that pPII is enthalpically stabilized whereas β-strand is entropicallys stabilzed. These values are all somewhat larger than those obtained by Oh et al.(186), who simultaneously analyzed UVCD and NMR data of cationic AAA using an iterative non-statistical approach to find reference J_{pPII} and J_β values. Their analysis yielded ΔH= -10.63 kJ/mol and ΔS= -50.2 J/mol*K. In contrast, our values are only slightly lower than those obtained by the joint MD/NMR studies of Graf et al.(74) (ΔH =

-24.8 kJ/mol and ΔS = -62.2 J/mol*K), where each reference coupling constant was calculated by directly averaging the Karplus-derived coupling constants over all MD conformations within the sub-state. In view of the uncertainties of the obtained thermodynamic parameters our results and those reported by Graf et al. can be considered as being in reasonable agreement. In a previous study on solvation effects on the conformation of AAA (75), we employed a slightly different fitting approach, using central and C-terminal residue thermodynamic parameters to calculate the temperature dependence of the effective equilibrium constant and Gibbs free energy for the net pPII↔β-strand transition, which was then used to fit the $\Delta\varepsilon(T)$ data. However, we have since revised our $\Delta\varepsilon(T)$ fitting procedure to explicitly account for the contributions from the four different peptide conformations (see eq. 4.6). The thermodynamic values obtained with the revised equation are qualitatively similar to those listed by Toal et al. (75), although slightly less negative. As mentioned above, we calculated $\chi_{1,pPII}$=0.84 and $\chi_{2,pPII}$=0.67 for zwitterionic AAA from the respective experimental $^3J(H^NH^\alpha)$ coupling constants and conformational distributions. This yields Gibbs free energy differences at room temperature for the pPII↔β-strand transition of ΔG_1=-4.41kJ/mol and ΔG_2=-1.01kJ/mol, which are noticeably similar to those obtained for cationic AAA ($\chi_{1,pPII}$=0.84, ΔG_1=-4.44kJ/mol and $\chi_{2,pPII}$=0.67, ΔG_2=-1.71kJ/mol).

To further explore whether there is an influence of the terminal groups on the thermodynamics of zwitterionic AAA we used the ΔH_i values obtained from $^3J(H^NH^\alpha)$ (T) of cationic AAA to fit the corresponding data of zwitterionic AAA. The resulting fit is in

good agreement with the experimental data (Figure 5.7), indicating that the free energy landscape of un-blocked AAA is indeed very similar across all protonation states. From this fit to zwitterionic AAA data, we obtain slightly higher entropic contributions for both residues (i.e, ΔS_1=-55.3 J/mol*K for the central residue and ΔS_2= -32.3 J/mol*K for the C-terminal) as compared to cationic AAA (Table 5.4).

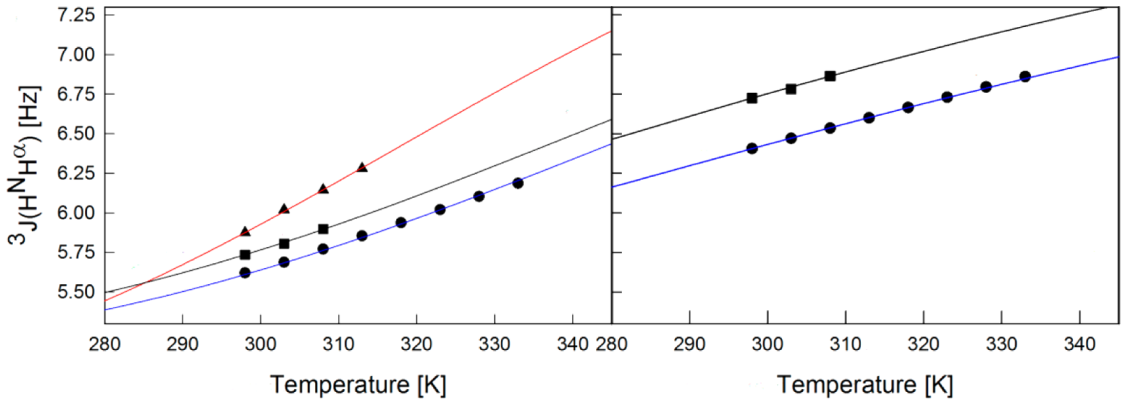

Figure 5.7. $^3J(H^NH^\alpha)$ [Hz] of the central (left panel) and C-terminal residue amide (right panel) plotted as a function of temperature for cationic AAA (circles), zwitterionic AAA (squares) and the AdP (triangles). The solid lines result from the two-state thermodynamic model fitting procedure described in the text.

Table 5.4: Spectroscopic and thermodynamic parameters derived from fitting the temperature dependence of the $^3J(H^N,H^\alpha)$ coupling constants for cationic AAA (AAA+), zwitterionic AAA (AAA+-), and the alanine dipeptide (AdP) using the two-state fitting procedure described in the text.

	^3J [Hz]	J_{pPI} [Hz]	J_β [Hz]	χ_1 (pPII)	ΔG (kJ/ mol)	ΔH (kJ/ mol)	ΔS (kJ/ mol)	χ_2, (pPII)	ΔG (kJ/ mol)	ΔH (kJ/ mol)	ΔS (kJ/ mol)
AAA+	5.61	5.02	9.18	0.86	-4.44	-20.6	-54.4	0.67	-1.71	-10.63	-29.92
AAA+-	5.74	5.09	9.18	0.84	-4.41	-20.6	-55.32	0.60	-1.01	-10.63	-32.27
AdP	5.87	4.63	9.17	0.74	-2.5	-25.2	-66.1	n/a	n/a	n/a	n/a

To fit the experimental H-NMR data for the AdP, we were restricted to a single set of $^3J(H^NH^\alpha)$ (T) data as AdP does not have a second amide proton coupled to a C_α proton. As described above, using the distribution obtained from amide I' profiles and the experimental room temperature value of $^3J(H^NH^\alpha)$=5.9 Hz, we obtain a pPII population of $\chi_{1,pPII}$=0.74, in agreement with our vibrational analysis. This population is associated with a Gibbs free energy difference between pPII and β-strand of ΔG_1= -2.5 kJ/mol (Table 5.4). The result of the final fit using dH_i as the sole free parameter is shown in Figure 5.7. From this fit, the corresponding ΔH_1 and ΔS_1 values were obtained (i.e., ΔH_1= -25.2kJ/ mol and ΔS_2= -66.1J/mol*K), which are comparatively larger than the respective values obtained for both protonation states of AAA (Table 5.4).

With the thermodynamic parameters for each peptide derived above, the analysis of the Δε(T) data could be carried out using the mole fraction weighted Boltzmann distributions for AAA and AdP respectively. The final fit to the experimental data is shown as solid lines in Figure 5.8. From this analysis we obtain the conformation-specific spectroscopic parameters, $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$, which are listed for each peptide in Table 5.5.

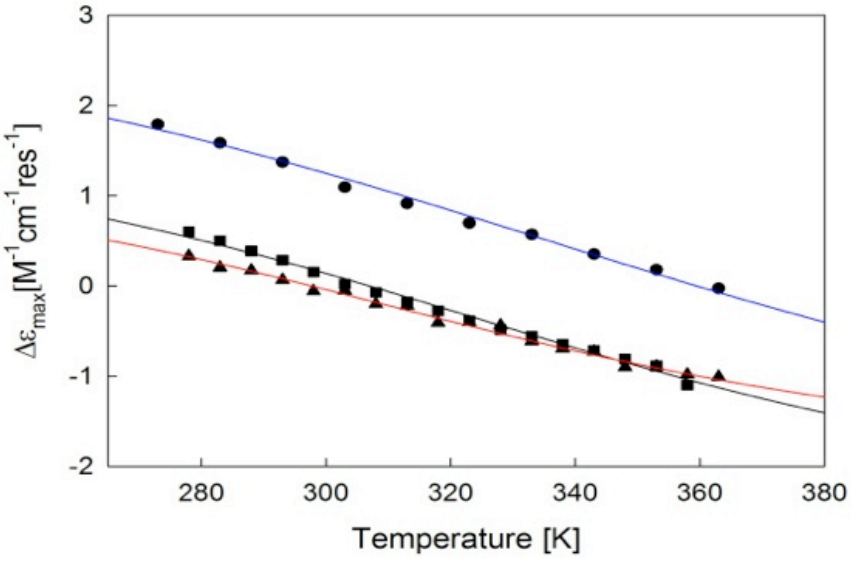

Figure 5.8: $\Delta\varepsilon_{max}$ of cationic AAA (circles), zwitterionic AAA (squares) and the AdP (triangles) measured at the maximum dichroism for each at room temperature (214nm, 219nm, and 215nm respectively) plotted as a function of temperature. The solid lines result from the fitting process described in the text.

Table 5.5: Intrinsic dichroism values, $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$, of cationic AAA, zwitterionic AAA and AdP obtained from the two-state thermodynamic fitting procedure described in the text. $\Delta\varepsilon$ values are in [$M^{-1}cm^{-1}res^{-1}$].

Peptide	$\Delta\varepsilon$(pPII)	$\Delta\varepsilon(\beta)$
cationic AAA	1.32	-1.55
zwitterionic AAA	0.84	-1.87
AdP	0.74	-2.06

5.2.4 MD reveals that the pPII content and hydration shell of AAA remains intact upon switching protonation states.

Contributions: MD work done in this section was primarily conducted by Derya Meral, under the guidance of Dr. Urbanc (Department of Physics, Drexel University), with analysis done in collaboration.

To investigate the ensemble differences of the three alanine-based peptides in atomistic detail we performed a series of all-atom MD simulations combining two of the currently available force fields (OPLS and AMBER03) with the three commonly used water models (TIP3P, SPCE, TIP4P). The AMBER03 force field was also used in combination with with the TIP4PEw water model. The decision to test multiple force-field/water models combinations stems from the poor reproduction of experimentally-obtained distributions for short peptides and unfolded proteins reported in many MD

studies (See *Chapter 1.2.5 of this thesis*). It is now well known that different force fields yield rather different conformational distributions, typically producing very low pPII propensities and overestimating the helical content, at variance with experimental results (95, 98, 100). In addition, the use of different water models for explicit solvation also leads to variable conformational preferences (115). Here, we chose to gauge which of the above mentioned force-field/water model combinations would predict conformational ensembles in the best agreement with experimental data, and then use this combination for a direct comparison of the two different alanine model systems.

In order to obtain conformational propensities, we defined the (ϕ,ψ) angles corresponding to the peak position for each major conformation (i.e., pPII-, β strand-, and helical-like) by first identifying the centers of each distribution in the MD-derived Ramachandran plots. The spread of each sub-population was then defined by ensuring that all respective conformations were included, similar to the method employed by to Gnanakaran and Garcia (105). Table 5.6 shows the resulting fractions of pPII, β-strand, and helical-like conformations sampled during all MD simulations. By comparing the results for different force-field/water model combinations, we noted that the OPLS force-field yielded the most accurate reproduction of the experimentally obtained conformational distributions and conformer statistical weights for cationic AAA. Not surprisingly, the fractions of pPII obtained with the OPLS force-field are still far below what we and others obtain experimentally. In addition, the helical content is far above what is expected for short peptides. This overestimation of the right-handed helical

content is characteristic of most MD simulations involving unfolded proteins or short peptides (101). Although the MD results are not in quantitative agreement with experiment, they are sufficient for our purposes of investigating the relative population differences and changes between the three alanine peptides as we are only interested in observing changes in populations.

Table 5.6: Fraction of pPII, β-strand and helical-like conformations obtained from MD simulations of cationic AAA, zwitterionic AAA, and AdP using the OPLS, Amber 03, and Amber 10 force fields with the TIP3P, TIP4P, and SPC/E explicit water models.

Force Field	Conformation Type	Cationic AAA				Zwitterionic AAA				Alanine Dipeptide			
		TIP3P	SPC/E	TIP4P	Tip4p-Ew	TIP3P	SPC/E	TIP4P	Tip4p-Ew	TIP3P	SPC/E	TIP4P	TIP4p-Ew
OPLS	pPII	0.6	0.67	0.62	--	0.54	0.53	0.57	--	0.44	0.47	0.45	--
	B-strand	0.12	0.11	0.15	--	0.13	0.12	0.15	--	0.12	0.11	0.14	--
	Helical-like	0.17	0.13	0.15	--	0.12	0.13	0.08	--	0.23	0.22	0.22	--
	remainder	0.105	0.09	0.06	--	0.21	0.22	0.2	--	0.11	0.2	0.19	--
Amber 03	pPII	0.58	0.61	0.59	0.39				0.39	0.37	0.41	0.34	0.39
	B-strand	0.22	0.19	0.24	0.15				0.15	0.16	0.16	0.17	0.15
	Helical-like	0.09	0.08	0.06	0.34				0.34	0.35	0.3	0.36	0.34
	remainder	0.11	0.12	0.11	0.12				0.12	0.12	0.13	0.13	0.12

Of all the water models under study, the SPC/E water model yielded the best agreement with experiment, i.e., the greatest sampling of pPII conformations for cationic AAA and AdP. Figure 5.9 shows the corresponding Ramachandran plots for cationic AAA, zwitterionic AAA, and AdP obtained using the OPLS force field with the SPC/E water model. All three plots clearly show a dominant sampling of the pPII conformations followed by β-strand-like (i.e., extended) conformations. Both protonation states of AAA had a relatively higher pPII fraction compared to AdP, again in agreement with our vibrational and thermodynamic results. It should be noted that MD simulations yielded a slightly higher pPII content for the cationic compared with the zwitterionic state of AAA. However, no increase in the β-strand or helical conformations was observed in the latter. Instead, the MD-derived Ramachandran plot in Figure 5.9 indicates a scattered population close to the bridge region between the β-strand and the right-handed helical region. An increased population in this region would cause a substantial increase of the $^{3}J(H^{N}H^{\alpha})$ in the zwitterionic state compared to the cationic state. Since we do not obtain this experimentally, we can conclude that this is an artifact of the MD simulation only.

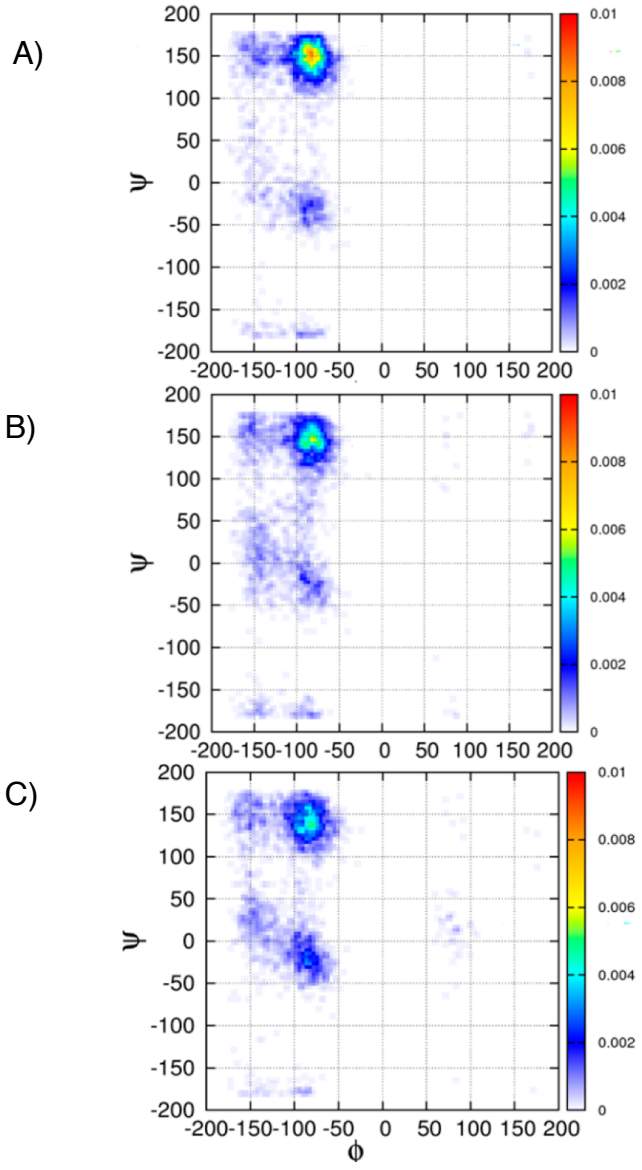

Figure 5.9. Ramachandran plots for (A) the cationic and (B) zwitterionic AAA and (C) AdP obtained by MD simulations using the OPLS force field and SPC/E water model.

To obtain information on the dynamics of the conformational ensemble, the time evolution of the dihedral angles ϕ and ψ was monitored (Figure 5.10) throughout the 50ns MD trajectory for cationic AAA and AdP. To calculate the lifetimes and hence the stability of each conformation, the time durations of the three conformations along the MD trajectory were extracted. Figure 5.11 shows the distributions of the time durations, N(t), for each of the three major conformations. The lifetime (τ) of each conformation was determined by fitting each curve with an exponential function. Notably, all time distributions shown in Figure 5.11 could be fit accurately with a single exponential function, except for the pPII distribution of AAA which required a bi-exponential fit. Table 5.7 shows the lifetimes for each major conformation sampled by AAA and AdP. In general, the pPII conformation persisted for the longest lifetime in both alanine-based peptides. For AAA, the bi-exponential fit yielded two average lifetimes of 15.8 ps and 181.8 ps. The two lifetimes likely reflect fluctuations of the water distribution within the hydration shell. For AdP, the pPII lifetime was 63.7 ps, approximately in-between the two lifetimes obtained for the same process in AAA. Although we did not observe the fast decay in the pPII conformation of AdP, this might have been due to the 20 ps time step used in the MD simulations. In both peptides, other than pPII, the helical conformation had the longest lifetime, followed by the β-strand conformations. Not surprisingly, the β/

pPII transitions were the most frequent for each peptide, in agreement with the largely two-state character of the obtained conformational ensembles.

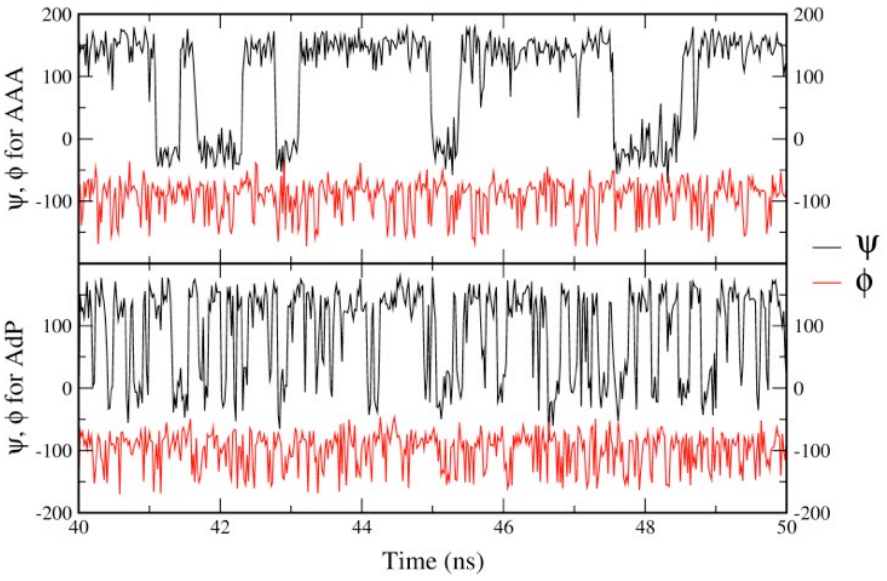

Figure 5.10: Time evolution of ϕ and ψ dihedral angles for AAA (top panel) and AdP (bottom panel) over the 40-50 ns time interval within the 50 ns MD trajectory.

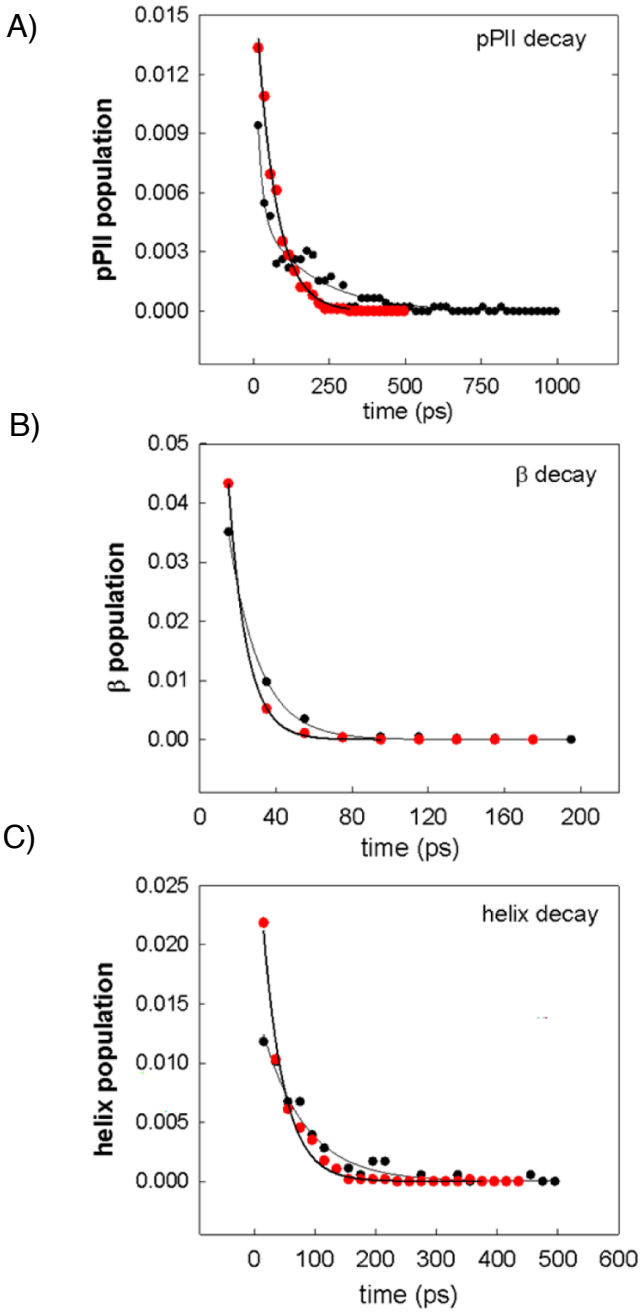

Figure 5.11. Distribution of durations, N(t), of the (A) pPII, (B) β-strand, and (C) helical conformations for cationic AAA (black circles) and AdP (red circles) derived by MD by counting the number of occurances of respective sub-poluations at each time interval. The solid line represents exponential fits.

It is noteworthy that the pPII distribution of duration times, $N_{pPII}(t)$, were dominated by the pPII→ β transition, as evidenced by the comparatively large effective rate (4.14×10^9 s^{-1} and 3.94×10^9 s^{-1} for Adp and AAA, respectively, Table 5.8). Similarly, the β decay was dominated by the β→pPII transition (4.0×10^9 s^{-1} and 4.10×10^9 s^{-1}, respectively). Such a fast exchange dynamics in cationic AAA has been obtained earlier by Mu and Stock (104). A comparison of the three lifetimes for AAA and AdP (Table 5.7), shows that all conformer lifetimes were significantly shorter for AdP. The large disparity between lifetimes of the three major conformations adopted by the two peptides would not necessarily be expected based soley on differences in conformational propensity. For instance, although the helical conformation had the lowest propensity for all peptides, it had a relatively long lifetime (70.4ps and 34.6 ps for AAA and AdP, respectively) as compared to the lifetime of β-strand (15.95 ps and 9.58 ps, respectively). This disparity of lifetimes between AAA and AdP and the stability of the three conformations can be explained by considering the role of the solvent in stabilization of pPII, β-strand, and helical conformations.

Table 5.7: Average lifetime (τ), and initial population (N_0), and R^2 parameters obtained from fitting duration distribution curves, N(t), for cationic AAA and AdP in each major conformation. Each curve was fit with a single-exponential function, except for the pPII curve of AAA which required a bi-exponential fit and 4 parameters.

	Conformation	N_0	τ (ps)	R^2
Trialanine	pPII	0.0084 / 0.0051	15.77 / 181.81	0.976
	β	0.0898	15.95	0.99
	Helix	0.0154	70.4	0.985
Alanine Dipeptide	pPII	0.0175	63.7	0.996
	β	0.2071	9.58	0.999
	Helix	0.0326	34.6	0.993

Table 5.8: The number of occurances, probability and effective rate constants for each transition type for AdP. The “bridge” region refers to the occupation in the area Φ<-125°, 50°>Ψ>-110°.

AdP Transitions	# occurances	probability	rate (s^{-1})
β →pPII	200	0.180	$4*10^9$
pPII → β	207	0.187	$4.14*10^9$
β →helix	47	0.042	$9.4*10^8$
helix → β	49	0.044	$9.8*10^8$
β →bridge	28	0.025	$5.6*10^8$
bridge → β	18	0.016	$3.6*10^8$
pPII →helix	126	0.114	$2.52*10^9$
helix →pPII	117	0.106	$2.34*10^9$
pPII →bridge	33	0.030	$6.6*10^8$
bridge →pPII	49	0.044	$9.8*10^8$
helix →bridge	119	0.108	$2.38*10^9$
bridge →helix	113	0.102	$2.26*10^9$

Table 5.9: The number of occurances, probability and effective rate constants for each transition type for AAA. The “bridge” region refers to the occupation in the area Φ<-125°, 50°>Ψ>-110°.

AAA Transitions	# occurances	probability	rate (s^{-1})
β →ppII	205	0.334	$4.1*10^9$
ppII → β	197	0.321	$3.94*10^9$
β →helix	3	0.005	$6*10^7$
helix → β	10	0.016	$2*10^8$
β →bridge	6	0.010	$1.2*10^8$
bridge → β	6	0.010	$1.2*10^8$
ppII →helix	20	0.033	$4*10^8$
helix →ppII	10	0.016	$2*10^8$
ppII →bridge	10	0.016	$2*10^8$
bridge →ppII	13	0.021	$2.6*10^8$
helix →bridge	68	0.111	$1.36*10^9$
bridge→helix	65	0.106	$1.3*10^9$

To more closely investigate the solvation of the three alanine peptides, we calculated the radial pair distribution functions g(r) between the amide proton of the central residue and water hydrogen and oxygen for AAA and AdP. Interestingly, for both protonation states of AAA, the g(r) plots are practically overlapped. Figure 5.12 shows the radial distribution functions for cationic AAA and AdP. For the sake of clarity, we omit here the corresponding g(r) plots for zwitterionic AAA as these were near identical to cationic trialanine. Most of the water oxygen atoms were at the hydrogen bonding distance (approximately 1.7 Å) for both protonation states of AAA. In addition, there is a rather intense second maxima in the g(r) for the water oxygen observed at approximately 3.2 Å, reflecting some degree of water ordering, resulting in a pronounced second hydration shell around the central amide atom of AAA. Again, we did not observe any significant differences between g(r) curves of protonated and zwitterionic AAA, indicating that the hydration shells remained intact upon switching the protonation state. For AdP the H_2O-HN distance with the highest water density was increased to about 2Å and is noticeably less pronounced (by a factor of 3), suggesting a relatively limited hydration of, and weaker hydrogen bonding to, the alanine residues in blocked peptides. This indeed would affect the propensity of the central alanine residue, specifically decreasing the pPII preference for AdP, in agreement with our experimental results. In addition, and perhaps more importantly, the second hydration shell present in AAA was not observed in the dipeptide analogue. The decreased density of water around AdP and the absence of the second hydration shell indicate a much less ordered solvent structure in

AdP (relative to AAA). This more disordered solvent structure around AdP was also reflected in a broader distribution of the distance between the central C_β atom and the C-terminal amide nitrogen atom (Figure 5.12C), which had additional peaks at larger distances relative to AAA (Figure 5.12B). The highly ordered solvent structure around AAA and the increased H-bonding capacity can be thought of as effectively increasing the activation barrier between conformations, which indeed explains the aforementioned longer conformational lifetimes obtained for AAA.

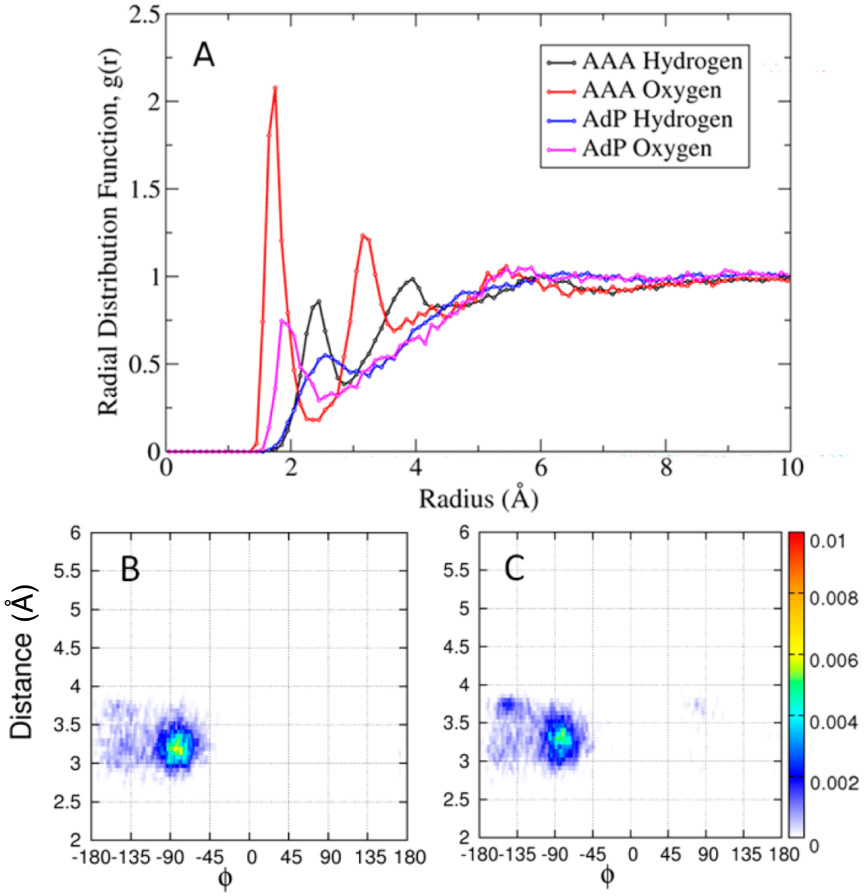

Figure 5.12. Radial distribution functions, g(r), of water molecules (using H- and O-atoms of water) around the amide proton of the central residue of cationic AAA and AdP derived by MD. Distributions of the (B) cationic AAA and (C) AdP conformations with respect to the dihedral angle ϕ and the distance between the nitrogen atom of the third residue and the side-chain atom C_β of the central residue in AAA and the corresponding distance in AdP.

5.3 CONCLUSIONS

Taken together, our experimental and MD data show no substantial decrease of the pPII population upon deprotonation of trialanine's terminal groups, in contrast to what was recently reported in literature. Conformational distributions were obtained from a global analysis of amide I' bands of IR, polarized Raman, and vibrational circular dichroism spectra and a set of 6 different J-coupling constants. Our combined simulation of these amide I' profiles and J-coupling constants reveals that the conformational distribution of the central residue in AAA predominantly adopts the pPII conformation (χ_{pPII}=0.84), followed by the β-strand conformation (χ_{β}=0.08), along with small admixtures of right hand helical-like (χ_{α}=0.04) and ɣ turn-like conformations (χ_{γ}=0.04), in all protonation states. The proximity of the end groups did not seem to affect this high pPII preference of alanine. Remarkably, the entire conformational distribution, defined by a superposition of Gaussian functions representing the maxima and widths of each sub-state in (ϕ, ψ) space remain quantitatively similar in all protonation states of the unblocked tripeptide. In contrast, the pPII fraction of the AdP is slightly lower than what is observed for AAA in all protonation states. Thermodynamic analysis of the alanine-based peptides reveals that the free energy landscape of the pPII-β equilibrium as well as the enthalpic stabilization of the pPII conformation is nearly invariant to the terminal charge. Interestingly, the conformational ensemble of the alanine dipeptide resembles closely that of the unblocked GAG model peptide with a pPII propensity of χ_{pPII}=0.74. Investigation of valine and aspartic acid dipeptides in comparison to their GxG

counterparts further reveals that the conformation of the target amino acid is the same in both model systems regardless of the dominant conformational ensemble of the guest amino acid. These results indicate that the end groups beyond neighboring Cα moieties do not have an experimentally significant influence on the intrinsic conformation of non-terminal residues in peptides.

Results of MD simulations on cationic and zwitterionic AAA also show that these peptides share similar Ramachandran plots. Both protonation states of AAA have a relatively higher pPII content as compared to the AdP, corroborating our experimental findings. In addition, radial distribution functions derived by MD simulations indicate that the amide proton of the central residue in AAA is on average in much closer proximity to water molecules, which may explain the higher enthalpic gains and stabilization of the pPII conformation for AAA relative to the AdP.

We have shown that the hydration shell around AdP is much less ordered than around AAA. Consequently, the pPII preference in AdP is diminished relative to that of AAA, effectively decreasing the activation barrier of the pPII-β transition. Indeed, favorable backbone-solvent interactions in aqueous solution have been cited numerous times in literature as the primary source for stabilization of the pPII conformation (72, 75, 76, 93, 119, 194) (see *Chapter 1.2.6 of this thesis*). We thus hypothesize that the choice of water model in MD simulations is instrumental for the reproduction of the experimentally-observed conformational propensities in small peptides. In particular, polarizable water models and force field may prove critical in this regard (96).

Recent efforts directed towards defining a pPII propensity scale in various host-guest peptides are part of the broader goal to establish a physical basis for the experimentally observed pPII preference in unfolded states. As such, this study was directed toward determining whether there was any appreciable difference in the conformational ensemble upon protonation/deprotonation of terminal groups in short unblocked peptides. The findings reported herein for AAA do not indicate any difference in this regard. Further, since the blocked forms of amino acids show remarkable similarity to the unblocked GxG peptides, we conclude that these are both suitable model systems for investigating unfolded states, conformational preference, and the pPII bias specifically.

CHAPTER 6. SOLVATION STUDIES

Reproduced in part, from:

I. Toal, S., Verbaro, D., Schweitzer-Stenner, R. "The Role of Enthalpy-Entropy Compensation in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides." J. Phys. Chem. B., 2014, 118, 1309-1318

II. Toal, S., Amidi, O., Measey, T., Schweitzer-Stenner, R. "Conformational Changes of Trialanine Induced by Direct Interactions between Residues and Alcohols in Binary Mixtures of Water with Glycerol and Ethanol." J. Am. Chem. Soc., 2011, 133, 12728–12739

6.1 Thermodynamics of GxG Peptides Points to the Role of Solvation for Governing Amino Acid Biases in the Unfolded state

6.1.1 Introduction

Despite the increasing relevance of characterizing local conformational ensembles in the unfolded state, the role that solvation plays in governing and modifying ensembles remains unclear. As discussed earlier (see *Chapter 1 of this thesis*), our studies (52, 54, 55, 202) along with others (49, 51, 132, 133) have shown that contrary to random coil theories, distribution functions obtained for the amino acid residues in the unfolded state are far more restricted and unique to each residue. Specifically, distributions are often bimodal in the upper left quadrant and allow one to discriminate between sub-distributions of pPII and β-strand like conformations. Generally, the combined conformational space of pPII and β-strand was found to be much more restricted than a typical random coil distribution in the upper left quadrant in the Ramachandran plot.

We recently reported detailed conformational analyses of GxG model peptides in water based on a global analysis of amide I' bands of IR, Raman, and vibrational circular dichroism spectra as well as of 2D NMR J coupling parameters (54, 55, 141, 198). Thus, we could derive residue-level information from each spectroscopic technique and combine them with statistical models that describe the entire conformational manifold of guest residues in the peptides. The obtained results allow us to sub-divide the amino acid residues investigated here into categories according to their pPII propensity. Alanine was found to stand out with an exceptionally high intrinsic pPII propensity of 0.72. Category number two contains residues, such as L and K, with a modest dominance of pPII over a substantial β-strand fraction. The third category contains branched aliphatic and aromatic groups (V, I, F), which display comparatively higher β strand and turn propensities. Surprisingly, we found a fourth category of peptides which contains residues with short polar and/or ionizable central residues (D,T, N, C) and a higher-than-average content of turn-like conformations.

The physical origin behind this restricted conformational sampling of the Ramachandran space by practically all amino acid residues in general, and of their different conformational distributions in particular, remains largely unclear. Most theories, however, and a limited number of experiments emphasize the role of water solvation particularly for the stabilization of pPII, while they still disagree on the underlying mechanism. As outlined in *Chapter 1 of this thesis*, these theories suggest a

variety of causes encompassing direct hydrogen bond bridge formation, electrostatic and steric effects, and optimal water packing in the hydration shell (76, 93, 119).

Many of the attempts to elucidate the conformational biases of amino acid residues generally focus on Gibbs free energy differences between populated conformations (120, 121). However, the reliability of any thermodynamic model should be checked with regard to its capability to rationalize concomitant changes in enthalpy (internal energy) and entropy. For instance, it is well documented that pPII conformations are stabilized enthalpically, while more extended β-strands are entropically favored (203). This indicates at least a partial enthalpy-entropy compensation for the equilibrium between these conformers. Early theories on structural transitions of biomolecules suggest that direct and indirect solute - solvent (water) interactions could bring about a nearly exact thermodynamic compensation of very large enthalpic and entropic changes associated with conformational transitions of biomolecules (i.e. between pPII and β in the case of amino acid residues), thus yielding very low Gibbs energies (204-207). Hence, all the important thermodynamic information which should provide physical insight into why certain conformational propensities are stabilized becomes lost within minor ΔG changes.

This study was aimed at elucidating the individual thermodynamic contributions underlying conformational propensities in order to shed light on the factors driving local order in the unfolded state. If a common thermodynamic cause for the different propensities of amino acid residues exists, enthalpy and entropy should be nearly

perfectly linear, thus reflecting a very narrow distribution of compensation temperatures at which a substantial part of the total Gibbs energy disappears due to enthalpy-entropy cancelation (207). In order to check whether this scenario applies to individual amino acid residues in unfolded pepitdes, we determined the enthalpic and entropic contributions to the Gibbs energy landscape of GxG model peptides in aqueous solution by a global analysis of their temperature dependent UVCD and HNMR spectra. The analysis was based on the aforementioned set of experimentally determined conformational distributions, which we recently obtained for these peptides at room temperature using 2D NMR and vibrational spectroscopy (54, 55, 141, 198). We herein identify a common nearly exact enthalpic-entropic compensation effect (208), which for a majority of amino acid residues in aqueous solution causes an iso-equilibrium at near physiological temperatures (207). This is a rarely observed relationship which shows that these amino acid residues share a common compensation temperature at which the solvation Gibbs energy is zero. Only three amino acid residues, A, D and T, deviate from this picture, most likely due to additional solvent peptide and intrapeptide interactions. We propose the average compensation temperature of the investigated set of peptides as a measure of the goodness of the solvent which determines its capability to stabilize a statistical coil over partially or completely folded states of proteins.

6.1.2 Results

6.1.2.1 Thermodynamics of Conformational Preference

To characterize conformational ensembles of each amino acid within the model GxG system in thermodynamic terms, we analyzed the temperature dependence of the respective UVCD and ^{1}H NMR spectra by using the method outlined in *Chapter 4 of this thesis*. The thermodynamic analyses of non-alanine GxG peptides is complicated by the presence of turn-like conformations, which for aspartic acid constitute 32% of the conformational ensemble (55). Such a large fraction of turn conformations necessitates an extension of the simplified two-state pPII-β model used for alanine based peptides. Recently, however, Rybka et al. showed that these turn-like conformations are generally thermostable (141). If this notion applies to all residues, a pseudo two-state thermodynamic model with temperature independent fractions of turn conformations is a suitable approximation. The formalism of the pseudo two state thermodynamic model is outlined in detail in *Chapter 4 of this thesis.*

To check the validity of a pseudo two-state model, we first measured the far UVCD spectra as a function of temperature for all investigated GxG peptides. These spectra are exhibited in Figure 6.1. At low temperatures, spectra of all peptides display a negatively biased couplet with the negative maximum at approximately 195 nm, which is diagnostic of a substantial sampling of pPII conformations (58, 64). With increasing temperature, the dichroism at 215 nm ($\Delta\varepsilon_{215}$) and the negative maximum ($\Delta\varepsilon_{195}$) decrease, indicating a conformational redistribution towards more β-strand-like

conformations (186). The pPII content as indicated by $\Delta\varepsilon_{195}$ and $\Delta\varepsilon_{215}$ differs for each peptide at each temperature, confirming the notion that each amino acid residue has distinct conformational preferences in the unfolded state. The pPII preference for each residue at temperatures lower than or equal to room temperature qualitatively agrees with those explicitly derived by Hagarman et al. (54, 55), i.e. the pPII preference decreases in the following order: A, M> L, E, R, K> >I, V, S> D,N, T. Amino acids with high turn propensities (D, N, T, S) additionally exhibit a shallow negative maximum at approximately 220 nm. However, the temperature dependence of this marker is modest, and indeed, all CD spectra exhibit a clear isodichroic point (≈ 204 nm) confirming that the conformational ensemble at all investigated temperatures is dominated by two states (pPII and β) and a pseudo two-state model considering only exchanges between pPII and β-strand is a suitable approach.

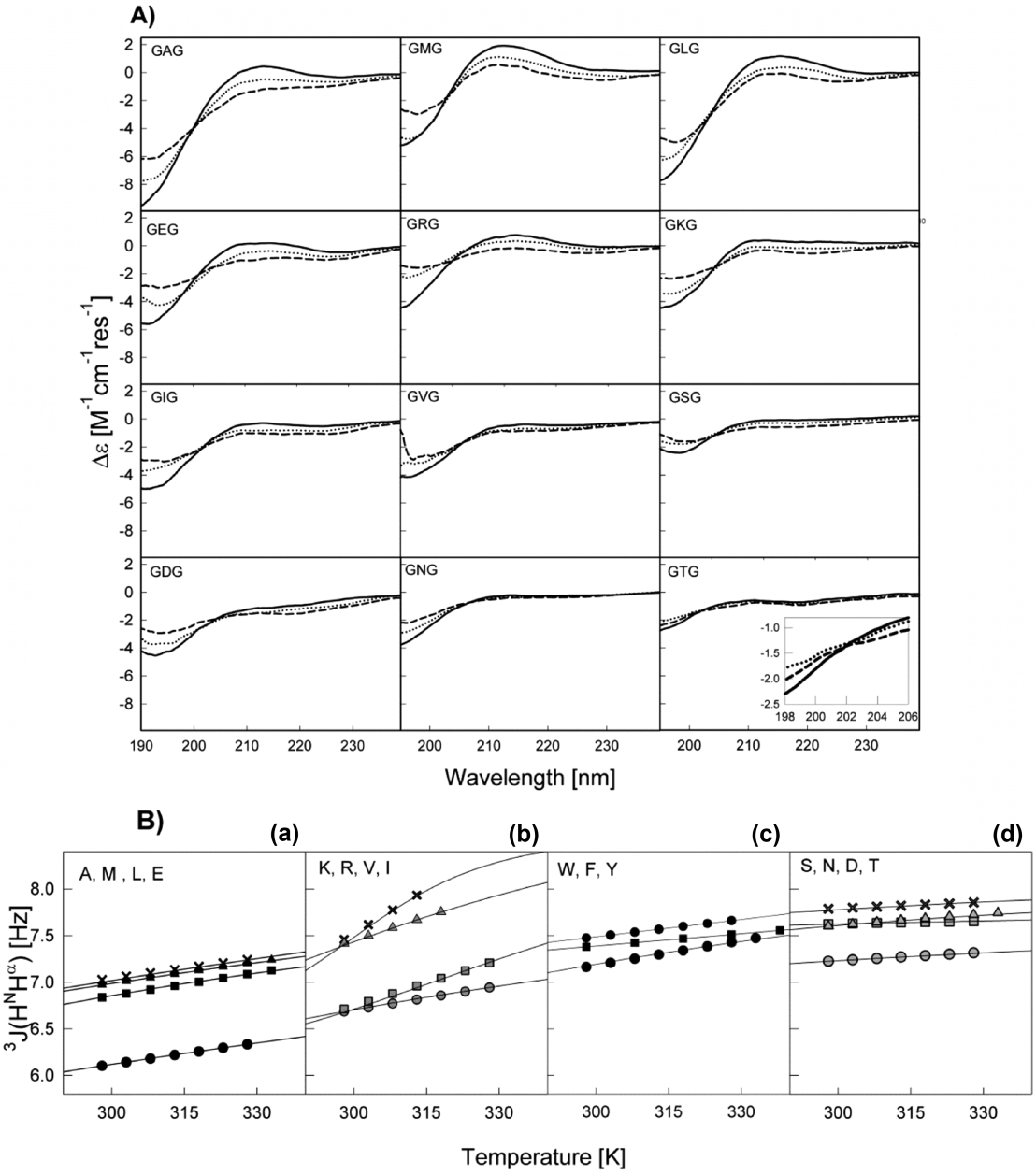

Figure 6.1: (A) UV-CD spectra as a function of temperature for GAG, GKG, GVG, and GDG. For clarity, only spectra taken at 10 °C (solid line), 50 °C (dotted line), and 90 °C (dashed line) are shown. (B) $^{3}J(H^{N},H^{\alpha})$ [Hz] coupling constants as a function of temperature for all amino acid residues. The solid lines result from the thermodynamic model described in the text. Panel a displays alanine (circles), methionine (squares), leucine (triangles), glutamic acid (crosses). Panel b displays lysine (circles), arginine (squares), valine (triangles), isoleucine (crosses). Panel c displays tryptophan (circles), phenylalanine (squares), tyrosine (triangles). Panel d displays serine (circles), asparagine (squares), valine (triangles), isoleucine (crosses).

To gain site-specific information, the $^{3}J(H^{N}H^{\alpha})$ coupling constants were measured as a function of temperature. These constants reflect the average ϕ value of the central x-residue according to the Karplus relationship (163) and are thus useful observables for assessing residue-level conformational populations. In line with the above results, we constructed a simple pseudo-two state thermodynamic model that considers 1) redistributions among extended structures (pPII-β) as temperature dependent and 2) temperature independent turn populations as outlined in *Ch 4 of this thesis*. Using this model we obtained very good fits to the experimental data (Figure 6.1B). As shown in Figure 6.1B, the $^{3}J(H^{N},H^{\alpha})$ constants for all residues increases with temperature, indicating a conformational re-distribution from pPII to β, as reported in the literature (51). Amino acid residues with relatively high pPII content at room temperature (A, M, L, E) display modest changes in $^{3}J(H^{N}H^{\alpha})$ with temperature. In contrast, for amino acid residues that have a more balanced pPII-β distribution (R, V, I) we observed larger changes in the ^{3}J-value with temperature (Figure 6.1B). The β-branched residues valine and isoleucine in particular shift rapidly towards more extended β-strand conformations upon heating. Notably, amino acids that have been previously shown by us to contain uniquely high turn-like content (S, N, D, T) (55, 141), exhibit rather weak temperature dependencies of ^{3}J values (Figure 6.1B). This may be a direct result of their relatively large intrinsic turn population, which according to Rybka et al would provide a temperature insensitive contribution to ^{3}J (141).

The thermodynamic values ΔH and ΔS obtained from our fits, and the corresponding room temperature ΔG values are displayed in Figure 6.2. Immediately, one can see from the wide distribution of thermodynamic values that amino acids substantially differ substantially with regard to their enthalpic and entropic contributions to the pPII/β equilibrium. Generally, our data indicate that enthalpic and the entropic contributions far exceed the Gibbs energy at room temperature, which is diagnostic of a near-exact enthalpy-entropy compensation (207).

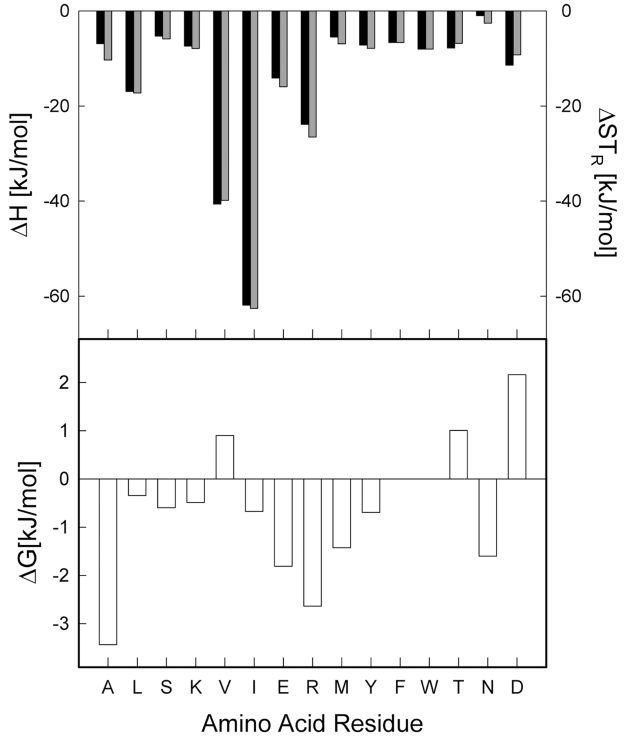

Figure 6.2. ΔH (gray bars) and T ΔS (black bars) values (upper panel) and ΔG (lower panel) obtained for the indicated residues (upper Panel).

6.1.2.2 Enthalpy-Entropy Compensation

A closer look at Figure 6.2 reveals that, not only are the ΔH and ΔS values of different residues vastly dissimilar, but in addition, individual ΔH/ΔS values for each amino acid residue compensate each other to an extent which minimizes the Gibbs free energy for all residues. A balance between entropic losses and enthalpic gains upon structure formation is ubiquitous to protein folding processes and all types of interactions involving weak, van der Waals governed forces (205). However, exact enthalpy-entropy compensation (in contrast to a mere correlation) is an interesting phenomena *per se* and implies that $\Delta H = T_c \Delta S$ and hence $\Delta G = 0$ at $T = T_c$. T_c is termed the compensation temperature. With regard to the GxG peptides investigated in this study, T_c could then be identified with the average transition temperature for the pPII↔β equilibrium. Alternatively, only a fraction of the Gibbs energy could be subject to compensation, in this case (204):

$$\Delta H = \alpha + T \Delta S \qquad \text{(eq. 6.1)}$$

Therefore, an exact statistically significant linear fit to ΔH/ΔS values indicates that all investigated processes can be related to exactly the same compensation temperature T_c and Gibbs free energy, whereas statistically significant data scattering reflects a distribution of similar compensation temperatures. To determine the category of enthalpy-entropy compensation reflected by our thermodynamic data, we plotted the enthalpy and entropy differences between pPII and β for all amino acid residues investigated. As

shown in Figure 6.3, we obtained a nearly perfect linear relationship when fitting the data with eq. (6.1). The fit yielded a T_c-value of 295±2 K and an α-value of -0.77kJ/mol. The high correlation coefficient of 0.98 indicates that the compensation temperatures for the individual GxGs should be very similar. These values will be considered further in the discussion.

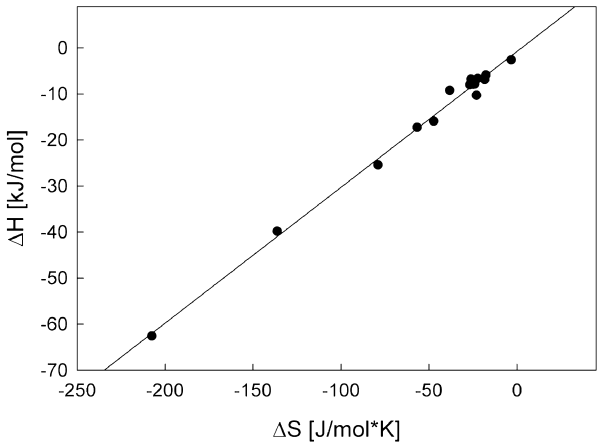

Figure 6.3. Plot of ΔH versus ΔS values obtained from a thermodynamic analysis of $^3J(H^NH^\alpha)$ (T) data of all investigated amino acid residues. The solid line results from the linear least-squares fit.

6.1.2.3 Iso-equilibrium Points

Technically, it is difficult to establish a common compensation temperature for a set of correlating ΔH, ΔS pairs owing to their statistical uncertainties, which *per se* rule out a perfect correlation. The high value obtained for the correlation coefficient in Figure 6.3 suggests that our data pairs might be sufficiently accurate and precise to achieve this goal. In this case, we should be able to identify a temperature at which at least a substantial number of the investigated amino acid residues exhibit nearly the same Gibbs energy difference between pPII and β (207). In order to determine whether such an iso-equilibrium point exist for the amino acid residues, we calculated the mole fraction ratio of pPII and β for all peptides as a function of temperature which is displayed in Figure 6.4. We indeed found that there are two closely spaced iso-equilibria representing two sub-sets of amino acid residues.

The first iso-equilibrium cluster (IE1) occurs at approximately 302K and contains the amino acid residues L, V, I, S, K, Y, W, and F; the second iso-equilibrium point (IE2) occurs at a slightly higher temperature (approximately 312K) and contains E, R, M, N (Figure 6.4 A and B, respectively). Notably, IE1 contains mostly amino acids with short bulky side chains (L, I, V) as well as the aromatics (Y, W, F); with serine as an exception. This equilibrium point is characterized by nearly equal mole fractions of pPII and β ($\chi_{\mathrm{pPII,\ IE1}}/\chi_{\beta,\ \mathrm{IE1}}\approx 1$) and hence indicates a rather iso-energetic free energy landscapes in the upper left quadrant of the Ramachandran plot at the corresponding temperature

(302K). IE2 is characterized by $\chi_{pPII,\,IE2}/\chi_{\beta,IE2}\approx 2$, indicating a much stronger bias for pPII for these amino acids.

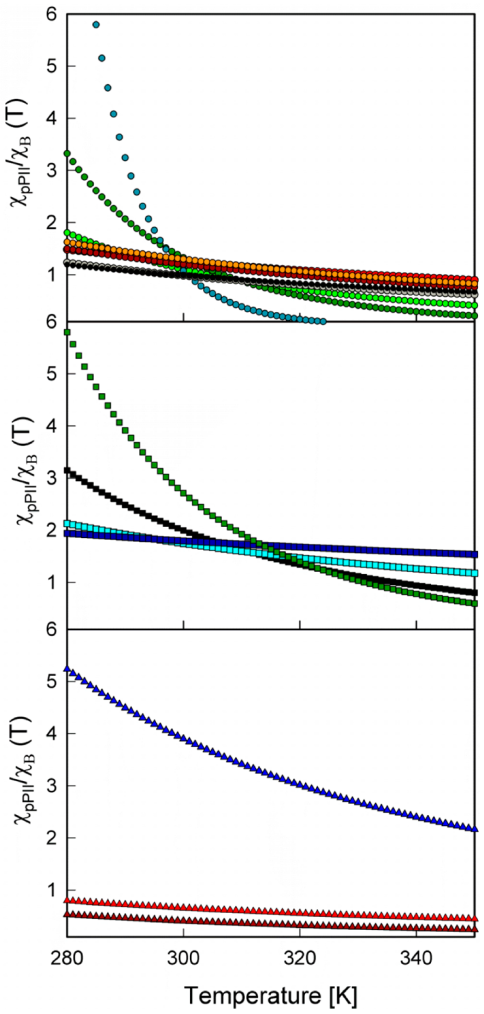

Figure 6.4. Temperature dependence of χ_{pPII}/χ_β for all investigated amino acid residues. (upper panel) The first iso-equilibrium cluster (IE1) containing amino acid residues I (turquoise), V (green), L (light green), Y (orange), S (red), W (gray), F (black). (middle panel) The second iso-equibrium cluster (IE2) containing R (green), E (black), M (aqua), and N (blue) residues. (lower panel) Remaining χ_{pPII}/χ_β for alanine (blue), aspartic acid (dark red), and threonine (red).

In order to determine the average compensation temperature for these two subset we plotted the respective ΔH and ΔS values separately and subjected them to a linear regression. The resulting ΔH and ΔS relationships for each subset are shown in Figure 6.5 and exhibit nearly perfect linearity for both fits. The respective compensation temperatures obtained from these fits are T_{c1}=297±4 K and T_{c2}=305±3K, the respective axis values are α_1= -0.3±0.3 kJ/mol and α_2=1.4±0.2 kJ/mol. The corresponding correlation coefficients are 0.999.

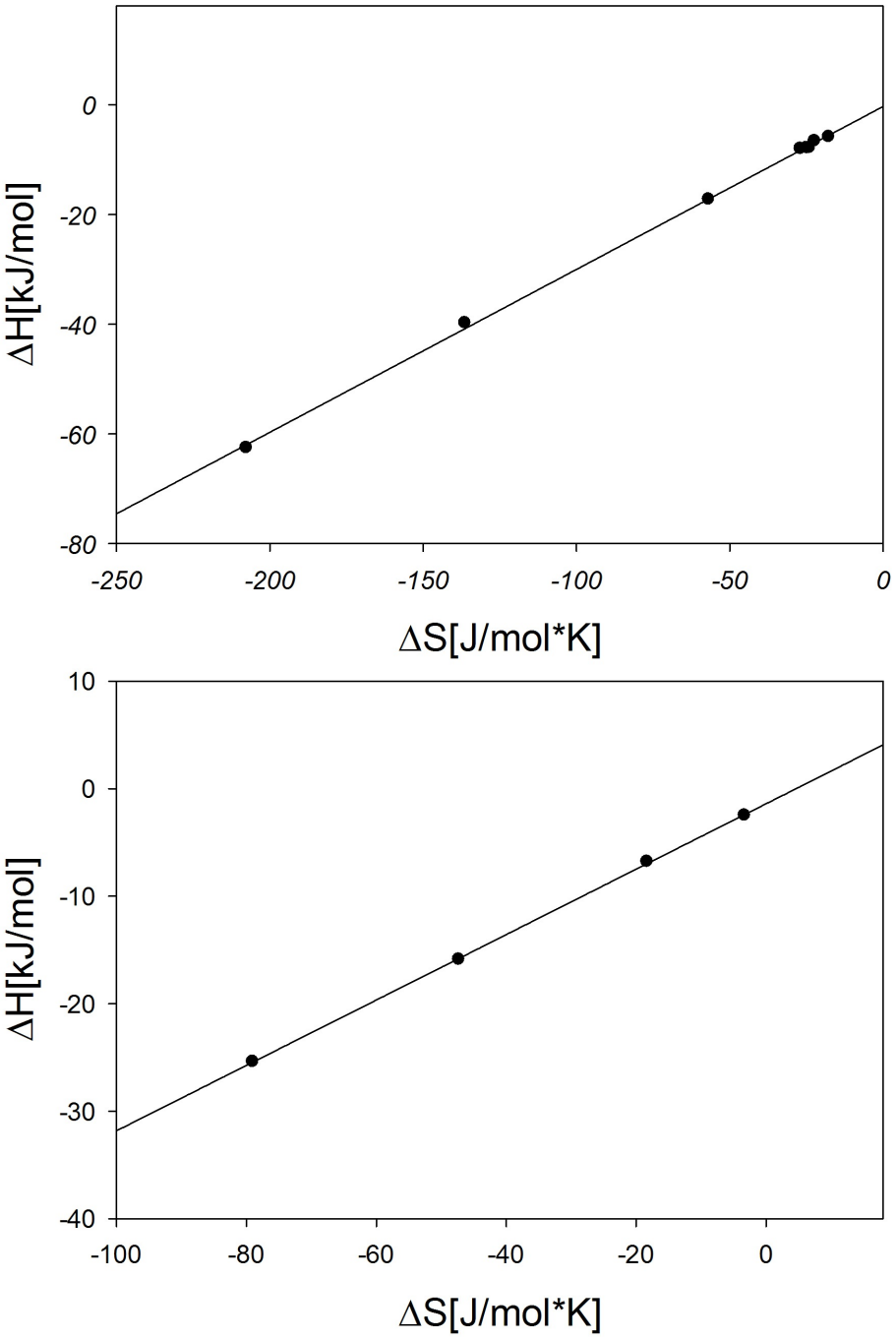

Figure 6.5: Enthalpy - Entropy correlation for amino acid residues contributing the isoequlibria 1 (upper figure) and 2 (lower figure). The solid lines are the result of linear regression.

When these two isoequilbrium clusters are removed from analysis, we are left with only three amino acid residues, namely, alanine, aspartic acid, and threonine (Figure 6.4C). These real compensation temperatures differ from the lower T_c-value obtained from the fit to eq. 6.1. Only in the case of a perfect fit (correlation coefficient of 1) would this T_c-value coincide with the temperature associated with a single iso-equilibrium point.

6.1.3 Discussion

6.1.3.1 The Existence of Statistically Significant ΔH-ΔS Compensation

The combined iso-equilibria and near-exact ΔH-ΔS compensation among the investigated GxG peptides implies that there is a common mechanistic effect behind conformational preference in the unfolded state. Enthalpy-entropy compensation effects have been reported in a wide array of different protein folding and binding phenomena (205, 207, 209). However, the statistical accuracy, the existence itself of 'extra-thermodynamic' enthalpy-entropy compensation effects, and the physical basis for such effects has been a matter of debate in a range of literature (207, 210). It can be shown that if the compensation temperature is close to the experimental temperature, that compensation effects may be due to error propagation (207). The compensation temperature obtained in this study (295.2K) is lower than the range of experimental temperatures (298-328K). Krug et. al presented an analysis of such compensations and

derived tests for distinguishing between extra-thermodynamic factors and statistical compensation (210). To assign a compensation effect as likely occurring from extra-thermodynamic factors, one has to compare the slope (β) with the harmonic mean of experimental temperatures (T_{hm}). Krug et al. additionally showed that statistical analysis allows one to determine an approximate 95% confidence interval for the slope β. In general, if the compensetion slope (T_c) falls within this range, one may not rule out statistical compensation due to error propagation. Following this protocol we obtain the 95% confidence limit of β to be (295 K, 63.3 K). Comparing this to the slope resulting from linear regression of ΔH/ΔS values (Tc=295.2K), we see that the compensation temperature lies just outside the error interval on the upper boundary of β. Therefore, this analysis suggests that the obtained ΔH/ΔS compensation is likely not a statistical artifact. In this context, it is important to also note that the critical assessment of linear enthalpy-entropy relationships provided by Krug et al. is based on van't Hoff and Arrhenius plots, which indeed are prone to strong statistical correlations between ΔH and ΔS, however this is not how the thermodynamics were obtained.

Recently, Liu and Gao (207) showed that if the experimental errors along ΔH and ΔS plane are significantly less than the range of the experimental values, the data are indicative of a real compensation effect. In order to determine these errors we calculated the error of the ^3J (T) values for each GxG peptide according to standard propagation of error techniques. The error in each ^3J series is determined by:

$$\delta J = \sqrt{(\delta m * T)^2 + \delta b^2 + (\delta n * T)^2 + \delta c^2} \qquad \text{(eq. 6.2)}$$

where δm and δn are the standard errors for the linear slopes of the higher and lower frequency peaks of the amide proton doublet, and δb and δc are the standard errors associated with the respective intercepts. We then calculated the error in ΔH and ΔS by obtaining the upper and lower error limits for each ΔH/ΔS pair by re-fitting the "error propagated ^{3}J(T)" data. For illustration, Figure 6.6 shows the experimentally obtained ^{3}J (T) with error bars as well as the "error propagated ^{3}J(T)" used to determine δΔH/δΔS errors for GAG and GIG. Thus, we were able to obtain δΔH and δΔS values for each amino acid residue (Table 6.1). As shown below, the largest error is seen for isoleucine with δΔH=18kJ/mol and δΔS=20.7 J*mol/K which is still well within the range of experimental values. Hence, we can consider our ΔH/ΔS compensation to be statistically significant.

In addition, the iso-equilbria we obtained indicates that at least a large subset of the investigated amino acid residues share a common transition temperature, and hence the ΔH/ΔS compensation is real and justified. Although one expects, to a certain degree, that enthalpy and entropy will be opposing driving forces in conformational transitions, an exact linear compensation for similar, though not identical, thermodynamic processes (i.e. in our case the pPII/β equilibrium of different residues) is rarely observed. If, for instance, compensation temperatures for related processes were slightly different, the ΔH and ΔS values would not be exactly reproduced by the linear fit, they may still show significant correlation. The obtained T_c-value in this case would then be considered as

representative of a set of slightly different underlying compensation temperatures. Therefore, one may say that the better the correlation obtained, the more similar are the respective compensation temperatures and the more likely it is that a common mechanism is operative.

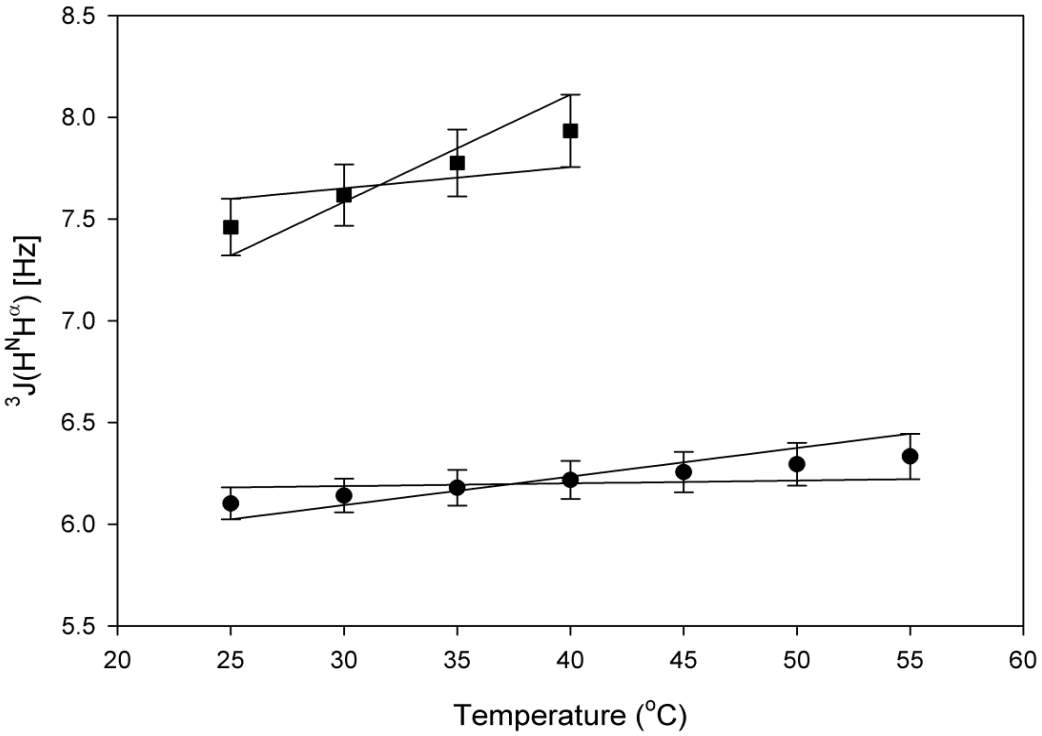

Figure 6.6: $^3J(H^NH^a)$ and associated error as a function of temperature. The solid lines reflect the "error propagated $^3J(T)$" values used to determine $\delta\Delta H$ and $\delta\Delta S$ values.

Table 6.1: Thermodynamic errors, δΔH and δΔS for each investigated peptide.

Peptide	**δΔH kJ/mol**	**δΔS J/mol K**
GAG	1.0	3.4
GLG	0.96	3.1
GSG	1.2	3.7
GKG	0.91	2.96
GVG	8.1	12.2
GIG	18	20.7
GEG	1.23	3.01
GRG	2.1	3.2
GMG	3.64	12.2
GYG	0.82	2.5
GFG	1.2	4.5
GWG	0.96	3.5
GTG	5.2	3.6
GNG	4.1	4.8
GDG	3.9	2.9

6.1.3.2 ΔH and ΔS are Varied Among Residues

The respective ΔH and ΔS values of those amino acid residues exhibiting near exact compensation behavior are vastly different. For instance, the enthalpic stabilization of pPII for alanine is rather modest, but sufficient to outweigh the entropic losses upon conformational ordering to the more compact pPII structure. Surprisingly, the respective enthalpic gain of -10 kJ/mol produced by a β→pPII transition in alanine is substantially lower than those of other amino acid residues such as valine (-35kJ/mol) and isoleucine (-70kJ/mol, which all have a preference for β. However, as shown in Figure 6.2, the respective entropies for these residues favor β conformations to such an extent that they overcome (most of the) enthalpic stabilization of pPII even at room temperature. For valine, the entropic stabilization of β-strand like structures even overcomes enthalpic stabilization conferred in pPII structures at room temperature. The remaining residues all exhibit comparable enthalpic differences between pPII and β with an average enthalpy of approximately 5kJ/mol. The pPII/β equilibria of valine and isoleucine, two aliphatic amino acids with C_β branched side chains, involve particularly large ΔH/ΔS values compared to non-branched leucine. This result is surprising and will be addressed in further detail below.

6.1.3.3 Decomposing ΔG to Reflect Solvent Mediation

The near-exact ΔH/ΔS compensation as well as the derived iso-equilibria both point to a common mechanistic effect driving conformational propensities in the unfolded state. The physical origin behind exact thermodynamic compensation is a matter of very multifaceted debates, but it is clear that protein/peptide hydration and solvent reorganization play a key role (204-208, 211, 212). For instance, Lumry and Rajender described a two- state hydration model in which the reaction/transition of a protein is coupled to a reaction/transition of water (204). In this context, the compensation behavior observed here for pPII↔β transitions would be rationalized as a result of direct interaction of hydration shell water with the peptide. Grunwald and Steel proposed a similar theory in which compensation is linked to changes in solvent reorganization – distinguishing between environmental processes in the solvent and nominal reactions of the solute (206). At the real compensation temperature the free energy change of the environmental process is zero and hence the ΔG associated with the reaction is solely due to the nominal process. To interpret our data we develop a similar concept in that we first assume the total Gibbs energy follows the following balance:

$$\Delta G_{\beta} = \Delta G_{\beta_c} + \Delta G_{\beta_{PS}} + \Delta G_{\beta_S} \qquad \text{(eq. 6.3)}$$

where $\Delta G_{\beta c}$ is the conformational Gibbs energy difference *in vacuo* in the absence of intramolecular hydrogen bonding, $\Delta G_{\beta PS}$ is the peptide solvation Gibbs energy, which encompasses backbone and side chain hydration. $\Delta G_{\beta S}$ reflects relaxation processes in the hydration shell which are caused by pPII↔β transitions.

In line with our data, environmental (solvent processes) terms are subject to total enthalpy-entropy compensation at certain compensation temperatures $T_{c,PS}$ and $T_{c,S}$. Therefore, we can formulate eq.(6.3) in more explicit terms as:

$$\Delta G_{\beta} = \Delta G_{\beta_C} + \Delta S_{\beta_{PS}} \cdot \left(T_{c_{PS}} - T\right) + \Delta S_{\beta_S} \cdot \left(T_{c_S} - T\right) \qquad \text{(eq. 6.4)}$$

where $\Delta S_{\beta PS}$ and $\Delta S_{\beta S}$ are the entropy differences of peptide-solvent interactions, and changes in the configuration of the hydration shell, respectively, which are caused by the increase of the side chains solvent accessible surface upon pPII $\rightarrow\beta$ conformational transitions (120). Ben-Naim showed that the purely conformational enthalpies and entropies $\Delta H°_{\beta}$ and $\Delta S°_{\beta}$ do not compensate and that $\Delta H_{\beta c} \gg T\Delta S_{\beta c}$ at room temperature (212). Following Grunwald and Steel(206), Ben Naim (212), and Qian and Hopfield (211), we assume the Gibbs free energy due to relaxation of the hydration shell is zero (i.e., $\Delta G_{\beta S} = 0$, $T_{C,S} = T$) for all temperatures, even though the respective ΔH and ΔS values can be rather large. Hence, we can omit the third term and re-write eq. (6.4) as:

$$\Delta G_{\beta} = \Delta H_{\beta_C} + \Delta S_{\beta_{PS}} \cdot \left(T_{C_{PS}} - T\right) \qquad \text{(eq. 6.5)}$$

Comparing this equation with the generalized linear equation for entropy-entropy compensation given in eq 6.1, we can then identify the general compensation temperature T_C with $T_{C,S}$ and the constant α with $\Delta H_{\beta c}$. Thus, our original linear fit to ΔH vs. ΔS data

(Figure 6.3) yields an average $\Delta H_{\beta c} = -0.770$ kJ/mol, indicating that the change in conformational free energy in the unfolded state favors pPII distributions slightly. The very small value we calculate for α (i.e. $\Delta H_{\beta c}$) is diagnostic of what Movileanu and Schiff termed a near-exact compensation (208). Individual analysis of ΔH/ΔS values for iso-equilibrium cluster 1 (IE1) and 2 (IE2) reveals that IE1 residues (L, V, I, S, K, Y, W, and F) exhibit a lower average conformational enthalpy of $\Delta H_{\beta c,IE1} = -0.324$ kJ/mol, indicating an even more exact thermodynamic compensation driving propensity among these residues. This low conformational enthalpy is also is in line with the rather iso-energetic character of this point ($\chi_{pPII,\ IE1}/\chi_{\beta,\ IE1} \approx 1$). On the other hand, cluster 2 residues (E,R,M,N) have a slightly larger conformational enthalpy $\Delta H_{\beta c,IE2} = -1.49$kJ/mol.

Re-examining the series of ΔS_β and ΔH_β values displayed in Figure 2 we observe the largest deviations from the linear compensation fit for alanine ($\Delta S_\beta = -23.0$ J/mol·K, $\Delta H_\beta = -10.3$ kJ/mol) and aspartic acid ($\Delta S_\beta = -38.0$ J/mol·K, $\Delta H_\beta = -9.2$ kJ/mol). This makes sense as these residues were found to lie outside of the iso-equilibria points (Figure 4c) where $\Delta G_{\beta c}$ for the series are approximately equal.

6.1.3.4 Rationalizing the Range of ΔH ΔS Values: Branched Aliphatics - A case study

Although the above described model indeed accounts for the obtained enthalpy-entropy compensation, it fails to explain the striking differences among ΔH/ΔS values of (predominantly) aliphatic residues with C_β branched (V,I) and linear side chains (K,L,M). The different ΔH and ΔS values of L and I residues are particularly surprising since their side chains are very similar, exhibiting comparable surface areas and hydrophobicities, which are both larger than the corresponding values for valine. The same can be said about the so-called conditional hydrophobic accessible surface areas, which Fleming et al. introduced as a crucial determinant for the Gibbs energy of residue conformations (120). These authors suggested that this surface is generally reduced in pPII compared with β strand conformations hence stabilizing the former via the hydrophobic effect. This would indeed increase the entropic and enthalpic difference between pPII and β-strand, but if the solvent accessible surface area is a measure for this effect, one would expect a hierarchy I≈L >V, contrary to our observation A similar hierarchy would be expected for solvent relaxation processes associated with a change in solvent accessibility. In addition to side-chain solvent interactions one could invoke variations of backbone-solvent interactions due to side chains as the dominant contribution to both ΔH and ΔS values of pPII/β equilibria, in agreement with Avbelj and Baldwin (121, 132). This notion is corroborated by the observed influence of amino acid side chains on the hydration of the N-terminal amide protons in blocked dipeptides, which Bai and Englander determined with hydrogen exchange experiments (213). The respective Gibbs energy associated with

the HD-exchange shows a V≈I>L hierarchy, indicating that V and I are more effective in shielding the peptide from hydration.

A complete theory which accounts for how enthalpy-entropy compensation governs amino acid conformation might need an extension beyond the consideration of peptide solvation to account for the differences observed between residues. It is possible, for instance, that the obtained differences in ΔH-ΔS values at least partially reflect different conformational restrictions for branched and linear side chains in pPII and β-strand conformations. To check for this possibility we analyzed the rotamer library of amino acid residues that Dunbrack and Karplus obtained for all amino acid residues from a large set of protein data bank files (214). Based on the positions and widths of distributions reported in our earlier studies (54, 55), we counted the number of side chain rotamers for the following regions of the Ramachandran plot from of Dunbrack and Karplus: $180^o \geq \psi \geq 160^o$ and $-100^o \geq \varphi \geq -120^o$ for β-strand of V and I, : $160^o \geq \psi \geq 140^o$ and $-100^o \geq \varphi \geq -120^o$ for β-strand of K, L and F, : $180^o \geq \psi \geq 160^o$ and $-60^o \geq \varphi \geq -100^o$ for pPII-strand of V, I, K, L and F. Utilizing this library, we found that only V and I have a larger number of rotamers in the β-strand (6 and 4 for V and I) than in the pPII region (3 for V and 0 for I), while the numbers are very similar for K, L and F (16 and 16, 16 and 15, 6 and 6, respectively). This observation suggests that the rotational degree of freedom of branched side chains is much more restricted in the pPII than in the β-strand region of V and I, whereas no such discrepancy exists for linear side chains. The fact that V and I still share an iso-equilibrium point with four other amino acid residues (including L) suggests

a nearly ideal enthalpy-entropy compensation so that these steric effects remain absent in the total Gibbs energy balance. Taking this into account, our Gibbs free energy model (eq. 5) has to be extended in order to account for enthalpy-entropy compensation brought about by steric restrictions of branched side chains:

$$\Delta G_{\beta} = \Delta H_{\beta_C} + \Delta S_{\beta_{PS}} \cdot \left(T_{C_{PS}} - T\right) + \Delta S_{\beta_{St}} \cdot (T_{c_{St}} - T) \qquad \text{(eq. 6.6)}$$

where $\Delta S_{\beta St}$ is the new conformational entropy term which is predominantly associated with differences between side chain entropies.

6.1.3.5 Conformational Entropy Reduction

Reproduced in part, from:

I. Schweitzer-Stenner, R., Toal, S. "Entropy Reduction in Unfolded Peptides (and Proteins) due to Conformational Preferences of Amino Acid Residues" Biophysical Journal, Submitted

Protein folding is a transition from a disordered to an ordered state. It is generally thought to involve a massive reduction of conformational entropy, which is compensated by a gain in enthalpy, and for hydrophobic residues, by a gain in solution entropy (4). Any reliable assessment of the thermodynamics of folding has to rely on an estimation of how much conformational entropy is eliminated by the folding process. It is obvious that the existence of conformational preferences of amino acid residues should substantially reduce the conformational entropy of unfolded peptides and proteins compared with a

local random coil scenario, hence increasing the stability of the native state relative to the unfolded state. Local order caused by conformational preferences would also reduce the entropy of IDPs, which should be advantageous for promoting disorder→order transitions caused by IDP binding to biomolecules (13). However, a quantitative picture connecting residue level order/disorder and conformational entropy has still to emerge and quantitative estimations about the suspected entropy reduction are rare in the literature. To explore the entropy reduction imposed by residue conformational bias we compare the conformational entropy of GxG peptides to that of a an ideal random coil ensemble. Figure 6.7 shows an artificial Ramachandran plot, which mimics a classical local random coil. Contrary to the experimentally conformational ensembles obtained for GxG peptides the random coil distribution assumes substantial sampling of right handed helical and similarly structured turn-like conformations. As in the classical plots of Ramachandran et al (5), there is no clear distinction between pPII and β-strand, since they display a rather homogeneous sampling of a broad region in the upper left quadrant of the Ramachandran plot.

Figure 6.7 also exhibits corresponding plots for GAG, GVG, GLG and protonated GDG, which represent four different types of amino acid residues with regard to their intrinsic propensities (54, 55). Alanine is peculiar in that it exhibits a significantly above the average intrinsic pPII propensity (i.e., 0.72). Leucine shows a more balanced pPII/β-strand ratio with pPII still being more populated. Valine has a slight preference for β-strand because it is heavily entropically favored over pPII. Protonated aspartic acid has an

above average propensity for various turn-like conformations stabilized by intra-peptide hydrogen bonding between side chain and amide proton.

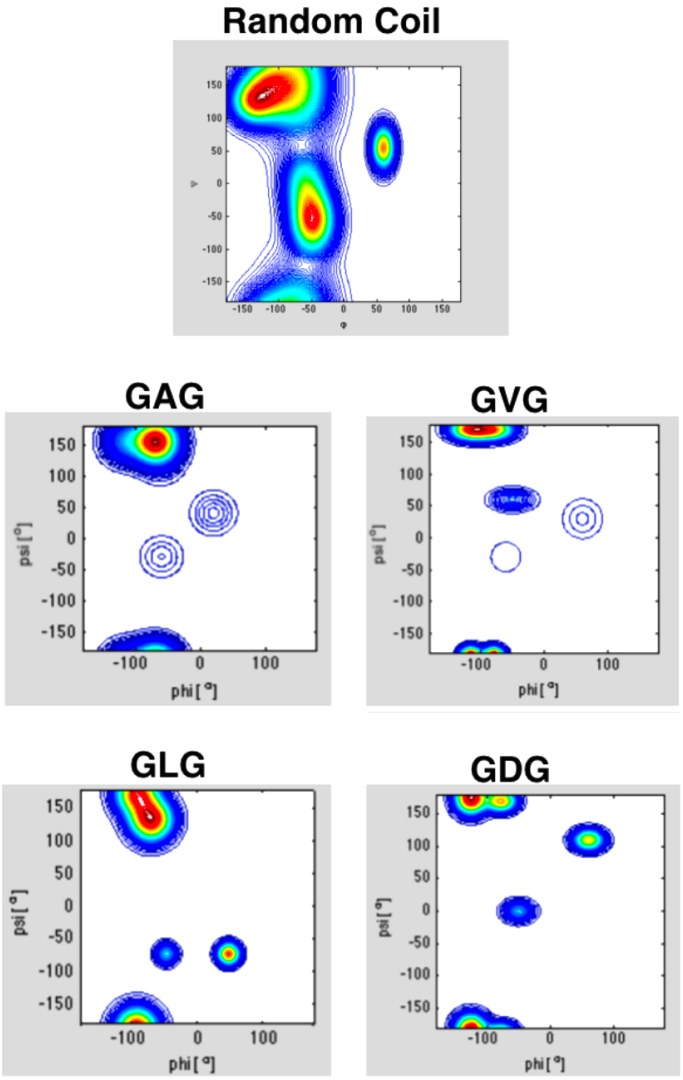

Figure 6.7: Ramachandran plots of a random coil-like ensemble compared to GAG, GVG, GLG, and protonated GDG.

To calculate the conformational entropy of the amino acid residue x in GxG for distributions such as those shown in Figure 6.7, we utilize the well know Boltzmann definition of entropy:

$$S_{conf} = R \int_{-\pi}^{\pi} P_x(\phi,\psi) \ln P_x(\phi,\psi) d\phi d\psi \qquad \text{(eq. 6.7)}$$

where R is the gas constant and $P_x(\phi\psi)$ is the mole fraction weighted conformational distribution for a respective residue (see *Chapter 2 of this thesis*, eq 2.30). The final result of a numerical integration of eq. 6.7 depends on the size of the micro-states into which the Ramachandran space is sub-divided.* Here, we use rather small sized segments with $\Delta\varphi=\Delta\psi=2.0^o$ which lies well within the regime in which S_x depends linearly on $\Delta\varphi$ and $\Delta\psi$. Since we are interested in entropy differences between S_{conf} and the entropy S_{rc} of the random coil distribution shown in Figure 6.7, we do not need to worry about absolute values:

$$\Delta S_{conf} = S_{rc} - S_{conf} \qquad \text{(eq. 6.8)}$$

entropy used the obtained thermodynamic data for GxG peptides to estimate the temperature dependence of the conformational entropy.

Figure 6.8 exhibits the differences between Helmholtz energy contributions $T\Delta S_x$ at 298 K for the investigated residues in the GxG model peptides. The significance of the obtained values is judged by the extent by which they exceed the thermal energy RT

indicated by the upper horizontal line. For aspartic acid (D), we calculated $T\Delta S_x$ for the fully protonated and the fully ionized molecule. At room temperature, all $T\Delta S_x$ values are negative which indicates that the conformational entropies of all amino acid residues are less then that of a random coil, as expected. While the residue specific $T\Delta S_{conf}(x)$ contributions to the free energy are within the energy available at room temperature, it should be noted that these are per residue values which, when considering the length of IDP regions in proteins, will accumulate and induce substantial variations in IDP conformational entropy depending on amino acid sequence.

All $T\Delta S_x$ values exceed the thermal energy, though to a different extent. The average deviation from random coil is -3.84kJ/mol. Not surprisingly, the largest deviation from random coil entropy is obtained for alanine with ΔTS_x=-4.72kJ/mol indicating a quite restricted conformational sampling at room temperature. The ΔTS_x varies significantly between residues, with differences reaching up to approximately 2kJ/mol. Alanine as well as aspartatic acid, asparagine, serine, threonine and methionine exhibit $T\Delta S_x$ values above 4kJ/mol, i.e. they lie in the 2RT regime (marked by the lower horizontal line in Figure 6.8). This is a reflection of their relatively high level of order (i.e. large degree of conformational bias) in the unfolded state.

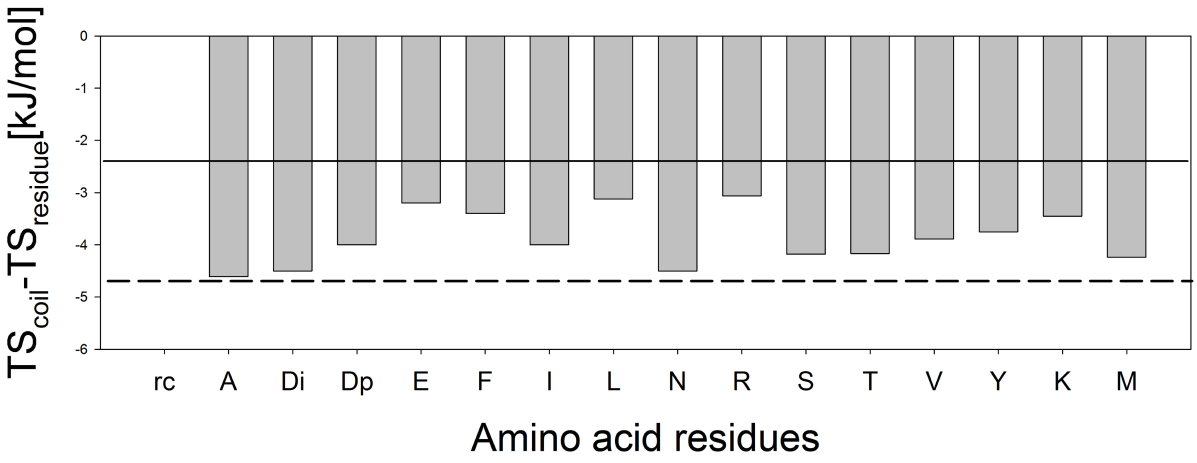

Figure 6.8: The difference between conformational entropy contributions to free energy of the indicated amino acid residues relative to a hypothetical random coil distribution. Single letters denote amino acid residues in GxG peptides. The index i indicates the ionized state of the residue. The solid line represents the available energy at RT, and the dashed line represents 2RT.

Interestingly, we found that for a majority of the residues investigated the conformational entropy relative to random coil, $\Delta S_{conf}(x)$, is at least partially correlated with the entropic difference between the subpopulations pPII and β (ΔS_{pPII-B}) derived from our earlier discussed thermodynamic analysis of HNMR data (Figure 6.9). We obtain a correlation coefficient of 0.63 when omitting the branched aliphatics isoleucine and valine which, as previously discussed, have abnormally high ΔS_{pPII-B} values in favor of β conformations. As the experimentally derived entropy difference between pPII and β conformations decreases along the series of amino acid residues, the relative conformational entropy $\Delta S_{conf}(x)$ increases, indicating larger deviations from a random

coil conformation.* This suggests that thermodynamic properties of underlying residual structure play a key role in modulating conformational entropy in the unfolded state.

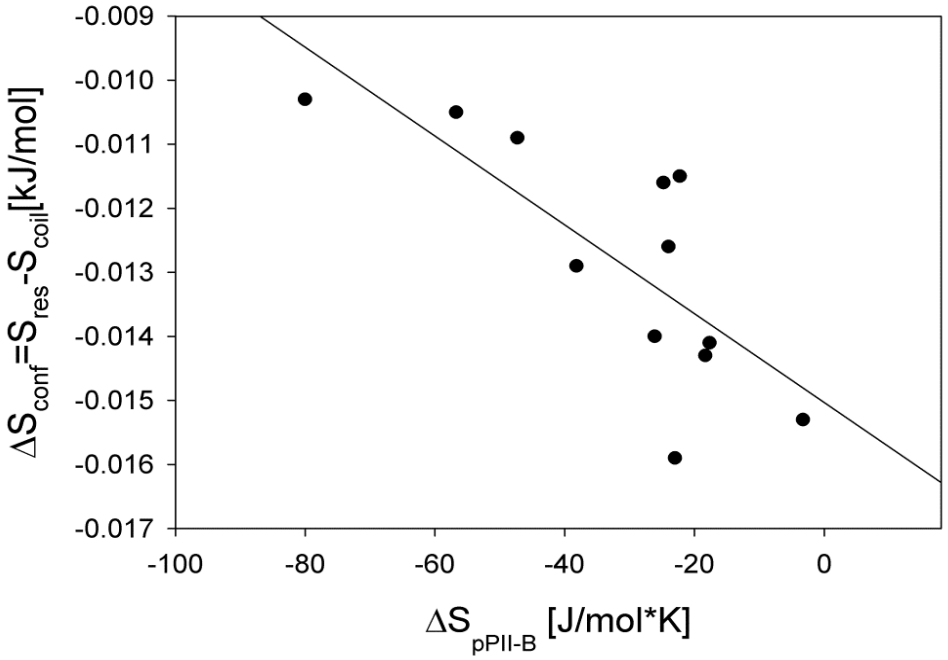

Figure 6.9: Relationship between $\Delta S_{conf}=S_{res}-S_{coil}$ and the experimentally derived $\Delta S_{pPII-\beta}$. The solid line reflects a linear least squares fit with a correlation coefficient of 0.63.

It has been previously shown *in this thesis* that an increase in temperature affects nearly exclusively the equilibrium between pPII and β of the central residue of GxG, whereas the population of turns is practically temperature independent (141). From the obtained pPII-↔β thermodynamic parameters we were able to calculate the temperature dependence of the conformational populations ($p_x(\phi,\psi)$). From these distributions, we

then calculated ΔS_{conf} for each peptide at various temperatures. In order to consider the increased sampling of conformational space within the individual (pseudo)potential wells for each conformation with rising temperature, we multiplied the Gaussian bandwidths of the respective sub-distribution with a factor $(T/T_R)^{1/2}$(215). We used the thus obtained distribution functions to calculate $T\Delta S_{conf}$ for five different temperatures between 290 and 360 K. As shown in Figure 6.10, the entropy difference between the conformational distributions of x in GxG and the assumed random coil distribution at room temperature decreases with increasing temperatures reflecting the increased available conformational space for each residue.

The temperature at which a given amino acid residue's conformational entropy contribution to the free energy reaches the random coil value at room temperature is different for each species. It is highest for isoleucine, for which random coil entropy is not reached within the considered temperature regime. This reflects the rather high β-strand propensity of this amino acid residue at high temperatures (216). For a majority of the amino acids residues, however, $T\Delta S_{xonf}$ becomes even positive at $T \geq 340K$, indicating a higher degree of disorder. These results suggest that with respect to their conformational entropy unfolded peptides and proteins might indeed become also locally random coil-like at high temperatures, even though distributions are still different from those of canonical random coil ensembles owing to the very low occupancy of right handed helical conformations. Taken together, the above analysis reveals that the intrinsic conformational entropy is residue specific and substantially lower then suggested by the

local random coil model. This indicates that particularly for IDP-type peptides (A_β, salmon calcitonin) and relatively small foldable proteins the total conformational entropy might depend on the amino acid composition. With regard to IDPs, the composition of sequences which are involved in e.g. protein-protein and protein-DNA interactions, might matter. Conformational entropies of IDPs and unfolded proteins of similar size might differ, because the former are less heterogeneous than the latter with regard to their amino acid composition.

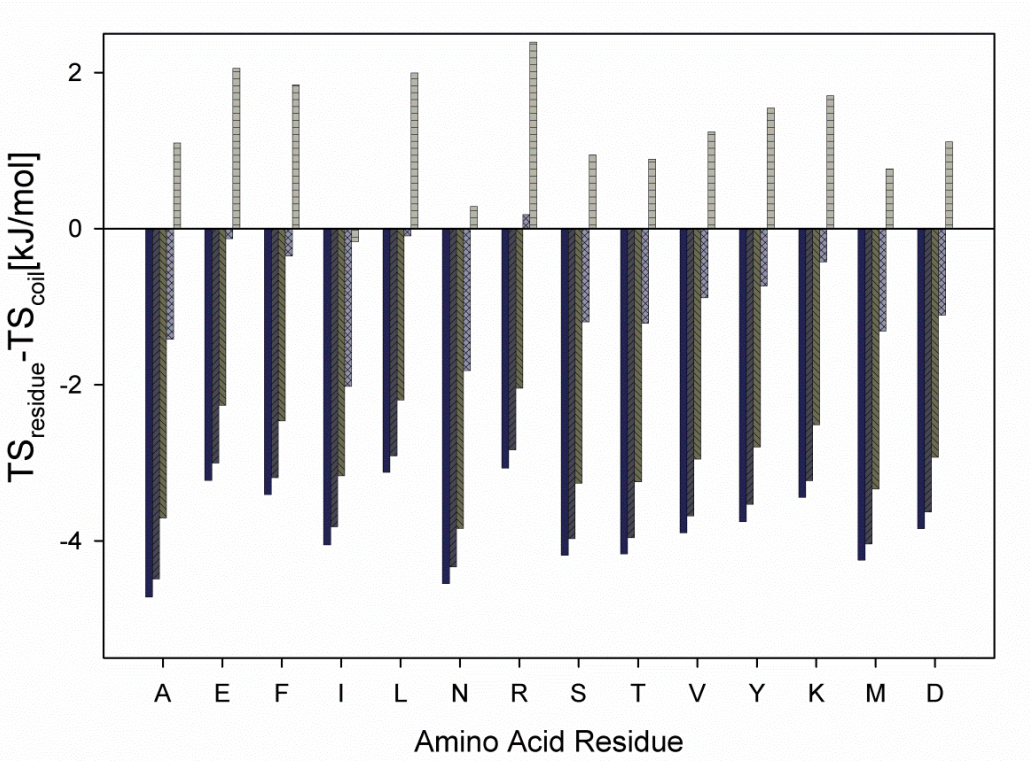

Figure 6.10: Relative conformational entropy contributions to free energy, $T\Delta S_{conf}$ at 298° K(dark grey), 301°K, 311°K, 340°K, and 370°K (light grey) derived from conformational distributions of the central x-residues in GxG peptides and the thermodynamics of pPII-β strand transitions.

The average $T\Delta S_{conf}$ for the values displayed in Figure 6.8 lie well in the range of values that Baxa et al. obtained from coil library biased MD simulations for ubiquitin (3.0 and 4.5 kJ/mol·residue) (217). We therefore wondered whether the obtained $T\Delta S_x$ values exhibit some correlation with helix propensities of amino acids in alanine based peptides. Conjectures in this direction are motivated by the fact that alanine, the amino acid residue with the highest propensity for pPII in the unfolded state, also exhibits the largest helical propensity in folded proteins (218). It has been shown by Schweitzer-Stenner et al. that the pPII propensity of the majority of the investigated residues exhibit a linear correlation with the respective Zimm-Bragg parameter s=[fraction helix]/[fraction coil] in alanine-based host guest systems (219). Figure 6.11 shows a plot which displays the relation between the corresponding –RT*ln(s) values of the investigated amino acid residues and their conformational entropy. The correlation is moderate (r=0.69). Interestingly, however, alanine (ΔS_x=-4.6 kJ/mol; -RT*lns=-3.8 kJ/mol) is clearly apart from any linear regression curve of the data (we assumed T=298 K for the plot). As for our earlier study, the s-values were taken from a very comprehensive host-guest study of Chakrabartty et al.(218)

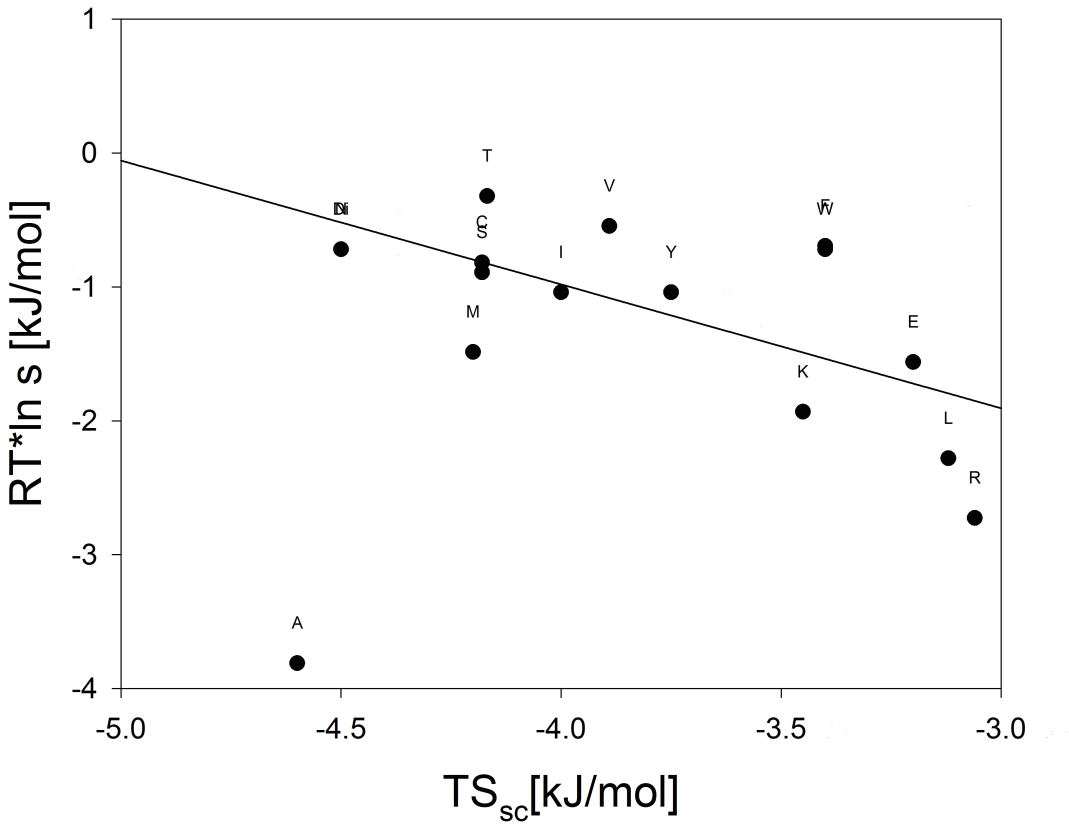

Figure 6.11: Comparison of the per-residue Gibbs energy of helix formation (-RT ln(s), s is the Zimm Bragg parameter) and $T\Delta S_{conf}$ of 15 amino acid residues in GxG dissolved in aqueous solution at T=298 K. These entropy values were derived from experimentally obtained conformational distributions. The solid line is the product of a linear regression.

It could be that a more meaningful correlation could be obtained by comparing -RT*ln(s) with the loss of backbone and side-chain conformational entropy concomitant with the transition from a statistical to a helical distribution. The reduced side chain entropy in the helical state has been proposed by Creamer and Rose as a key determinant for the helical propensity of amino acid residues in alanine based peptides (220). This conformational entropy reduction ΔS_{conf}* can be written as

$$\Delta S_{conf}^{*} = \left(S_{\alpha,conf} - S_{rc} \right) - \Delta S_{conf} + \Delta S_{conf,chain} \qquad \text{(eq. 6.9)}$$

where $S_{\alpha,conf}$ is the conformational entropy of the residue in a right-handed helical conformation. This was calculated by assuming that each residue can sample a helical conformational space describable by a Gaussian distribution with $\sigma_{\varphi}=\sigma_{\psi}=10 cm^{-1}$. This helical distribution resembles those obtained from crystal structures and MD simulations (221). The entropic loss of side chains, $\Delta S_{conf,chain}$ has been reported by Blaber et al. based on the analysis of crystal structures of proteins (222).

Figure 6.12 shows a plot of –RT*ln(s) versus the derived (helical corrected) ΔS_{conf}^{*}. A first glance at the data reveals no significant correlation. However, a closer look suggests the existence of two subsets of data for for which a good correlation might exist. The smaller of these subset contains the branched and bulky residues V, I, W and T; the larger subset contains the remaining investigated residues besides alanine which (not surprisingly) does not fit into any of these two sets. The branched residues have a lower

helical propensity. We subjected both sets to linear regressions and obtained r-values of 0.96 and 0.87, which both indicate good correlations.

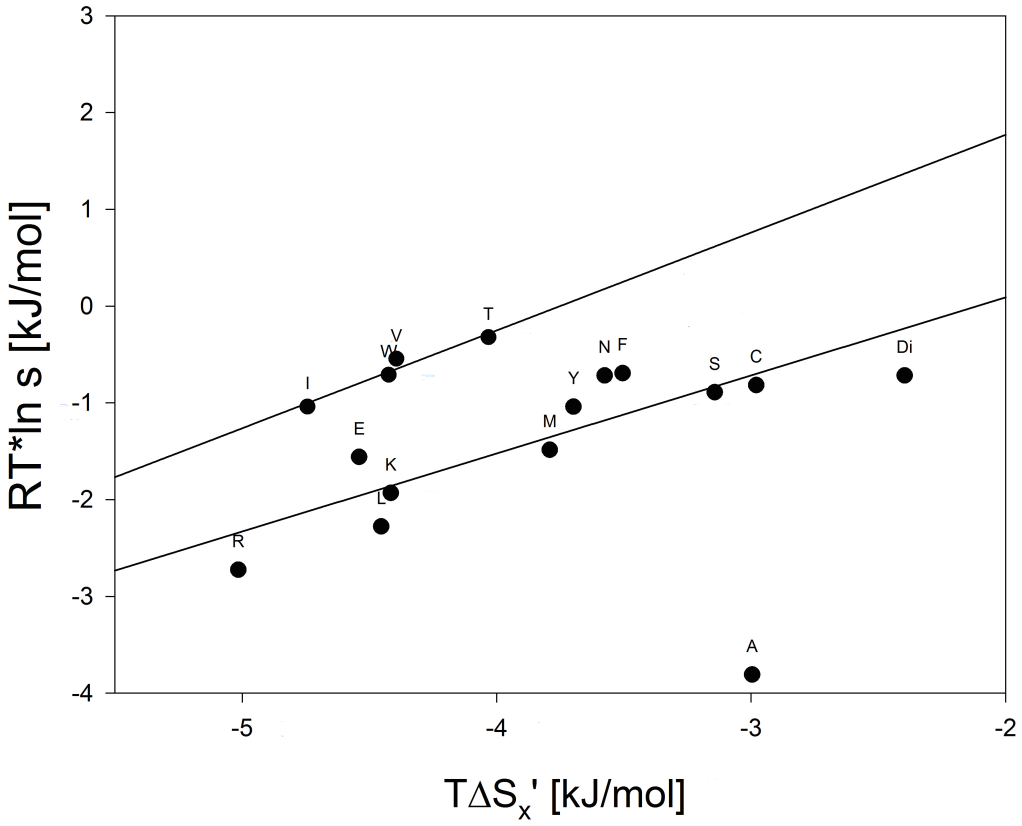

Figure 6.12: Comparison of the per-residue Gibbs energy of helix formation (-RT ln(s), s is the Zimm Bragg parameter) and the conformational entropy loss of 15 amino acid residues upon a transition from a coil-like to a right handed helical conformation. The temperature value is 298 K. The conformational entropy loss was calculated as the sum of backbone side chain entropy changes as described in the text. The solid lines represent linear regression of the two clusters of data.

To appreciate the significance of these values one should keep in mind that correlations between Gibbs enthalpies and entropies are notoriously bad even in the case of excellent correlations between enthalpy and entropy (210). The correlations obtained for both data sets clearly suggest that the helical propensity actually decreases with decreasing difference between the conformational entropy. Alanine is an exception, as it exhibits a high helix propensity and a very limited loss in conformational entropy. This is somewhat counter-intuitive, as one might expect that a reduction in entropy loss should yield a higher helical propensity. However, what looks like a surprising result actually makes physical sense, if one considers the well established fact that folding/unfolding processes in general and helix⇔coil transitions in particular are governed by enthalpy-entropy compensation which becomes exact at the folding temperature. For the helix⇔coil transition of different alanine based peptides one can assume that the folding temperatures are similar. In such a case enthalpy and entropy are linearly related. In the ideal case of identical transition temperatures one can write:

$$\Delta H_{conf} = T_c \Delta S_{conf}^{*} \qquad \text{(eq. 6.10)}$$

where T_c is the compensation temperature. The Gibbs energy is thus written as:

$$\Delta G_{conf} = \left(T_c - T\right) \Delta S_{conf}^{*} \qquad \text{(eq. 6.11)}$$

Hence, an increasingly negative ΔS_{conf}^{*} causes an increasingly negative ΔG_x for $T < T_c$ and thus an increasing helix propensity, as reflected by our data.

The total per-residue entropy contribution to the Zimm-Bragg parameter s cannot be totally accounted for by conformational entropies. Another contribution certainly arises from hydrophobic effects, i.e. the partial burial of side chains in the helical conformation. Blaber et al. (222) estimated the respective Gibbs energy associated with a helix ↔ coil transition for all amino acid residues. One can assume that this contribution is mostly entropic in nature at room temperature (4). We thus added these hydrophobic contributions reported by Blaber et al. to $T\Delta S_x$* to obtain new entropy values $T\Delta S_x$**. However, a plot of –RT*ln (s) versus $T\Delta S_x$** did not reveal any correlation (data not shown). This suggests that adding hydrophobic contributions practically eliminates the surprising positive correlation between the loss of conformational entropy and helical propensity.

The behavior of alanine deserves some additional comments. Many attempts have been made to explain the high helical propensity of this residue in folded proteins. In addition, results reported in our studies clearly show that alanine does not fit into any of the categories for other amino acids in the unfolded state. First, alanine has an abnormally high pPII propensity in the unfolded state (0.72 in GAG, > 0.8 in AAA, > 0.6 in polyalanines). With regard to GxG, its ΔG(T) curve does not share an iso-equlibrium point with other residues, most likely owing to the fact that its enthalpy-entropy compensation temperature is higher and pure conformational energy difference (i.e. *vacuo*) between pPII and β-strand are still slightly in favor of the former). In this study, we show that the high helical propensity of alanine t involves a comparatively small loss

of conformational entropy upon the transition from a coil to a helical distribution. We, in accordance with others, attributed the high pPII propensity to an optimal hydration of the backbone, which is only guaranteed with alanine as side chain (105, 194). In the helical state alanine guarantees minimal interactions between side chains (compared to more bulky residues) and is likely to still permit backbone hydration thus minimizing disfavorable enthalpic contributions to the Gibbs energy and allowing the hydrogen bonding contribution to dominate.

What is the general significance of this and related thermodynamic studies on short peptides for the understanding of unfolded peptides/proteins and the energetics of protein folding? With regard to the unfolded state it becomes clear that in principle the total entropy depends on the amino acid sequence, but this might become irrelevant for longer proteins with comparable mixtures of aliphatic, polar and ionizable side chains. The average conformational entropy for these three groups is slightly different; from our data we obtained $<\Delta TS_x>_{aliph}$=4.0 kJ/mol (V,I,L,M,A), $<\Delta TS_x>_{arom}$=3.5 kJ/mol (Y,W,F), $<\Delta TS_x>_{charged}$=3.5 kJ/mol (K,E, ionized D, R) and $<\Delta TS_x>_{pol}$=4.2 kJ/mol (protonated D, S,T and N). IDPs are known to contain more charged and less aliphatic and aromatic residues than foldable proteins. If we e.g. assume that an IDP has 10 more charged residues and 10 less aliphatic residues than a foldable protein of comparable size this would provide the IDP with 5kJ/mol of extra entropic contribution to the Gibbs energy. Since IDPs typically have more polar than aromatic groups, the actual entropy gain might

be smaller. Therefore, the sequence will matter for only comparatively short IDPs or within IDP segments, particularly if the composition is not very heterogeneous.

A stronger sequence influence could arise from the conformation and residue dependence of the solvation enthalpy and entropy between pPII and β reported ($\Delta S_{pPII-\beta}$, $\Delta H_{pPII-\beta}$) in *section 6.1.2 of this thesis*. This should have a significant effect particularly on the stabilization of unfolded states at high temperatures, at which entropically favored β-strand structures become increasingly populated. To illustrate this point we calculated the changes in average enthalpy and entropy associated with a temperature increase from 298 to 350 K by utilizing the following relationships:

$$\left\langle \Delta\xi_{pPIi-\beta} \right\rangle = \frac{\Delta\xi_{pPII-\beta} e^{-\Delta G_{pPII-\beta}/RT}}{1+e^{-\Delta G_{pPII-\beta}/RT}} \qquad \text{(eq. 6.12)}$$

where $\Delta\xi_{pPII-\beta}=\Delta H_{pPII-\beta}$, $\Delta S_{pPII-\beta}$,R is the gas constant and T the absolute temperature. Figure 6.13 shows the difference between $<\Delta H_{pPII-\beta}>$ and $T<\Delta S_{pPII-\beta}>$ values calculated for 298 and 350 K for all amino acids investigated. The obtained diagram is reminiscent of the large differences between the $\Delta H_{pPII-\beta}$ and $\Delta S_{pPII-\beta}$ values reported by Toal et al. (216). While the changes are in the range between 0 and -5 kJ/mol for most amino acid residues, L, R and particularly I and V stand out in that the data indicate their solvent free energy becomes substantially more entropic at high temperature. This implies that unfolded states of peptides and proteins with a preponderance of these amino acids (e.g.

in aliphatic peptides with a high propensity for self-aggregation) are substantially more entropic and less enthalpic than polypeptides with more polar or ionizable side chains.

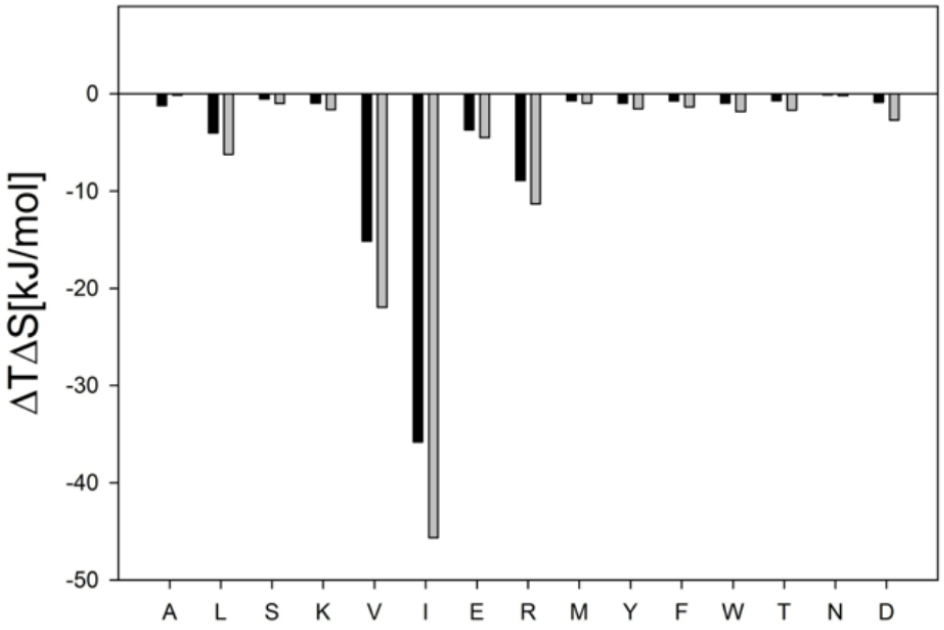

Figure 6.13: Plot of differences between $<\Delta H_{pPII-\beta}>$ and $T<\Delta S_{pPII-\beta}>$ of GxG peptides calculated for temperatures of 350 and 298 K by utilizing the $\Delta H_{pPII-\beta}$ and ΔS_{pPII-}

6.1.4 Conclusions

These results of these thermodynamic investigations on GxG peptides are interesting and very significant for numerous reasons. Firstly, the co-existence of both compensation effects and iso-equilibria is rare, and indeed this is the first study, that captures and explains this effect for residue level transitions. As discussed above, their occurrence indicates a common source behind promotion of conformational preferences in the unfolded state, which we can directly link to the competing action of peptide solvation and intrapeptide hydrogen bonding. Solvent reorganization might contribute significantly to enthalpy and entropy, but does not affect the conformational equilibrium. Secondly, the coalescence of χ_{pPII}/χ_{β} (i.e. the iso-equilbria) suggests that amino acid composition, and hence individual propensities in the unfolded state, become much less significant near physiological temperatures (310 K). Thus, within the confines of our obtained pPII/β distributions, and in the absence of nearest-neighbor interactions (i.e. solely intrinsic effects), one aspect of the classical random coil model becomes nearly valid at compensation temperatures, namely the independence of conformational ensembles on the amino acid sequence of peptides and proteins. Therefore, only the different turn-like fractions among the residues matter at these compensation temperatures. Thirdly, the deviation of alanine, aspartic acid, and threonine from iso-equilibrium is very interesting. Aspartic acid and threonine are among a sub-class of amino acids that have unusually strong preferences for turn-like conformations which in some cases have been found to be stabilized by side-chain backbone H-bonding (55).

How these propensities for turn-like structures affect the thermodynamics of the pPII-β equilibrium is unclear. We conjecture that the polar character of the side chains can significantly affect the overall hydration shell of the peptide thus causing a change of the compensation temperature. One might be tempted to consider interactions between the side chains and the terminal groups as well (so called end effects) (192). However, this can be definitely ruled out for GDG, since Rybka et al. (141) showed that conformational distributions of protonated GDG, AcGDG and the aspartic acid dipeptide are indistinguishable.

The third amino acid residue that does not share any iso-equilibria with other residues is alanine. The conventional explanation, based on several MD simulations would invoke a particularly efficient hydration shell packing as the stabilizing factor of pPII (119). However, in order to invoke protein-hydration as the driving force for pPII stabilization one would have to assume that the compensation temperature for alanine hydration is much higher for any of the other residues (an estimation puts it above 400 K, Figure 6.4). To further study the effect of solvation on the conformational preference of alanine, we performed a separate investigation of alanine based peptides in various co-solvent systems, as discussed below in Chapter 6.2.

In addition, we found that the conformational preferences of amino acid residues lead to a substantial decrease of conformational entropy. The corresponding contributions to the free energy at room temperature lie between 1 and 2 RT. The entropy reduction effect is most pronounced for alanine and protonated aspartic acid. Upon increasing

temperature the backbone entropy of GxGs approaches the entropy of the assumed random coil distribution. We used the obtained conformational backbone entropies to calculate the respective loss associated with a coil → helix transitions and found that it does correlate moderately with the respective Gibbs energy of helix formation (i.e. -RT*ln s). However, upon adding literature values for the loss of side chain entropy to our backbone conformational entropy values we obtained rather good correlations for two different subsets of amino acid residues. Taken together, we show that conformational propensities of amino acid residues in GxG can be used for deriving a realistic estimation of the backbone entropy of unfolded peptides and proteins.

6.2 Effects of alcohol co-solvents on the conformation of AAA

6.2.1 Introduction

The physical reason for the preferred sampling of the pPII trough by alanine has been extensively debated in the literature. As outlined in detail *Ch 1 of this thesis*, most theories have in common that they emphasize the role of water and solvation for the stabilization of pPII (76, 119). If water is indeed pivotal for the sampling of pPII conformations in unfolded peptides and proteins, the conformational distributions of the latter should be substantially altered by the addition of co-solvents. Theoretical and experimental studies aimed at describing the effect of certain co-solvents on protein/ peptide conformations in aqueous solution have been performed thus far without providing an unambiguous picture. It is well established that protein stability can be enhanced with the admixture of so-called kosmotropic co-solvents, such as polyols (223, 224). The thus induced stability can be described in terms of shifting conformational equilibria to more energetically favorable states. In addition, certain polyols such as glycerol have been shown to prevent protein aggregation, apparently by inhibiting partial unfolding or misfolding of aggregation-prone peptide segments (225).

Typically, co-solvent-induced stabilization is discussed in terms of preferential hydration of the protein backbone by water or, similarly, preferential exclusion of co-solvent . Specifically, Gekko and Timasheff proposed that protein stabilization is a result of the preferential hydration of its backbone via exclusion of the glycerol co-solvent from the immediate domain of the protein (223, 225). More recently, Head-Gordon and co-

workers analyzed data from neutron diffraction experiments on N-acetyl-leucine-methylamine (NALMA) in water and in a 1:5 glycerolwater mixture with an empirical potential structure refinement algorithm (226). Their results also suggest that glycerol preserves the peptide hydration shell. In addition, this group performed MD simulations on the aforementioned system and observed that the population of water molecules in the solvation shell of the hydrophobic leucine residue increases in the presence of glycerol, while the peptide-H_2O hydrogen bonds are maintained (227). These findings were found to be consistent with their results obtained from neutron diffraction data. However, a rather different picture emerged from a recent MD simulation of Vagenende et al. who investigated the proximity of glycerol molecules at the (hen egg white) lyzosyme-solvent interface in a glycerolwater mixture (228). Their results suggest that two modes of protein-glycerol interactions are operative. One of them involves electrostatic interactions of a glycerol molecule with the protein surface, which induces an orientation of the solvent molecule so that further interactions between glycerols and protein groups are sterically inhibited. The second mode of interaction causes glycerol to preferentially interact with aliphatic groupings on the backbone, acting as an amphiphilic interface between the local hydrophobic surfaces on the protein and the polar water solvent. This type of interaction has direct consequences for stabilizing aggregation-prone hydrophobic peptide segments in solution.

Clarifying this issue is of great importance for a variety of reasons. First, glycerol is used in numerous biophysical experiments on proteins (particularly those carried out at subzero temperatures) including protein-folding studies (229). It is also routinely employed for protein crystallization (230) and in the food industry (231). As indicated above, it stabilizes the folded state of proteins. However, if, in contrast to widespread thinking, glycerol does indeed interact directly with proteins and peptides, it is also necessary to determine how the cosolvent affects the Gibbs energy landscape of the unfolded state, which is much more exposed to the solvent than the folded state. In addition, investigation of the effect that small co-solvents have in the presence of water on the conformational distribution of alanine should shed light on how water itself stabilizes the pPII conformation.

To tackle this issue, we investigated the influence of glycerol and ethanol on the conformational distribution of cationic trialanine in water. Ethanol was selected as an additional alcohol cosolvent for this study to make contact with a recent NMR/CD study on short unfolded peptides in binary mixtures of water and various alcohols (126). Moreover, we deemed it useful to compare two co-solvents which differ in terms of their bulkiness and their number of aliphatic and polar groups. For this study, ^{1}H NMR and UVCD spectroscopy was combined to determine how the conformational distribution of cationic trialanine in water is modified in mixtures of water with different amounts of ethanol and glycerol. The resultant temperature dependence of the maximum dichroism

measured at 215 nm and of the $^{3}J(H^{N}H^{\alpha})$ coupling constants of two amide protons was subjected to a global analysis based on simple two-state models, from which the respective thermodynamic parameters (enthalpy and entropy) as well as the conformation specific spectroscopic parameters (i.e., average $\Delta\varepsilon$ and $^{3}J(H^{N}H^{\alpha})$ of pPII and β-strand sub-distributions) were obtained. In addition, IR and polarized Raman spectra of trialanine for selected binary mixtures of the above cosolvents were measured in oder to identify possible interactions between the alcohol and peptide molecules. Our results clearly show that polar cosolvents affect the energy landscape of peptides. Moreover, they provide strong evidence for the notion that they interact directly with the aliphatic side chains. Our NMR data furthermore suggest that the thermodynamics of these interactions are different for the central and C-terminal residue of the peptide.

6.2.2 Results

6.2.2.1 Trialanine in Water

In a first step, we measured the far UVCD spectra of cationic trialanine in water as a function of temperature between 10 and 90°C, which are shown in Figure 6.14. The inset depicts the difference spectrum calculated by subtracting the spectrum recorded at 10°C from that measured at 90° C. At low temperatures, the pronounced positive maximum of the UVCD spectrum at approximately 215 nm is diagnostic of a dominant sampling of pPII-like conformations, in agreement with what has been now well-

established in the literature (68, 70, 191, 232). The difference spectrum indicates a change of the conformational distribution from pPII-like to more extended β-strand-like conformations, again in agreement with earlier results. The maximum dichroism (Δε) at approximately 215 nm decreases with increasing temperature as shown in Figure 6.15, also reflecting this shift from pPII to β states.

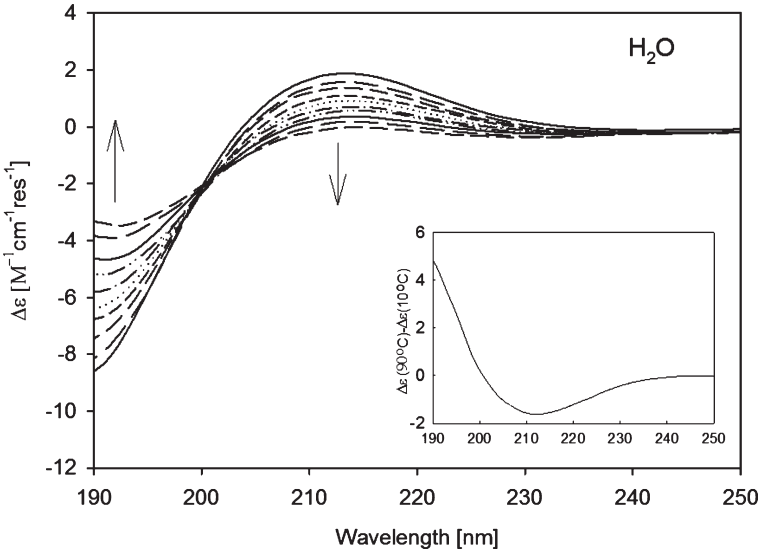

Figure 6.14. Temperature-dependent UV-CD spectra of cationic AAA in H_2O at pH 2.0. Arrows indicate increasing temperature from 0 to 90 C. Inset: CD difference spectrum obtained by subtracting the spectrum measured at 10 °C from the spectrum recorded at 90 °C.

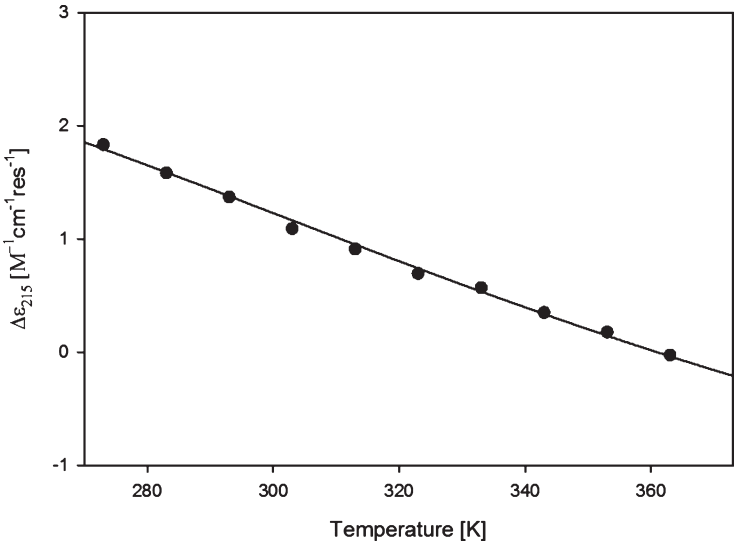

Figure 6.15. Maximum dichroism ($\Delta\varepsilon_{215\ nm}$) obtained from the UV-CD spectrum of cationic AAA in H_2O plotted as a function of temperature from 273 to 363 K.

To obtain residue-level details about the peptide conformations, we measured the ^{1}H NMR spectrum for cationic trialanine in water and determined the C-terminal and N-terminal $^3J(H^NH^\alpha)$ coupling constants as a function of temperature (Figure 6.16). Here, we utilize the nomenclature of Oh et al., who termed the coupling constant reflecting the ϕ-angle of the central alanine residue of the peptide as "N-terminal" (196). This reflects the fact that the N-terminal alanine is not associated with an amide proton. The experimentally determined value for the N-terminal coupling constant increases in the temperature range studied from 5.53 to 6.14 Hz in very good agreement with experimental studies by Graf et al. (74) These values are also indicative of a change in the

conformational equilibrium between pPII to more β-like distributions. The experimental coupling for the C-terminal ranges from 6.4 to 6.86 Hz, which is also in agreement with literature values(232).

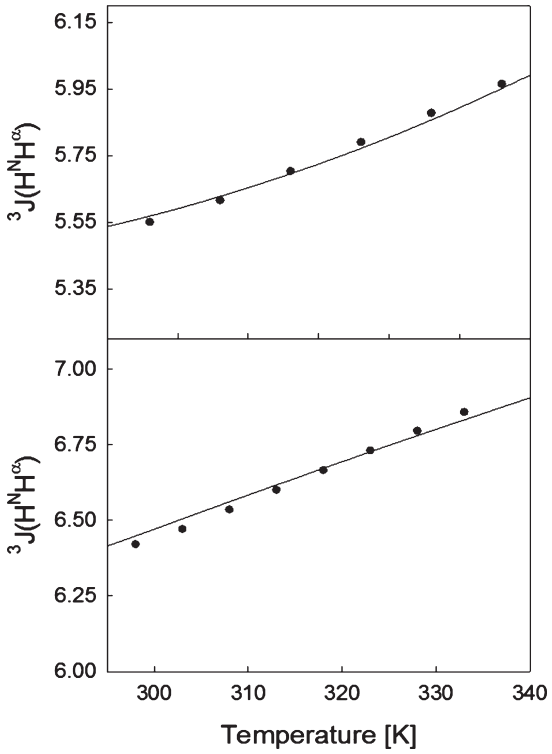

Figure 6.16. Temperature dependence of the $^3J(H^NH^\alpha)$ coupling con-stants of cationic AAA in H_2O obtained from the NMR signals of the N-terminal (upper panel) and C-terminal (lower panel) amide protons. The solid lines represent fits of a two-state model.

To extract the free energy landscape for trialanine, we employed a global fitting procedure to analyze the temperature dependence of the conformationally sensitive maximum dichroism $\Delta\varepsilon$ (T) and the $^{3}J(H^{N}H^{\alpha})(T)$ values with a two-state pPII-β model. The analysis was based on the conformational distribution obtained for the central residue of trialanine at room temperature (56, 70). This way, the average $^{3}J(H^{N}H^{\alpha})$ coupling value pertaining to each sub-distribution (i.e., J_{PPII}, J_{β}) could be determined utilizing the newest version of the empirical Karplus relationship (160, 163). In general, the method employed within this study is exactly the same as that employed in *Chapter 5 of this thesis* for alanine based peptides and described in full in *Chapter 4 of this thesis*. In the fitting procedure we consider only two states, and ignore the very low population of turns (<8%) for trialanine. Using the two-state model, we obtain excellent fits to the experimental $^{3}J(H^{N}H^{\alpha})(T)$ for both N- and C-terminal of trialanine in water as seen above in Figure 6.16. The thermodynamic values for ΔH and ΔS obtained from this analysis are listed in Table 6.2.

Table 6.2: Spectroscopic and thermodynamic parameters derived from fitting the temperature dependence of the $^3J(H^N,H^\alpha)$ coupling constants for cationic AAA in H_2O using the two-state fitting procedure described in the text. The $^3J(H^NH^\alpha)$ is the coupling constant obtained experimentally at room temperature

	$^3J(H^NH^\alpha)$ [Hz]	$^3J_{pPII}$ [Hz]	$^3J_\beta$ [Hz]	$\Delta\varepsilon_{pPII}$ [$M^{-1}cm^{-1}$ res^{-1}]	$\Delta\varepsilon_\beta$ [$M^{-1}cm^{-1}$ res^{-1}]	χ_{pPII}	ΔG [kJ/ mol]	ΔH [kJ/ mol]	ΔS [J/ mol*K]
AAA, N-terminal	5.61	5.02	9.18	1.32	-1.55	0.86	-4.44	-20.6	-54.4
AAA-C-terminal	6.40	5.28	9.27	1.32	-1.55	0.72	-2.32	10.15	26.29

6.2.2.1 Thermodynamics of AAA in Binary Mixtures

To examine the effect that addition of alcohol co-solvents has on the conformational distribution of alanine residues in trialanine we chose to implement 5%, 30%, and 60% co-solvent binary mixtures with water. The UV-CD spectra of three ethanol/water and three glycerol/water mixtures as a function of temperature are shown in Figure 6.17. Although we also measured the CD spectra for a 60% ethanol/water (v/v) we could not obtain resolved H NMR spectra for this mixture, so we omit this mixture from our analysis. In the region below 200 nm, some of these spectra are somewhat noisier than those of AAA in pure water, which might be due to impurity scattering or high

solvent absorptivity. However, this was not much of a concern because we used the maximum Δε at approximately 215 nm for our thermodynamic analysis.

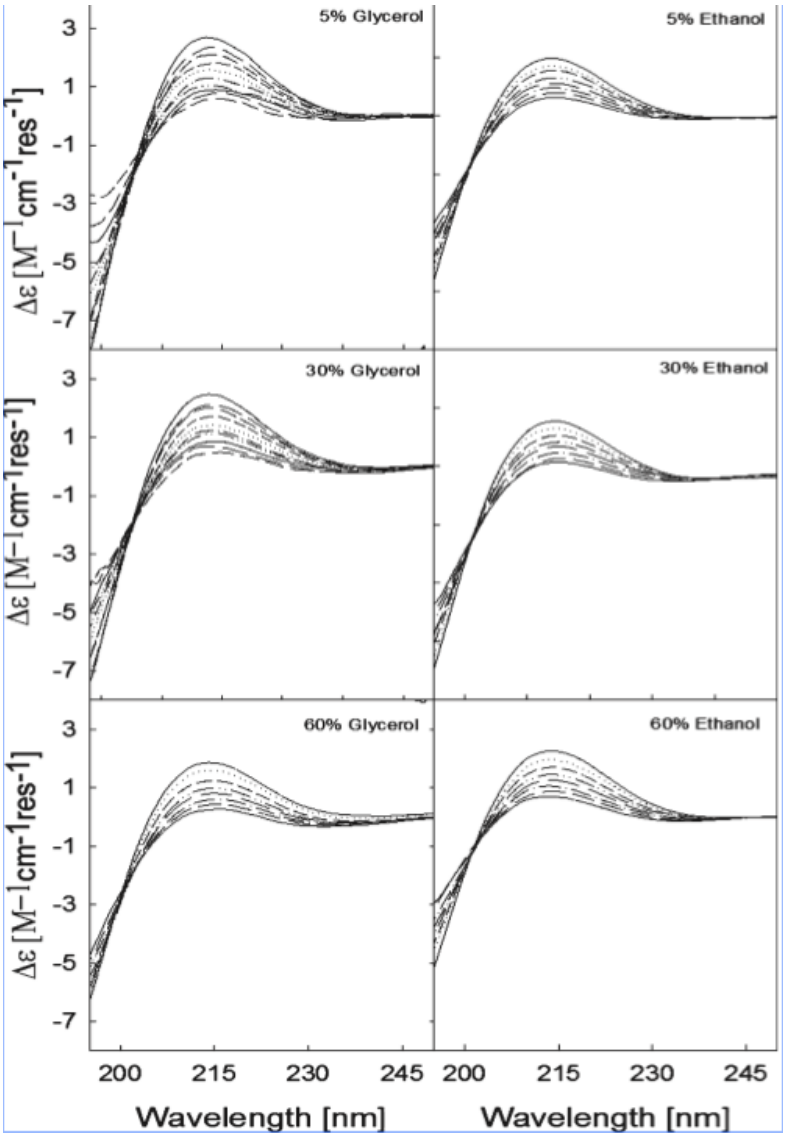

Figure 6.17. Temperature dependence of UVCD spectra of cationic AAA in glycerol/ H_2O and ethanol/H_2O binary mixtures. Upper panel: 5% binary mixtures. Middle panel: 30% binary mixtures. Lower panel: 60% binary mixtures.

The corresponding $\Delta\varepsilon(T)$ plots, as well as the temperature-dependent $^3J(H^NH^\alpha)$ coupling for each binary mixture, are shown in Figure 6.18 and Figure 6.19, respectively. The solid lines all result from fitting the experimental data using the aforementioned two-state model fitting routine. The temperature-dependent decrease in the maximum dichroism $\Delta\varepsilon_{215}$ indicates decrease pPII content as a function of increasing temperature, in line with what was found above for trialanine pure water. The $^3J(H^NH^\alpha)$ coupling constants for all solvent mixtures increase as a function of temperature, also signifying an increased sampling of β-like conformations in the ensemble. Interestingly, at room temperature the experimentally determined $^3J(H^NH^\alpha)$ values for AAA in ethanol/water and glycerol/water at the same volume percent are nearly identical, i.e., $^3J(H^NH^\alpha)$= 5.45 Hz for 5% admixtures alcohol and $^3J(H^NH^\alpha)$ = 5.77 Hz for 30% admixtures alcohol. However, the temperature coefficients for the $^3J(H^NH^\alpha)$ values of each mixture are different for the two alcohols, thus indicating a different Gibbs energy landscape and more pronounced differences between corresponding conformational mixtures at higher temperatures. The increase of $^3J(H^NH^\alpha)$ with temperature is generally more pronounced for water/ethanol as compared to water/glycerol mixtures, indicating that the conformational distribution of AAA shifts toward higher β content for ethanol more efficiently then for AAA in glycerol mixtures.

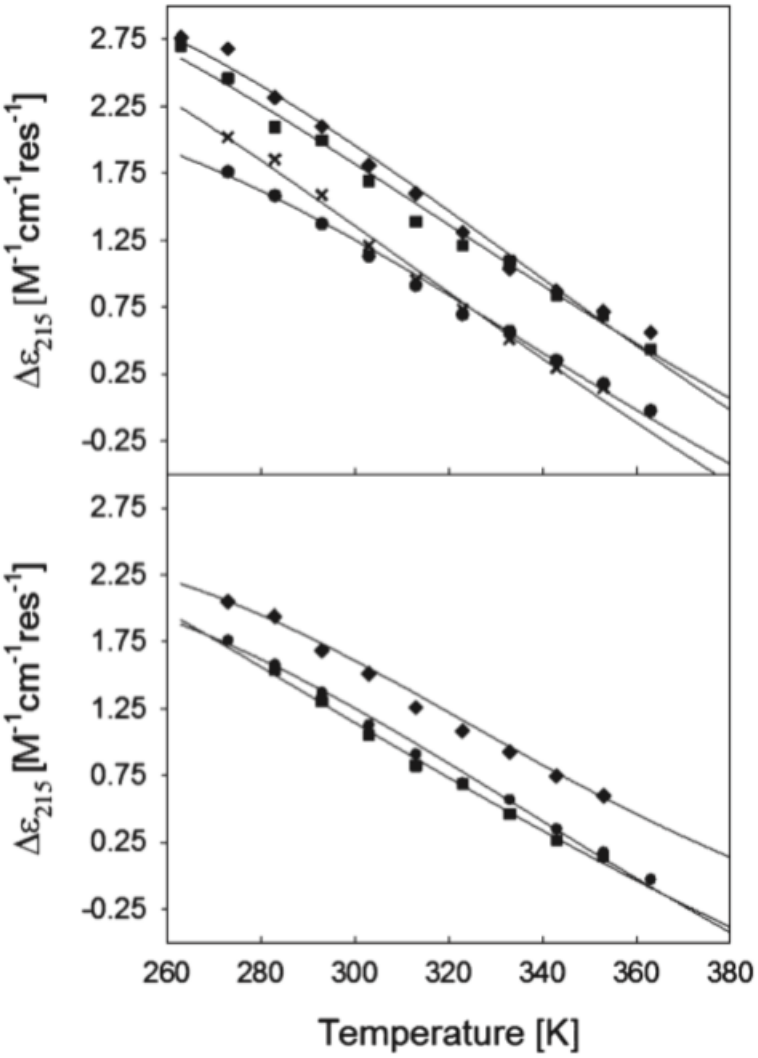

Figure 6.18. Maximum UV-CD ($\Delta\varepsilon_{215}$) signal plotted as a function of temperature from 273 to 363 K for AAA in binary mixtures. The solid lines result from fitting procedures described in the text. Glycerol binary mixtures (upper panel): D_2O (circles), 5% glycerol (diamonds), 30% glycerol (squares), and 60% glycerol (crosses). Ethanol binary mixtures (lower panel): D_2O (circles), 5% ethanol (diamonds), 30% ethanol (squares).

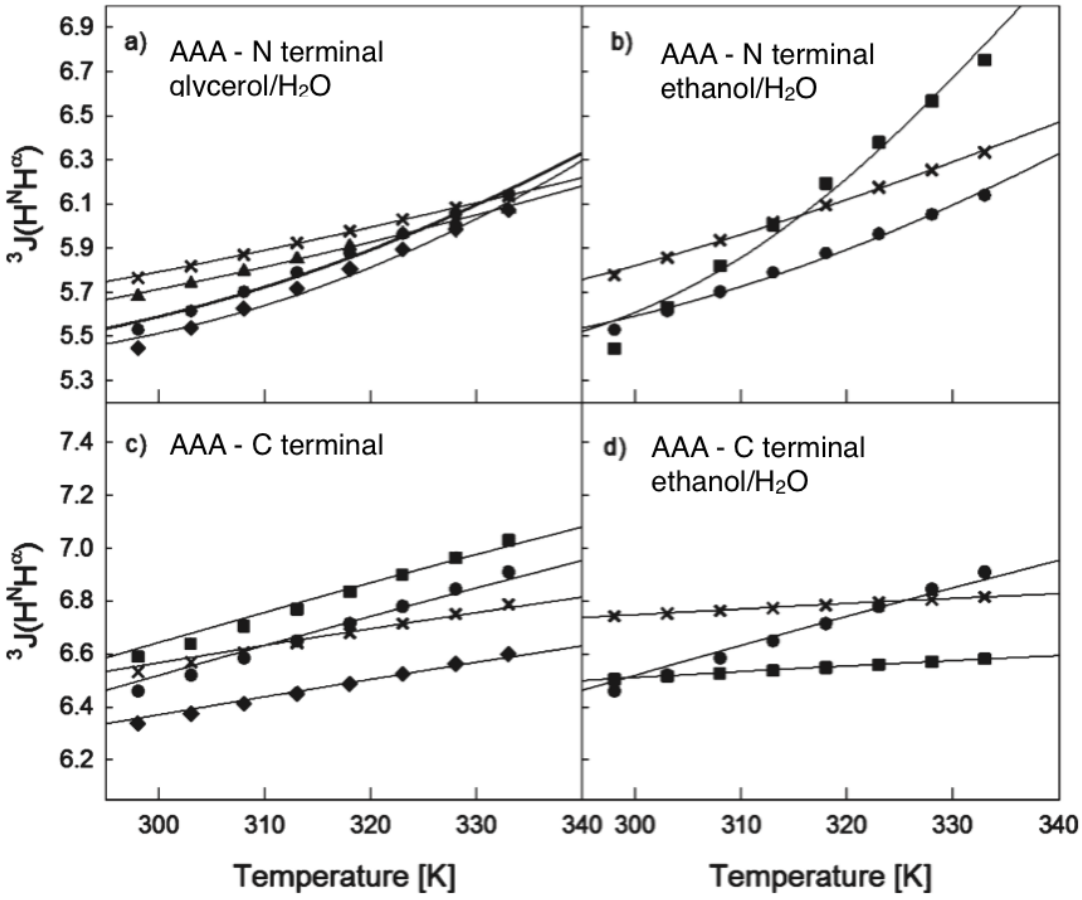

Figure 6.19. $^{3}J(H^{N}H^{\alpha})$[Hz] of N-terminal (upper panels) and C-Terminal (lower panels) plotted as a function of temperature from 298 to 333K for AAA in binary mixtures. The data for glycerol/H_2O mixtures are shown in the left panel, and the data of the corresponding ethanol/H_2O mixtures are displayed in the right panels.Parts (a) and (c) show the data for H_2O (circles), 5:95 glycerol:H_2O (diamonds), 30:70 glycerol:H_2O (crosses), 60:40 glycerol:H_2O (triangles). Parts (b) and (d) show the data for H_2O (circles), 5:95 ethanol:H_2O (squares), and 30:70 ethanol:H_2O (crosses).

The resulting thermodynamic and conformation specific spectroscopic parameters obtained from the fitting routine for all binary mixtures are all listed in Table 6.3 and 6.4. We interpreted the obtained changes of $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$ as indicative of changes of the distribution of pPII and β-strand-like conformations, which could involve changes of the coordinates of the distribution center as well as alterations of the distribution widths. However, a successful fitting of our data did not require changes of J_{β} and only rather minor changes of J_{pPII}, which seems to be at variance with the changes observed for $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$. One possible explanation of these conflicting observations is that the obtained changes of Δε solely reflect variations along the ψ-coordinate, which would not affect $^{3}J(H^{N}H^{\alpha})$. In a theoretical study by Woody et al. it was shown the positive maximum in the far UVCD spectrum of pPII indeed varies with the angle ψ (64). Additionally, we could invoke changes of the halfwidth of the Gaussian distribution along ϕ and ψ. Changing the halfwidth with respect to ϕ would not drastically change $^{3}J(H^{N}H^{\alpha})$ because the relationship between $^{3}J(H^{N}H^{\alpha})$ and ϕ is nearly linear for ϕ-values between 70° and 100° (i.e. between pPII and β-strand distributions). The Δε values in Table 6.3 suggest that even small amounts (i.e., 5%) of the two co-solvents cause significant changes of $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$. Glycerol is more effective in changing $\Delta\varepsilon_{pPII}$ than ethanol, whereas the latter has a rather strong effect on $\Delta\varepsilon_{\beta}$.

Table 6.3. Spectroscopic Parameters Obtained from Global Fitting of the Temperature Dependence of $^3J(H^NH^\alpha)$ Coupling Constants Obtained for N- and C-Terminal Amide Protons and of $\Delta\varepsilon_{215nm}$ of Cationic AAA in Binary Mixtures. $^3J(H^NH^\alpha)$ (T_R) is the room temperature experimental value.

Solvent	$^3J(H^NH^\alpha)$ (T_R)	$J_{N,pPII}$ [Hz]	$\Delta\varepsilon_{pPII}$[$M^{-1}cm^{-1}res^{-1}$]	$\Delta\varepsilon_\beta$[$M^{-1}cm^{-1}res^{-1}$]
5:95 glycerol/ H_2O	5.45	5.27	3.6	-5.10
30:70 glycerol/ H_2O	5.77	5.32	2.76	-4.93
60:40 glycerol/ H_2O	5.69	5.32	3.74	-7.78
5:95 ethanol/H_2O	5.45	5.23	2.58	-1.98
30:70 glycerol/ H_2O	5.78	5.33	3.53	-6.09

Table 6.4. Thermodynamic Values Obtained from Global Fitting of the Temperature Dependence of $^3J(H^NH^\alpha)$ Coupling Constants Obtained for N- and C-Terminal Amide Protons and of $\Delta\varepsilon_{215nm}$ of Cationic AAA in Binary Mixtures.

Solvent	$\chi_{N,\,pPII}$	$\Delta\Delta G_N$ [kJ/ mol]	$\chi_{C,\,pPII}$	$\Delta\Delta G_C$ [kJ/ mol]	$\Delta\Delta H_N$ [kJ/ mol]	$\Delta\Delta H_C$ [kJ/ mol]	$\Delta\Delta S_N$ [J/ mol*K]	$\Delta\Delta H_N$ [J/ mol*K]
5:95 glycerol/ H_2O	0.96	-0.76	0.75	-0.37	-4.80	4.38	-13.53	12.92
30:70 glycerol/ H_2O	0.89	1.69	0.69	0.36	14.8	-3.18	57.52	-0.19
60:40 glycerol/ H_2O	0.91	1.13	0.70	0.23	11.61	4.28	35.16	13.56
5:95 ethanol/ H_2O	0.95	-0.30	0.71	0.15	-18.74	7.98	-61.87	26.28
30:70 glycerol/ H_2O	0.89	1.71	0.65	0.83	9.25	8.33	25.31	25.19

The admixture of alcohol to water also has a substantial impact on the thermodynamic parameters of the system but surprisingly much less influence on the pPII propensities of the two alanine residues. The χ_{pPII} values slightly increase upon the addition of 5% alcohol and decrease to a somewhat larger extent if the alcohol content is 30% or greater. The discordance between the rather drastic changes of both ΔH and ΔS and the modest change of χ_{pPII} which we obtained even for a 5% admixture of alcohol reflects a rather interesting compensation between $\Delta\Delta H$ and $\Delta\Delta S$ (these differences are calculated by subtracting the respective value obtained in pure water from that in the binary mixture). This can be illustrated by comparing the respective values for 5% admixture glycerol. For the central alanine residue (N-terminal), we obtained $\Delta\Delta H_N = 4.8$ kJ/mol and $\Delta\Delta S_N = 13.5$ J/mol K. Hence, the gain in enthalpy (favoring pPII) upon moving from pure water to 5% glycerol is nearly compensated for by the increase of the entropic difference, which takes away nearly 4 kJ/mol at room temperature.

Surprisingly, the changes in thermodynamics determined are not monotonous with respect to increasing co-solvent fractions. As shown above in Table 6.3, even the addition of 5% alcohol causes the negative free energy difference (ΔG) to increase for both N-terminal and C-terminal residues. However, this effect becomes reversed for 30:70 alcohol water mixtures for which $\Delta\Delta G$ becomes positive. Interestingly, in a separate analysis we found that the temperature coefficients of the chemical shifts for both amide protons also show this non-monotonous behavior with respect to increasing co-solvent

concentration. Figure 6.20-21 shows the temperature dependence of the midpoint of each amide doublet as determined by averaging the two adjacent peak positions [ppm] obtained through Voigtian band deconvolution. Linear regression analysis of the resulting temperature dependent chemical shifts yields excellent correlation coefficients (<0.97). The temperature coefficients obtained through this process are plotted in Figure 6.22 as a function of solvent composition. Evidently, these temperature coefficients follow the same general non-monotonic behavior observed for the thermodynamic values with increasing solvent composition. This can be considered an independent result since thermodynamic values were obtained through ^{3}J analysis which are not necessary related to chemical shift values (i.e. ^{3}J and ppm are not generally correlated).

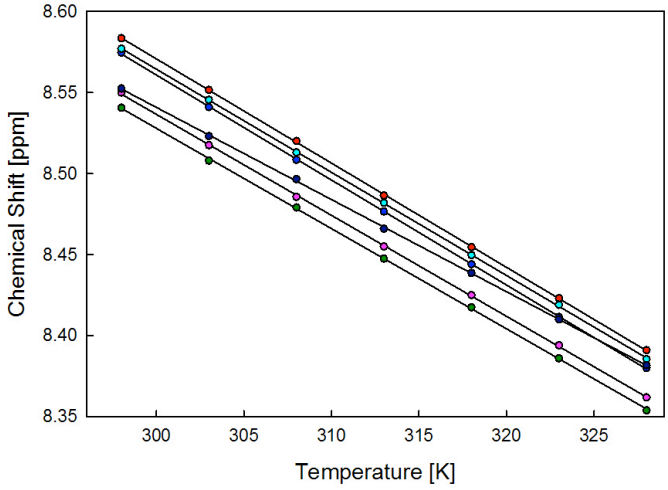

Figure 6.20: Chemical shifts [ppm] for each component of the A) N-terminal and B) C-terminal plotted as a function of temperature for AAA in all solvents. AAA in H_2O (red), 5% glycerol (blue), 5% ethanol (light blue), 30% glycerol (pink), 30% ethanol(dark blue), 60% glycerol (green).

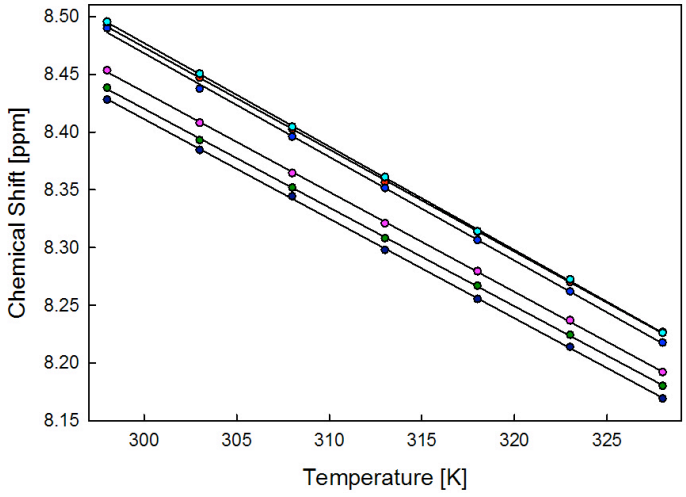

Figure 6.21: Chemical shifts [ppm] for each component of the C-terminal doublet plotted as a function of temperature for AAA in all solvents. AAA in H_2O (red), 5% glycerol (blue), 5% ethanol (light blue),30% glycerol (pink), 30% ethanol(dark blue), 60% glycerol (green).

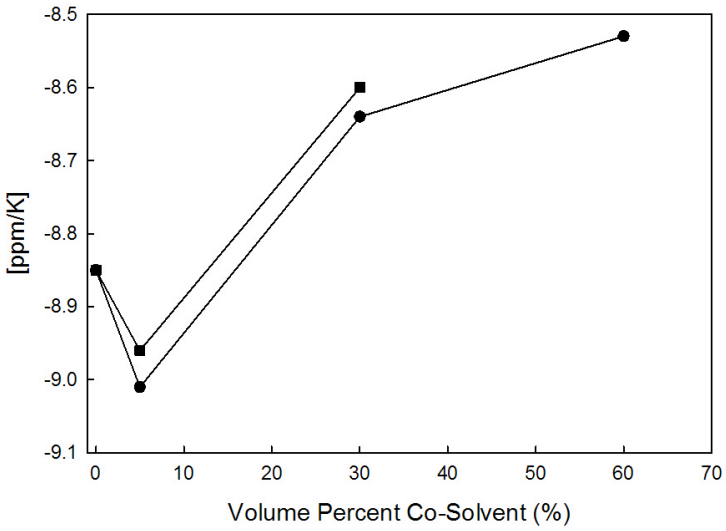

Figure 6.22: Resulting chemical shift temperature coefficients [ppm/K] for N-terminal (central residue) plotted as a function of solvent composition. Temperature coefficients in glycerol mixtures (circles) and ethanol mixtures (squares), with 0% admixture representing the H_2O solvent.6.2.2.2 Enthalpy-Entropy Compensation of Trialanine in Binary Mixtures

6.2.2.2 Enthalpy-Entropy Compensation of Trialanine in Binary Mixtures

The values for ΔΔH and ΔΔS inferred from the temperature dependence of $^3J(H^NH^\alpha)$ seem again to indicate an enthalpy-entropy compensation. To clarify whether this is characteristic for our peptide-solvent system, we plotted individual ΔH as a function of ΔS for both the N-terminal and the C-terminal residues for the milieu of binary mixtures investigated. The plots in Figures 6.23 and 6.24 suggest a linear relationship. Therefore, we fitted the afore-described eq. 6.1 to the experimental data. We obtained excellent correlation coefficients for both plots with somewhat different T_C and ΔH_C for the two residues (Table 6.4). This reflects the different peptide environment of the two residues (peptide-peptide for the N-terminal and peptide-carboxylic acid for the C-terminal residue). Further implications of the thus established enthalpy-entropy compensation will be discussed below.

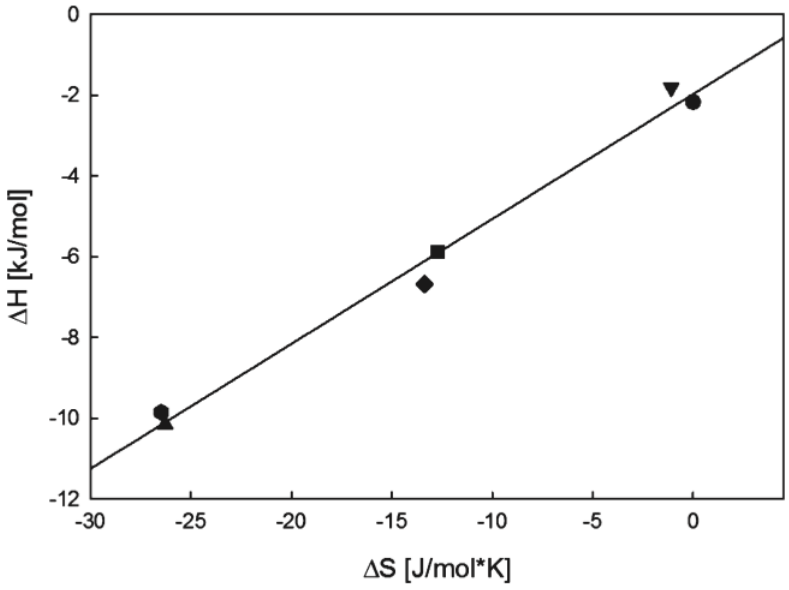

Figure 6.23. Correlation of ΔH and ΔS values obtained from a thermo- dynamic analysis of $^{3}J(H^{R}H^{N})$ coupling constants reflecting the ϕ-values of the central residue of AAA in water, glycerol/water, and ethanol/ water binary mixtures. Individual points are assignable as follows: pure H_2O (square), 5% glycerol (upside down triangle), 30% glycerol (hexagon), 60% glycerol (triangle), 5% ethanol (circle), and 30% ethanol (ellipse). The solid line results from a linear regression to the data.

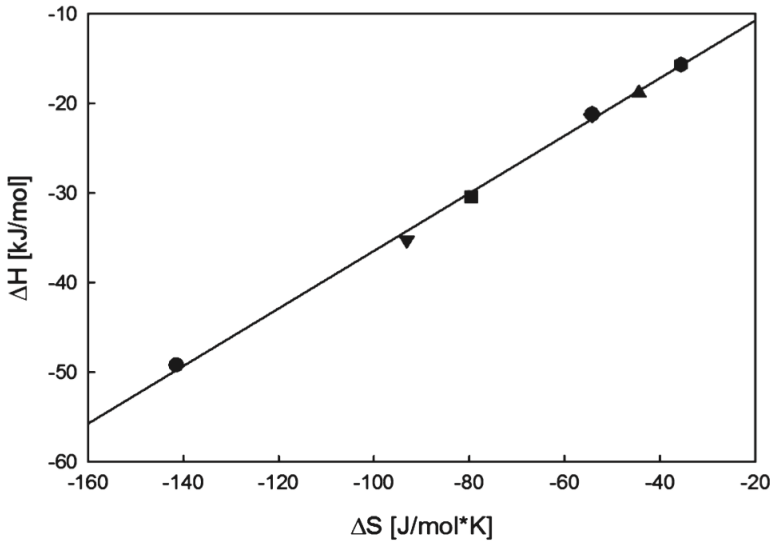

Figure 6.24. Correlation of ΔH and ΔS values obtained from a thermo- dynamic analysis of $^{3}J(H^{R}H^{N})$ coupling constants reflecting the ϕ-values of the C-terminal residue of AAA in water, glycerol/water, and ethanol/ water binary mixtures. Individual points are assignable as follows: pure H_2O (square), 5% glycerol (upside down triangle), 30% glycerol (hexagon), 60% glycerol (triangle), 5% ethanol (circle), and 30% ethanol (ellipse). The solid line results from a linear regression to the data.

Table 6.5: Results Obtained from the Linear Fit of enthalpy-entropy data

residue	ΔHo [kJ/mol]	T_R	R^2
N-terminal	-4.28	321.5	0.997
C-terminal	-1.96	309.2	0.989

Looking at the influence of cosolvent addition on the N- and C-terminal residue individually reveals that changing the alcohol content of the binary mixture affects the enthalpic and entropic contributions to the residues' Gibbs free energy rather differently. By adding 5% glycerol, for instance, the ΔG value of both residues increase. However, the enthalpy difference between central residue conformers increases, whereas it decreases for the C-terminal conformers. As indicated above, the entropy change reduces the respective changes of the Gibbs energy. This observation corroborates the notion that the residue-solvent interactions of the two residues are qualitatively different. However, the enthalpy-entropy compensation ensures that the Gibbs energies and thus also the mole fractions of pPII change in sync. Thus, our results show that similar mole fractions/ propensities of residues might conceal the full picture of the underlying thermodynamics.

6.2.2.3 Evidence for Preferential Binding

While important for any further modeling of peptide-solvent interactions, the obtained thermodynamic parameters alone do not assist in clarifying whether or not the co-solvent actually penetrates the hydration shell and interacts directly with the peptide. However, the observation that alcohol addition also changes the spectroscopic parameters $\Delta\varepsilon_{pPII}$ and $\Delta\varepsilon_{\beta}$ seems to suggest that such an interaction indeed occurs. Moreover, it is reasonable to argue that a model which considers only an indirect influence of the solvent on the peptide's energy landscape (i.e., by modulation of the hydration shell) might have difficulties to explain why small fractions of co-solvent can substantially affect enthalpy and entropy differences. These findings suggest that glycerol and ethanol are both effective in substituting water in the hydration shell of trialanine, thus even slightly stabilizing the pPII conformation when present at low concentrations. The notion of a direct interaction between co-solvent and peptide is further supported by the following argument. The model invoking an indirect co-solvent induced conformational stabilization is generally described in the literature by so-called "preferential hydration theories". Given that pPII conformations are known to be stabilized by water solvation (in fact pPII is not even a minimum on the free energy surface in vacuo (76)), preferential hydration theories would argue that the addition of co-solvent may increase water content on the surface of the peptide by preferentially excluding co-solvent from the hydration shell. However, this type of indirect interaction would lead to a monotonous relationship between co-solvent concentration and conformational change. This expectation is at

variance with our results, namely, the initial stabilization of pPII at 5% alcohol/ water mixtures followed by a shift to β-like conformations at a higher concentration co-solvent.

To investigate further the proposed direct interactions between alcohol and peptide, we measured the FT IR and Raman spectra of trialanine in water and in 5% alcohol/water (water means D_2O) mixtures. The respective spectra obtained after subtracting the respective solvent spectra are displayed in Figure 6.25.

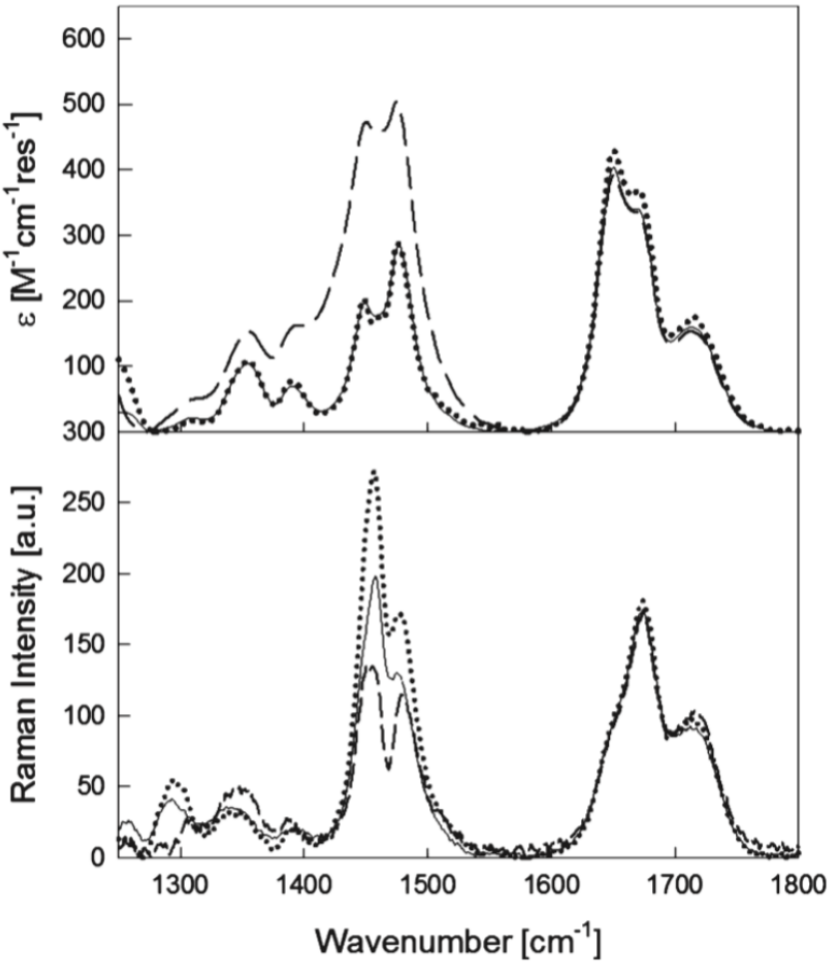

Figure 6.25. Infrared (upper panel) and Raman spectra (lower panel) of AAA in D_2O (solid line), 5% ethanol/D_2O (dotted line), and 5% glycerol/D_2O (dashed line).

As shown in Figure 6.25, co-solvent addition has only a very small influence on the amide I band profiles between 1600 cm^{-1} and 1700 cm^{-1}. This is consistent with the small change of the pPII population inferred from our CD and NMR data. However, we observed rather dramatic intensity changes of bands in the 1300 cm^{-1} - 1500 cm^{-1} region. With respect to IR, the addition of glycerol increases the intensity of nearly all bands in this region, whereas ethanol does not have a detectable effect. Both cosolvents, however, clearly affect the Raman bands particularly in the region between 1400 and 1500 cm^{-1}, which are assignable to CH_3 symmetric and antisymmetric bending modes. Bands between 1300 and 1400 cm^{-1} arise from rather complex modes with CH_3 symmetric bending and $Cas_{\alpha}H$ bending modes (in-plane and out-of-plane)n (169, 233). The substantial modification of intensity in this region of the vibrational spectra upon addition of 5% co-solvent strongly suggests that there are alcohol molecules in close enough proximity to trialanine's CH_3 group to allow for some non-bonded, van der Waals type interaction. This would be consistent with results of recent calculations of Vagenende et al. (228), who reported that the aliphatic part of glycerol can interact with aliphatic side chains. Thus, glycerol can indeed be thought of as acting as an amphiphilc interface between hydrophobic residues and the bulk water solvent. For preferential interaction of glycerol co-solvent to occur, the glycerol molecule must be oriented such that aliphatic groups point toward peptide hydrophobic groups and the three OH groups point toward

water solvent. The data in Figure 6.25 also shows that neither of the alcohols perturbs the environment of peptide's carbonyl groups.

To explain the rather strong influence of the co-solvent on IR and Raman intensities of methyl deformation modes, two options could be considered. First, one might consider drastic changes of the eigenvectors of the respective normal modes. This, however, should also cause wavenumber shifts which have not been obtained. The second option involves vibrational mixing between solvent and peptide modes, as observed for amide I and water bending modes (234). To check whether this hypothesis makes physical sense, we performed a DFT calculation for an alanine dipeptide-glycerol complex. The glycerol molecule was located close to the aliphatic CH_3 group of the central alanine residue. We used the canonical PPII coordinates ($\varphi = 70$, $\psi = 150$) as a starting point for a structural optimization of this peptide-solvent complex on a B3LYP6-31g** level of theory. The calculation was performed with the TITAN software of Schrödinger, Inc. The optimized structure was an extended β-strand with $\phi = 161$ and $\psi = 166$. This result shows again that the pPII conformation is not a dominant conformation in the absence of water. A normal mode calculation for the optimized structure clearly revealed vibrational mixing between the symmetric CH_3 bending mode and CH as well as the OH bending mode of glycerol. This is visualized in Figure 6.26, which compares the equilibrium structure of an alanine dipeptide-glycerol complex with a snapshot representing the maximal amplitude of the mixed methyl-glycerol CH deformation mode.

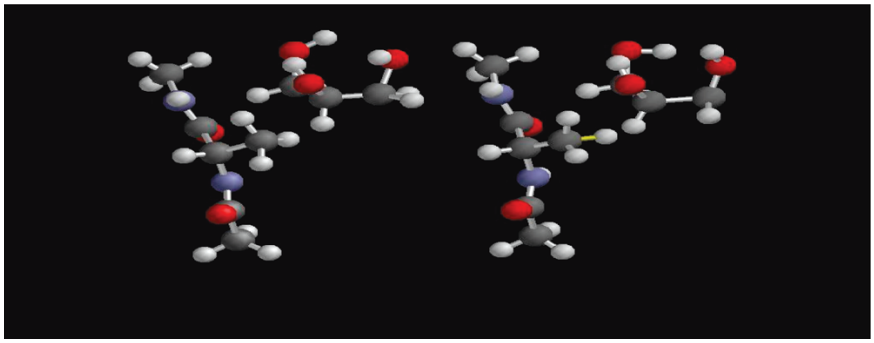

Figure 6.26. DFT calculated 1418 cm^{1} bending mode of a glycerol-alanine dipeptide complex. The right is the resting equilibrium state, and the left exhibits the maximal amplitude of the vibration.

6.2.3 Conclusion

After having provided various lines of evidence for a direct interaction between co-solvent and peptide molecules, we are now in the position to propose a model that could qualitatively describe the thermodynamic parameters derived from our data. The non-monotonous behavior observed for ΔH, ΔS, and the temperature coefficients of the chemical shifts with respect to their changes with increasing mole fractions of the employed co-solvents suggests different modes of peptide-alcohol interactions at low and intermediate concentrations of the latter. Since we are primarily interested in elucidating conformational propensities of nonterminal residues, we confine ourselves to considering only the data for the central residue. It seems to be reasonable to attribute these different steps to subsequent “binding” of alcohol molecules to the peptide. The term binding is

justified as long as it can be assumed that nonbonding interactions are operative. The first binding of an alcohol must be slightly stronger in pPII-like conformations. This provides the observed increase of the enthalpic stabilization of these conformations. The interaction potential for peptide co-solvent pairs is describable by an anharmonic potential (205). The corresponding vibrational energy levels are more widely spaced for the stronger bond. Thus, stronger binding to peptides in pPII conformations further increases the entropy difference between pPII and β-strand, as obtained. Higher alcohol concentrations facilitate the binding of a second and even third molecule, which, for, as of yet unknown reasons, is stronger in the β-strand conformation. This reduces the enthalpy difference between pPII and concomitantly decreases the entropy difference between pPII and β-strand by adding more entropy to the former than to the latter. This binding model explains the behavior of ΔH and ΔS as a function of the co-solvent fraction and the above established enthalpy-entropy correlation.

The broader conclusion that can be taken from these results on co-solvation, is that water solvation is indeed closely linked to the pPII preference of alanine. This is in agreement with studies discussed in *Chapters 5 of this thesis,* where it was shown that destabilization of the hydration shell is correlated with decreased pPII content in alanine residues.

CHAPTER 7: NEAREST NEIGHBOR INFLUENCES

Contributions: 2DNMR work done within this chapter was primarily conducted by S. Toal with contributions from Dr. Christian Richter, at Johann Wolfgang Goethe University, Schwalbe Research Group, Frankfurt Germany. Vibrational Analysis performed by S. Toal, Drexel University.

7.1 INTRODUCTION

The random coil model is based on the isolated pair hypothesis (IPH) which stipulates that the conformational space sampled by a residue in a polypeptide chain does not depend on the conformation or identity of its nearest-neighbors (22). If the IPH was indeed valid, the conformational distributions obtained of individual GxG model peptides could be used to predict conformations of ideal entropic IDPs in the absence of any local order stabilizing non-local interactions. However, computational and theoretical evidence have led a majority of researchers to question the IPH (151, 152, 155, 235, 236). The IPH hypothesis and evidence against it is discussed in detail in *Chapter 1.2.8 of this thesis*. Most notable evidence against IPH includes MD and bio-informatic work conducted by the Sosnick group who provided clear and residue specific evidence for nearest-neighbor (NN) interactions (107, 149, 154, 236). For instance, Jha et al. examined the conformational distributions of residues obtained by analysis of restricted coil libraries (i.e. all stable structures eliminated) as a function of NN identity (149). To increase the sampling size of the library while still retaining the dominant steric effects, the authors separated NNs into three groups according to the main steric properties of their side-chains: (a) β-branched side-chains (Val, Ile, and Thr); (b) aromatic side-chains (Trp, Phe, Tyr, and His); and (c) the remaining called alanine-like (except Gly and Pro). The authors

found strong nearest neighbor effects on individual conformational ensembles depending on NN type. Figure 7.1 shows this effect for alanine, where the Ramachandran plots for this residue flanked by the different residue classes are markedly different from a similar plot obtained by averaging over all potential neighbors. The biggest effect is seen when alanine is flanked by β-breached residues, in which the pPII propensity of alanine is reduced in favor of a more balanced extended conformation sampling and increased helical content.

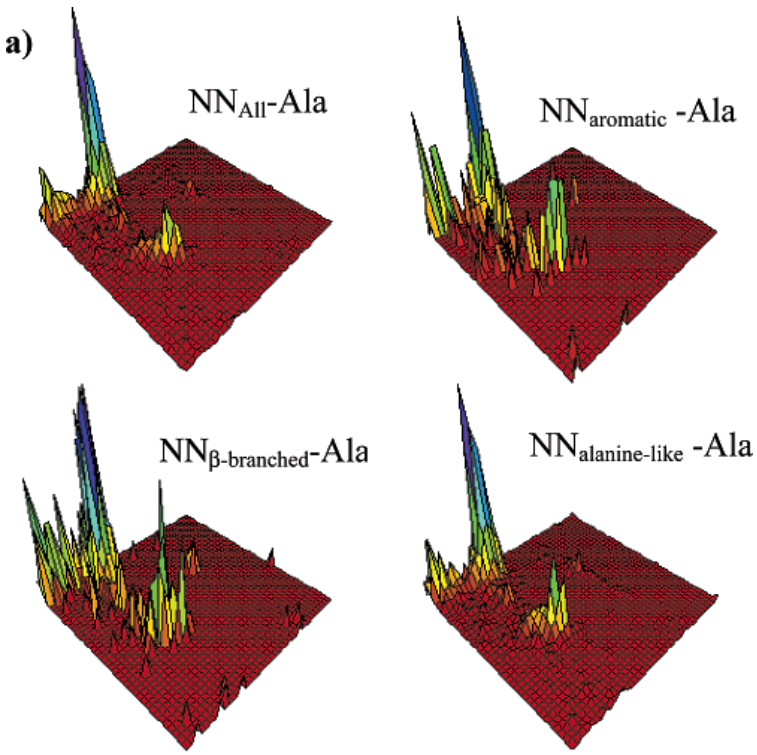

Figure 7.1 Ramachandran map for alanine from the coil library, averaged over all neighbors (except Pro and Gly), for cases in which the upstream residue is *β*-branched (Val, Ile, Thr), aromatic (His, Trp, Phe, Tyr), or the rest (alanine-like). Taken from ref (149) and modified.

Coil-libraries provide valuable information about conformational propensities of amino acid residues. However, in order to obtain the "restricted" libraries, which more accurately reflect conformation in the absence of stable secondary structures, one is sometimes left with only a limited set of data points (153). This is a problem for analyzing specific NN interactions via coil libraries as the number of occurrences of a particular NN motif is substantially reduced compared to averages over all neighbors. In addition, experimental intrinsic (i.e. in GxG motifs [55]) propensities and those obtained from coil libraries show marked differences. In a recent study, we compared experimental and coil library obtained conformational ensembles for the center residues in GYG, GFG, GIG, GVG, GRG, and GEG, which represent aromatic, aliphatic, and charged residues (153, 198). We found that coil library ensembles of corresponding tripeptide segments have a much larger fraction of type II'/type I/type III β turn conformations than distributions sampled by these tripeptides in water. These conformations are close to the trough of right-handed helical conformations located about the lower border of the upper left quadrant of the Ramachandran plot. The increased turn populations in restricted coil libraries suggest that even amino acid residues in coil regions of a protein experience non-local interactions that shift their distribution towards turns. Interestingly, we found this increase to be concomitant with asymmetric changes of the respective pPII/β-stand

populations implying that these non-local intra-protein interactions do not solely stabilize turn/helix conformations but in addition selectively destabilize either pPII or β. The thus obtained discrepancies between amino acid residue propensities in short peptides and in restricted coil libraries suggest that the latter do no constitute a truly ergodic canonical ensembles which are representative of corresponding peptide fragments in solution. Instead, these libraries likely reflect an ensemble of polypeptides subjected to statistical distribution of non-local interactions energies. Therefore, the use of coil libraries for the investigation of conformational propensities in the unfolded state and specifically the change in propensities with NN interactions is problematic.

Experimental results on nearest-neighbor interactions in short peptides are still rather limited in number. We previously discussed (see *Ch 5 of this thesis*) that alanine neighbors in AAA stabilize pPII of the central alanine residue at the expense of β-strand conformations. It has also been found that the high β-strand propensity of valine obtained by others (232), and us in the GVG model peptide (198), is enhanced by flanking valine residues in trivaline (VVV) (70). Similarly, Pizzanelli et al. has shown that the pPII fraction of phenylalanine in the GFG motif increases from 42% to 57% when flanked instead by alanine (i.e. in the AFA peptide) (237). In addition, we recently determined the conformational ensemble of ionized trilysine (KKK) and protonated as well as ionized triaspartic acid (DDD) and compared them with the respective GxG model (202, 238). The obtained conformational populations for the central residues for these peptides are all shown in Figure 7.2. Valine as terminal neighbors significantly increases the β-strand

content of valine from 46% to 68% mostly at the expense of pPII which concomitantly decreases from 50% to 16%. For trilysine, the two neighbors stabilize what Verbaro et al. (238) called a distorted pPII conformation with φ=-90 and ψ = -170, whereas K adopts a much more balanced pPII/β strand in a glycine context. Duitch et al. (202) found that ionized DDD contains practically no pPII, approximately 30% β-strand conformations, 30% right-handed helical conformations, and 40% turn-like conformations. Upon side-chain protonation, extended conformations dominate with the distribution containing equal fractions of pPII and β-strand without any detectable turn-like conformation. In GDG, however, the protonated state shows more pPII (though β is still dominant) and less helical/turn like conformations. In the ionized state, pPII is slightly dominant (59%), coexisting with 41% of β-strand. Hence ionized D as neighbor stabilizes β-strand populations of aspartic acid. These results clearly reveal strong nearest neighbor effects, which are apparently not limited to side chains with branched residues.

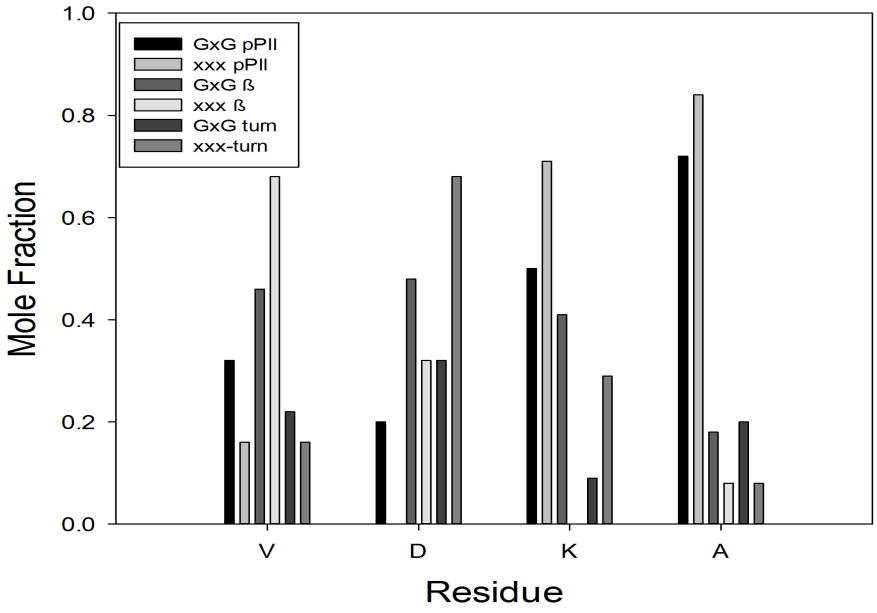

Figure 7.2: Representation of the molar fractions of amino acid residues indicated on the abscissa in GXG and XXX peptides. The code for the bars are defined in the inset of the figure.

It has been argued that the nearest-neighbor effect is mostly solvent mediated. Avbelj et al., for instance, calculated the electrostatic solvation free energy (ESF) for a model peptide system, acetyl-A_4XA_4-amide, where X is any amino acid (152). Electrostatic solvation results from the interaction between water and the partial charges on the peptide NH and CO groups. The authors found that substitution of the X residue

from alanine to valine causes reduction in ESF values of nearby peptide groups and and as a consequence destabilizes pPII compared with β. Figure 7.3 shows the change in ESF in both pPII and β for all amino acid residues. Taken together these results indicated the NN effects are linked to peptide solvation and depend on the type of NN as well as the its conformation.

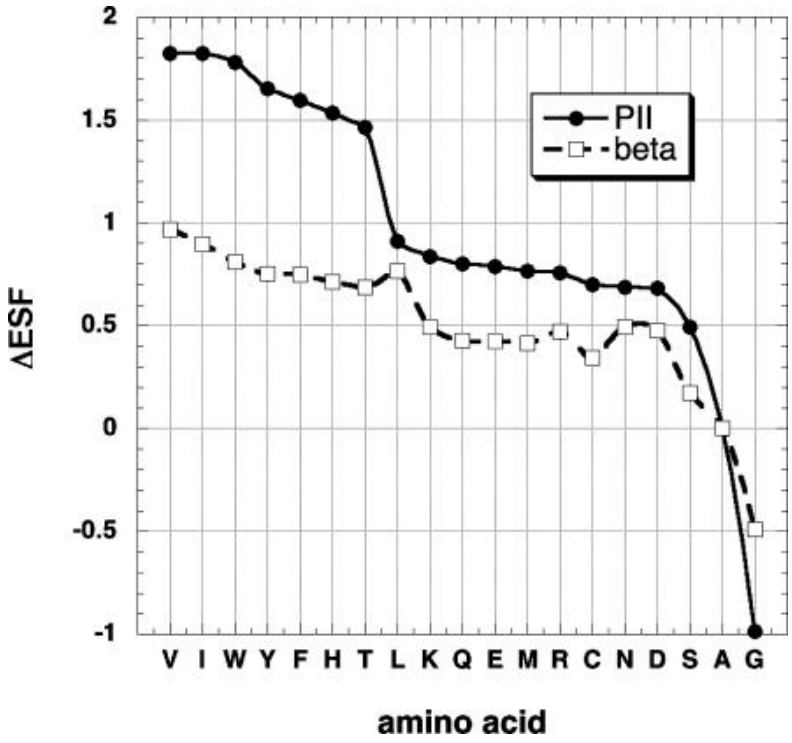

Figure 7.3: The overall change in ESF (ΔESF, kcal/mol) produced by an amino acid substitution in an (Ala)9 peptide plotted against amino acid type.

The above theoretical, bio-informatic and, in part, experimental results reveal a breakdown of the IPH in the unfolded state. Therefore, contrary to the random coil model, we cannot consider amino acid residues as isolated from their peptide environment. To move closer towards experimentally based structural predictions of IDPs the intrinsic preferences of amino acid residues must be complimented by experimental information on nearest-neighbor interactions. To this end, more systematic experimental investigations on NN effects are necessary. In order to investigate NN effects in the unfolded state we have conducted a combined 2D-NMR and vibrational study on selected "GxyG" host guest peptides: GDyG, GSyG, GxLG, GxVG, where x/y={A, K, L, V}. Aspartic acid and serine were chosen at the x-position due their noted ability to drastically change the distribution of alanine in xAy peptides in coil libraries. Lysine and valine were similarly chosen at the y-position as these amino acids have the largest effect on the C-terminal end of the peptide in the coil libraries of Sosnick et al. (153). The GxyG motif allows for determination of the conformational ensembles of both x and y residues simultaneously. We employed an algorithm based on excitonic coupling in order to simulate the amide I' band profiles of IR, isotropic and anisotropic Raman, and VCD as well as a full set of six conformationally sensitive J-coupling constants. All peptides were uniformly backbone labeled with ^{15}N and ^{13}C in order to obtain hetero-nuclear coupling constants: $^{3}J(H^{\alpha},C')$, $^{3}J(H^{N},C')$, $^{3}J(H^{N},C^{\alpha})$, $J(H^{N},C^{\beta})$, which depend on the angle ϕ, $J(NC^{\alpha})$ and $^{2}J(NC^{\alpha})$, which depend on ψ. The theoretical algorithm which combines NMR and vibrational parameters to extract conformational distributions is similar to that

employed for investigation of the intrinsic propensities of x-residues in GxG or homo-tripeptides (i.e. AAA) however with an extension over the three local amide I oscillators of GxyG. The theoretical formalism for this approach has been described in detail in *Chapter 2 of this thesis*. In addition to determining the conformational distributions of each guest residue of the selected GxyG peptides, the thermodynamics underlying the pPII⇔β equilibrium were also examined. To obtain enthalpic and entropic contributions to Gibbs free energy the temperature dependence of $^{3}J(H^{H}H^{C\alpha})$ coupling constant was analyzed in terms of a pseudo-two state model. In the following sections, we first present conformational distributions obtained from the combined vibrational and 2D-NMR analysis. We then dissect these results in terms of the trends of NN interactions according to amino acid type. Finally, we present the results of our corresponding thermodynamic analysis and discuss how these point to the role of solvent mediation in NN effects.

7.2 RESULTS

7.2.1 Vibrational and 2D-NMR Spectroscopic Results

For illustration, Figure 7.4 shows the experimentally obtained amide I' IR, isotropic and anisotropic Raman, and VCD profiles for a representative set of the total 14 investigated GxyG peptides. The solid lines in Figure 7.4 reflect the result of optimized amide I' simulation according to the derived excitonic coupling algorithm using 2D Gaussian conformational distributions as free parameters (see *Chapter 2.3 of this thesis*).

In order to simulate these bands and obtain structural information, in a first step we started with the known intrinsic conformational ensemble of the corresponding amino acid residue obtained from GxG model peptides (54, 55). For instance, in the GDAG peptide, the conformational ensemble of GDG and GAG were used as starting points in the simulation. If the resulting fit of amide profiles and NMR constants was poor, in a second step, we began to incrementally modify the ensembles to more closely reflect those reported in the NN specific tripeptide coil libraries of Sosnick et al. (153). For the GDAG peptide this means obtaining information on GDA and DAG from restricted coil library motifs. Within this step, special care was given to making small adjustments to the ensembles while interactively checking the resulting simulated amid I' profiles and NMR coupling constants for reproduction quality. In some cases, it was necessary to implement distributions that were not present in coil library tripeptide motifs. This is not surprising, as discussed earlier, these libraries constitute a qualitative look at NN interactions but are not sufficiently robust to capture quantitative differences in ensembles. As a final step, the distribution parameters were fine tuned to minimize the root mean square deviation between the six experimental and simulated J-coupling constants. As shown in Figure 7.3, all simulated band profiles are in good agreement with experiment. Table 7.1A-D lists the J-coupling constants obtained from various 2D NMR experiments as well as the corresponding optimized simulated J-coupling constants.

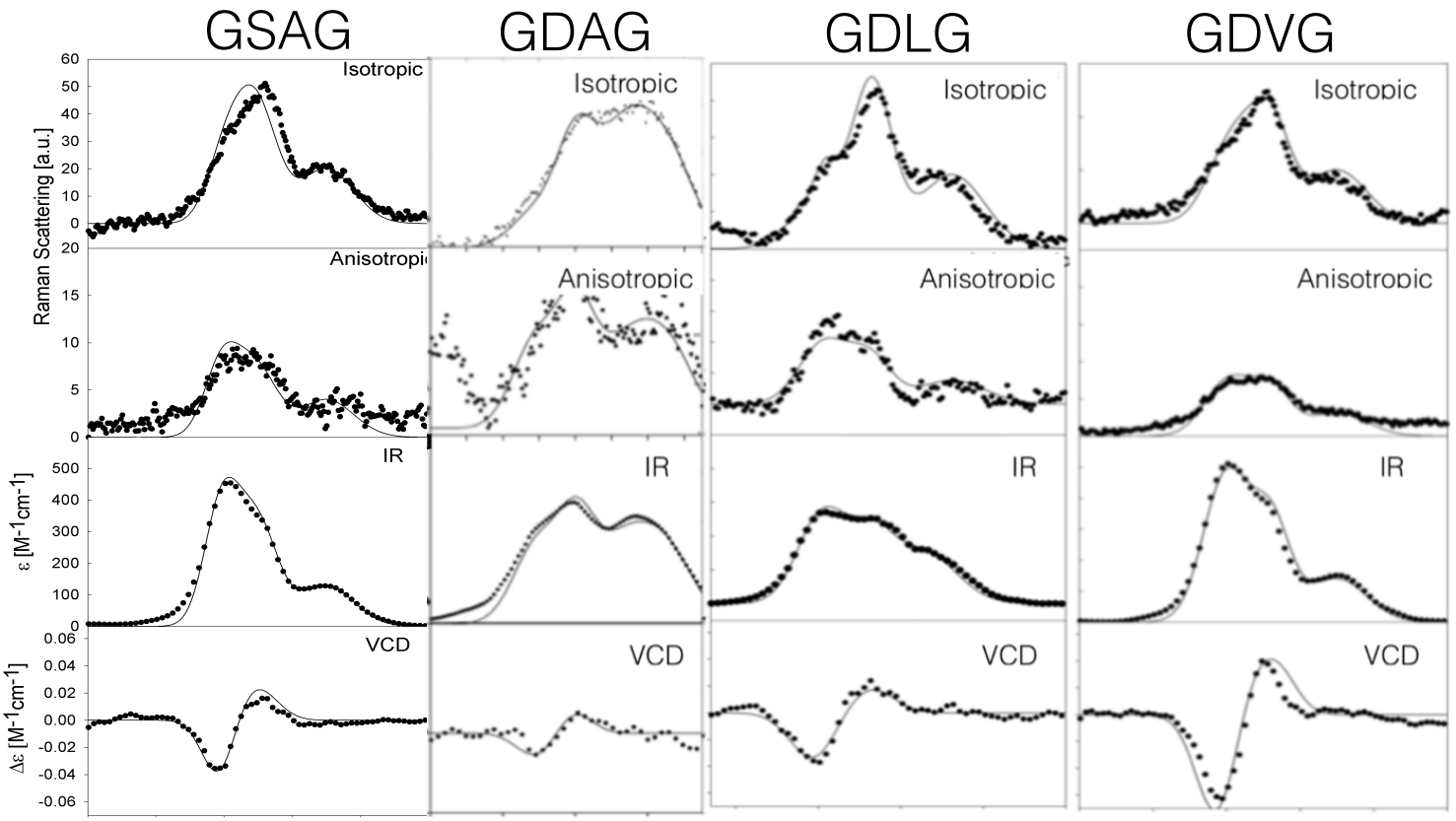

Figure 7.4: Isotropic Raman, Anisotropic Raman, IR, and VCD profiles in the amide I' region for noted GxyG peptides

Table 7.1A: Simulated (upper sub cell) and experimental (lower sub cell) NMR derived J coupling constants for all investigated residues within the chosen GDyG peptides.

Method	J-coupling	GDAG		GDKG		GDLG		GDVG	
		D :	A:	D :	K:	D :	L:	D :	V :
		sim	sim	sim	sim	sim	sim	sim	sim
		(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)
^{1}H-NMR	$^{3}J(H^{N}H^{\alpha})$ [Hz]	7.18	6.62	6.25	6.77	6.90	6.81	6.59	8.23
		7.22 ±0.01	6.63 ±0.04	6.19 ±0.2	6.63 ±0.28	6.97 ±0.04	6.78 ±0.01	7.08 ±0.06	8.25 ±0.06
HNC$^{\alpha}$-COSY	$^{3}J(H^{N}C^{i})$ [Hz]	1.08	1.24	0.72	1.1	0.92	0.90	0.91	0.81
		0.97 ±0.06	1.2 ±0.08	0.5 ±0.07	1.03 ±0.3	0.9 ±0.03	0.84 ±0.09	0.87 ±0.04	0.81 ±0.12
CO-coupled (H)NC$^{\alpha}$H$^{\alpha}$	$^{3}J(H^{\alpha}C^{i})$ [Hz]	3.01	2.50	2.82	2.2	2.65	2.45	3.00	2.80
		3.15 ±0.9	2.19 ±0.2	2.7 ±0.17	2.09 ±0.17	2.75 ±0.04	2.45 ±0.03	3.0 ±0.15	2.8 ±0.08
(C$^{\beta}$)-E.Cosy	$^{3}J(H^{N}C^{\beta})$ [Hz]	1.50	1.61	1.53	1.6	1.52	1.60	1.70	1.30
		1.54 ±0..5	1.98 ±0.08	1.47 ±0.06	1.54 ±0.06	1.57 ±0.05	1.66 ±0.02	1.54 ±0.08	1.28 ±0.05
J- Mod HSQC	$^{1}J(NC^{\alpha i})$ [Hz]	11.75	11.19	12.2	10.9	11.30	10.91	11.75	11.19
		11.81 ±0.13	11.37 ±0.11	11.93 ±0.12	11.27 ±0.18	11.34 ±0.11	10.58 ±0.1	11.9 ±0.12	11.23± 0.18
J- Mod HSQC	$^{2}J(NC^{\alpha i})$ [Hz]		8.50		8.70		n/a		8.80
			8.47 ±.16		8.74 ±0.17		n/a		8.83 ±0.18

Table 7.1B: Simulated (upper sub cell) and experimental (lower sub cell) NMR derived J coupling constants for all investigated residues within the chosen GSyG peptides.

Method	J-coupling	GSAG		GSKG		GSLG		GSVG	
		S :	A :	S :	K:	S :	L:	S :	V:
		sim	sim	sim	sim	sim	sim	sim	sim
		(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)
^{1}H-NMR	$^{3}J(H^{N}H^{\alpha})$ [Hz]	7.4	6.2	6.6	7.2	7.3	7.3	6.9	8.2
		6.86 ±0.07	6.35 ±0.03	6.62 ±0.04	7.22 ±0.03	6.86 ±0.02	7.34 ±0.01	6.95 ±0.07	8.43 ±0.04
HNC$^{\alpha}$-COSY	$^{3}J(H^{N}C^{i})$ [Hz]	1.0	0.8	0.9	n/a	1.0	0.8	0.9	0.8
		0.97 ±0.1	0.79 ±0.05	0.88 ±0.26	n/a	0.44 ±0.06	0.66 ±0.11	0.87 ±0.05	0.76 ±0.05
CO-coupled (H)NC$^{\alpha}$H$^{\alpha}$	$^{3}J(H^{\alpha}C^{i})$ [Hz]	3.2	3.2	3.3	3.2	2.4	2.4	3.3	3.1
		3.26 ±.03	3.2 ±0.07	3.52 ±0.03	3.52 ±0.03	2.48 ±0.03	2.59 ±0.04	3.52 ±0.03	3.52 ±0.03
(C$^{\beta}$)-E.Cosy	$^{3}J(H^{N}C^{\beta})$ [Hz]	1.8	2.0	1.8	1.6	1.5	1.6	1.9	1.5
		1.79 ±.0.06	2.05 ±0.06	1.86 ±0.06	1.54 ±0.13	1.79 ±0.67	1.60 ±0.06	1.92 ±0.06	1.41 ±0.13
J- Mod HSQC	$^{1}J(NC^{\alpha i})$ [Hz]	12.2	10.9	12.2	11.0	12.1	10.8	12.2	11.3
		12.2 ±0.12	11.3 ±0.15	12.3 ±0.13	11.3 ±0.18	12.2 ±0.12	11.15 ±0.02	12.2 ±0.09	11.3 ±0.2
J- Mod HSQC	$^{2}J(NC^{\alpha i})$ [Hz]		8.1		8.4		8.2		8.2
			8.04 ±0.16		8.38 ±0.17		8.21 ±0.15		8.39 ±0.20

Table 7.1C: Simulated (upper sub cell) and Experimental (lower sub cell) NMR derived J coupling constants for all investigated residues within the chosen GxLG peptides.

Method	J-coupling	GALG		GKLG		GLLG		GVLG	
		A :	L:	K :	L:	L1 :	L2:	V :	L :
		sim	sim	sim	sim	sim	sim	sim	sim
		(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)	(exp)
^{1}H-NMR	$^{3}J(H^{N}H^{\alpha})$ [Hz]	6.75	7.36	6.40	7.00	7.60	8.85	8.46	8.10
		6.8 ±0.1	7.4 ±0.11	6.44 ±0.06)	7.1 ±0.05	7.62 ±0.09	9.08 ±0.07	8.46 ±0.17	8.15 ±0.16
HNC$^{\alpha}$-COSY	$^{3}J(H^{N}C^{i})$ [Hz]	1.30	0.85	1.01	0.75	0.80	0.90	0.75	0.63
		1.36 ±0.17	0.53 ±0.06	0.96 ±0.13	0.51 ±0.06	0.77 ±0.06	0.84 ±0.06	0.77 ±0.3	0.58 ±0.06
CO-coupled (H)NC$^{\alpha}$H$^{\alpha}$	$iJ(H^{\alpha}C^{i})$ [Hz]	2.00	2.32	2.22	2.39	3.20	3.44	2.50	2.40
		2.08 ±0.19	2.29 ±0.29	2.20 ±0.03	2.37 ±0.03	3.52 ±0.03	3.52 ±0.03	2.53 ±0.03	2.37 ±0.06
(C$^{\beta}$)-E.Cosy	$iJ(H^{N}C^{\beta})$ [Hz]	2.30	1.75	1.69	1.72	1.52	1.80	1.60	1.49
		2.29 ±0.02	1.66 ±0.08	1.73 ±0.3	1.73 ±0.06	1.54 ±1.73	1.73 ±0.19	1.54 ±0.06	1.54 ±0.13
J- Mod HSQC	$^{1}J(NC^{\alpha i})$ [Hz]	11.30	11.34	11.90	11.54	10.90	11.20	11.10	11.45
		11.32±0.13	11.4 ±0.16	11.44 ±0.07	11.26 ±0.18	10.65 ±0.06	11.16 ±0.25	11.07 ±0.03	11.24 ±0.20
J- Mod HSQC	$^{2}J(NC^{\alpha i})$ [Hz]		8.35		8.50		8.41		8.21
			8.29 ±0.19		8.55 ±0.15		8.43 ±0.21		8.23 ±0.18

Table 7.1D: Simulated (upper sub cell) and Experimental (lower sub cell) NMR derived J coupling constants for all investigated residues within the chosen GxVG peptides.

Method	J-Coupling	GAVG		GKVG		GLVG	
		A :	V:	K:	V:	L :	V:
		sim	sim	sim	sim	sim	sim
		(exp)	(exp)	(exp)	(exp)	(exp)	(exp)
^{1}H-NMR	$^{3}J(H^{N}H^{\alpha})$ [Hz]	6.60	7.19	7.24	7.40	7.52	7.78
		6.6 ±0.12	7.15 ±0.09	7.21 ±0.12	7.42 ±0.13	7.49 ±0.09	7.77 ±0.15
HNC$^{\alpha}$-COSY	$^{3}J(H^{N}C^{i})$ [Hz]	1.10	1.08	0.90	1.00	86.00	0.75
		1.27 ± 0.03	1.16 ±0.06	0.87 ± 0.01	0.99 ± 0.04	0.81 ± 0.03	0.68 ±0.03
CO-coupled (H)NC$^{\alpha}$H$^{\alpha}$	$^{3}J(H^{\alpha}C^{i})$ [Hz]	2.50	2.29	2.70	2.10	2.70	2.32
		2.56 ±0.14	2.35 ±0.17	2.72 ± 0.04	2.06 ± 0.05	2.68 ±0.19	2.36 ±0.12
(C$^{\beta}$)- E.Cosy	$^{3}J(H^{N}C^{\beta})$ [Hz]	2.21	1.56	1.91	1.54	1.99	1.65
		2.4 ± 0.11	1.5 ±0.16	1.84 ± 0.03	1.52 ±0.11	1.936 ±0.02	1.52 ± 0.064
J- Mod HSQC	$^{1}J(NC^{\alpha i})$ [Hz]	11.43	11.29	11.33	11.55	11.10	11.31
		11.45 ±0.17	11.33±0.16	11.23 ±0.11	11.46 ±0.19	10.98 ±0.09	11.26 ±0.19
J- Mod HSQC	$^{2}J(NC^{\alpha i})$ [Hz]		8.39		8.58		8.50
			8.4 ±0.15		8.61 ±0.15		8.52 ±0.21

The amide I' profiles shown in Figure 7.4 look qualitatively similar with respect to their intensity distributions, but differences are noteworthy. Most of the investigated peptides exhibit some degree of non-coincidence between the isotropic Raman and IR profiles. In general, the high frequency band of amide I' profiles in isotropic Raman are more intense, whereas in the IR profile intensity is redistributed such that the low frequency band is the most intense. This type of non-coincidence is characteristic of a dominant sampling of extended conformations (i.e., pPII and β-strand) associated with the upper left quadrant of the Ramachandran plot. The origin of this re-distribution and non-coincidence for ensembles dominated by extended structures (i.e. pPII and β-strand) is the large strength of excitonic coupling in this region of the Ramachandran plot as shown in Figure 2.7 (176). Excitonic coupling, its dependence on dihedral angles, and its effect on the shape of amide I band profiles has been discussed in detail in *Chapter 2.3 of this thesis*. The largest non-coincidence was obtained for for GALG and GKLG for which the derived conformational ensembles were found to have large percentages of extended conformations. For instance, the lysine and leucine residues in GKLG have 90% and 95% extended populations, respectively. In contrast, peptides with conformational distributions dominated by turn-like structures, have much less redistribution in intensity resulting in coincidence between Raman and IR amide I profiles. This coincidence is typical for residues that sample regions of the Ramachandran space outside of the extended region. For instance, GDAG and GDVG peptides both have large percentages of turn-like populations and as a result exhibit the largest coincidence between band

profiles. Table 7.2a-d lists the resulting conformational distributions obtained for each x and y amino acid residue within all investigated GxyG peptide. The VCD signal emerges again as the most sensitive indicator of structural differences. While it displays a negative-positive couplet indicative of significant pPII population for most of the investigated peptides, the intensity of this couplet varies markedly among different peptides. In general, the series of alanine containing peptides GDAG, GSAG, GALG, and GAVG exhibit the largest VCD signals largely due to the high pPII content of alanine. GALG shows the largest observed VCD signal due to the high pPII content of alanine (65%) and, to a lesser extent, leucine (42%) within the peptide.

Table 7.2A: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for aspartic acid in the GDyG .peptide series. Upper subcells are (ϕ, ψ) positions, lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

Peptide	pPII	β	extended total	i+2 I/II'β turn	asx turn
GDG	(-78, 175)	(-125,175)		(-50, 0)	(60, 110)
	0.20	0.48	0.68	0.09	0.23
GDAG	(-78, 175)	(-132, 175)		(-50, 0)	(65, 150)
	0.24	0.48	0.72	0.09	0.19
GDKG	(-75, 150)	(-130, 145)		(-50, 0)	(60, 115)
	0.40	0.45	0.85	0.08	0.07
GDLG	(-75, 145)	(-130, 145)		(-50, 0)	(65, 130)
	0.4	0.43	0.83	0.07	0.10
GDVG	(-78, 150)	(-130, 145)		(-50, 0)	(65, 130)
	0.45	0.4	0.85	0.07	0.10

Table 7.2B: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for Serine in the GSyG .peptide series. Upper subcells are (ϕ, ψ) positions, Lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

Peptide	pPII	β	extended total	i+2 I/II'β turn	i+2 I'/II β turn	asx turn
GSG	(-79,165)	(-103,165)		(-50, 0)	(70, 0)	(70, 160)
	0.42	0.33	0.75	0.10	0.05	0.10
GSAG	(-75, 152)	(-118, 150)		(-50, -0)	(70, 0)	(70, 160)
	0.37	0.40	0.77	0.1	0.06	0.07
GSKG	(-75, 165)	(103, 165)		(-50, -15)	(70, 0)	(75, 145)
	0.31	0.44	0.75	0.1	0.05	0,04
GSLG	(-74, 155)	(-132,155)		(-50, -15)	(70, 0)	(65,160)
	0.37	0.42	0.79	0.1	0.05	0.06
GSVG	(-70, 155)	(-110,155)	0.88	(-50, -15)		(65,160)
	0.47	0.41		0.05		0.07

Table 7.2C: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for Alanine in the GAyG and GxAG .peptide series. Upper subcells are (ϕ, ψ) positions, Lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

	pPII	β	extended total	I'/III' β turn	i+2 I/II'β turn	i+1 II β turn	ɣ-turn
GAG	(-69,155)	(-115,155)			(-60,-30)	(-60, 120)	(20, 60)
	0.72	0.18	0.9		0.03	0.03	0.04
GDAG	(-69,155)	(-115,155)					(50,-50)
	0.62	0.23	0.85		0.05		0.10
GSAG	(-69,150)	(-115,140)		(60,50)	-60,-30		(-50,50)
	0.50	0.31	0.81	0.07	0.07		0.05
GALG	(-69,145)	(-115,140)			-55,30		(40,-60)
	0.65	0.29	0.94		0.06		
GAVG	(-75, 155)	(-120,155)		(50, -50)	(-60, -30)	(-60, 100)	
	0.4	0.48	0.88	0.05	0.05	0.02	

Table 7.2D: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for Alanine in the GKyG and GxKG .peptide series. Upper subcells are (ϕ, ψ) positions, Lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

Peptide	pPII	β	extended total	I'/III' β turn	i+2 I/II'β turn	i+2 I'/II β turn	i+1 II β turn
GKG	(-66,150)	(-115,145)			(-65, 35)		
	0.50	0.41	0.91		0.09		
GDKG	(-66,148)	(-115,135)		(70, 45)	(-65, 35)		(-60, 110)
	0.46	0.42	0.88	0.03	0.05		0.04
GSKG	(-66,145)	(-125,135)		(70, 45)	(-65, 35)		
	0.42	0.49	0.91	0.06	0.03		
GKLG	(-66,145)	(-120,140)		(70, 30)	(-65, 35)		
	0.52	0.38	0.9	0.04	0.06		
GKVG	(-66,145)	(-120,140)		(75, 45)	(-65, 35)		
	0.4	0.48	0.88	0.05	0.07		

Table 7.2E: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for Alanine in the GLyG and GxLG .peptide series. Upper sub-cells are (ϕ, ψ) positions, Lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

Peptide	pPII	β	extended total	I'/III' β turn	i+2 I/ II'β turn	i+2 I'/ II β turn	i+1 II β turn	asx turn
GLG	(-76, 145)	(-98, 160)			(-50, -40)		(-60, 110)	
	0.45	0.42			0.08		0.05	
GDLG	(-70, 152)	(-115, 145)			(-50, -40)		(-65, 112)	
	0.49	0.42			0.05		0.04	
GSLG	(-70, 152)	(-120, 142)	(70, 45)	(-70, 0)	(-50, -50)			
	0.48	0.40	0.05	0.04	0.03			
GLLG	(-70, 145)	(-120, 142)			(-50, -50)		(-65, 112)	(100, 115)
	0.40	0.49			0.05		0.02	0.04
GLLG	(-76, 145)	(-98, 160)		(-70, 10)	(-50, -50)		(-65, 112)	(100, 115)
	0.42	0.42		0.02	0.04		0.07	0.03
GALG	(-70, 150)	(-120, 140)		(-70, 10)	(-50, -50)		(-65, 112)	(100, 115)
	0.42	0.46		0.03	0.03		0.04	0.02
GKLG	(-70, 142)	(-120, 150)			(-50, -50)			
	0.43	0.52			0.05			
GVLG	(-70, 145)	(-123, 142)			(-50, -50)		(-67, 112)	(100, 115)
	0.40	0.48			0.05		0.05	0.02

Table 7.2F: Conformational distributions obtained from fitting amide I' band profiles and corresponding NMR coupling constants for Alanine in the GVyG and GxVG .peptide series. Upper subcells are (ϕ, ψ) positions, Lower subcells are respective mole fractions. The corresponding amino acid is highlighted in red.

Peptide	pPII	β	βt	extended total	I'/III' β turn	i+2 I/II'β turn	i+2 I'/II β turn	i+1 II β turn	ɣ-turn	asx turn
GVG	(-80, 170)	(-120, 170)			(60, 30)	(-60, -30)			(-60, 60)	
	0.32	0.46		0.78	0.07	0.04			0.11	
GDVG	(-74, 145)	(-120, 140)				(-60, -30)				(95, 130)
	0.3	0.50		0.8		0.10				0.1
GSVG	(-74, 152)	(-120, 147)				(-60, -30)	(60, 0)			(95, 130)
	0.28	0.51		0.79		0.05	0.06			0.10
GVLG	(-74, 152)	(-120, 145)					(60, 0)		(-60, 60)	
	0.33	0.48		0.81			0.11		0.08	
GAVG	(-72, 155)	(-120, 155)				(-60, -30)			(-60, 60)	(95, 115)
	0.39	0.45		0.84		0.09			0.05	0.02
GKVG	(-75, 155)	(-115, 155)	(-110, 100)		(65, 35)	(-60, -30)			(-60, 60)	
	0.44	0.4	0.03	0.84	0.07	0.02			0.04	

For visualization purposes, Figure 7.4 shows the Ramachandran plots that correspond to the numerical conformational distributions for all residues within the GxyG motifs listed above in Table 7.2a-f. Each plot is labelled as the corresponding tripeptide motif. For instance, the GDAG tetra-peptide studied can be broken down into two tripeptide motifs for studying NN effects: (1) GDA in which we observe the effect of alanine on aspartic acid and (2) DAG in which we observe the effect of aspartic acid on alanine. These conformational ensembles can then be compared to GDG and GAG, respectively, to infer NN influence compared to a purely glycine environment. Just a qualitative look at the highly variant Ramachandran plots in Figure 7.5 underscores the notion that there are unique preferences among amino acid residues in the unfolded state, and that these preferences are indeed modified by NN interactions. Details on trends observed for specific NN interactions are discussed in detail in the following section.

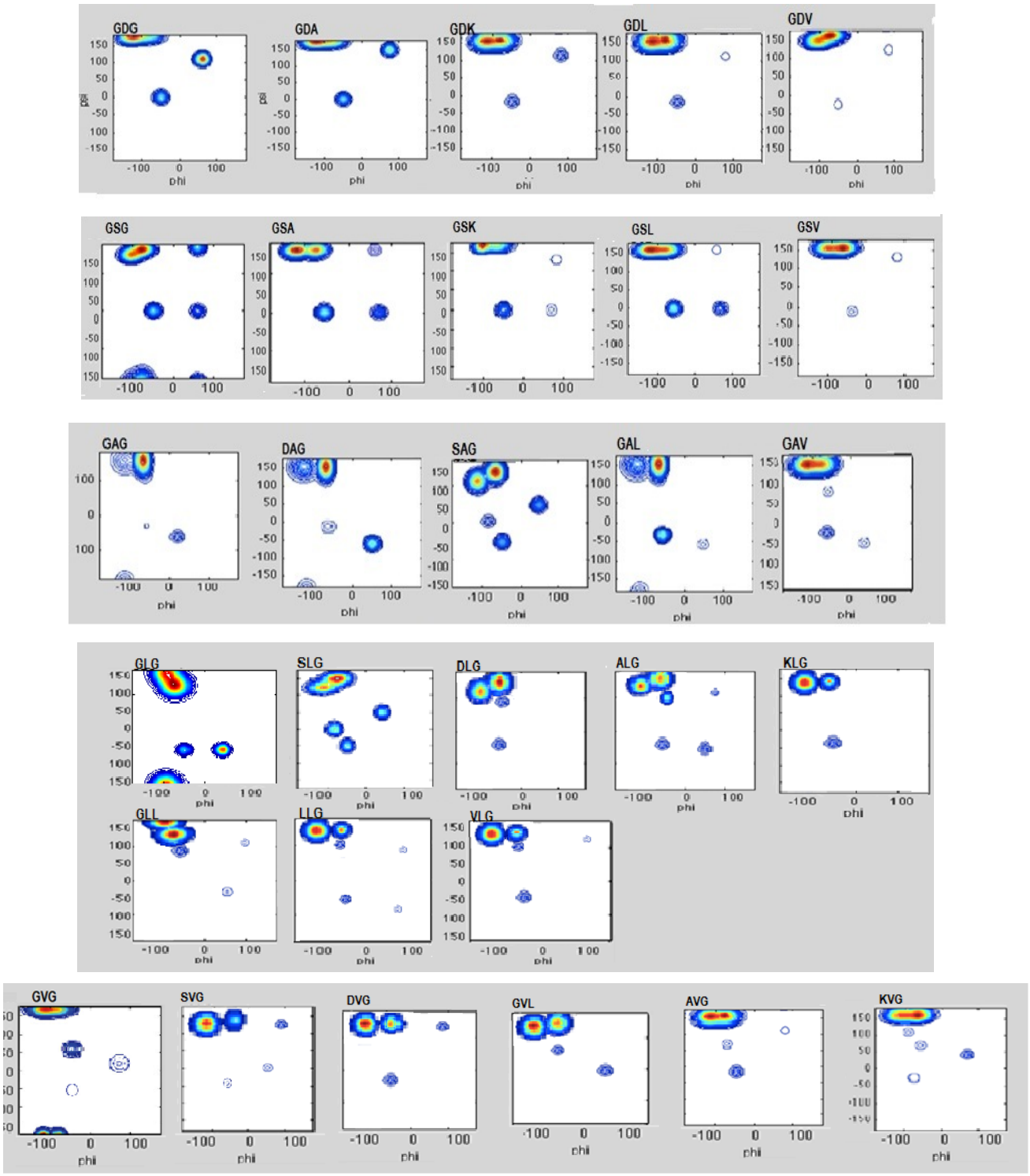

Figure 7.5: Ramachandran plots corresponding to the derived conformational ensembles for GxyG peptides, which show the neighbor effect on center amino acid in each series.

7.2.2 Trends in Nearest Neighbor Interactions

7.2.2.1 NN Effects on pPII propensity and Extended State Equilibrium of Alanine

As discussed throughout this thesis, alanine is special in its high intrinsic pPII propensity in the unfolded state as found by us using short alanine based peptides as well as by numerous others (56, 75). In GAG, alanine has been reported as having a pPII content of 72%. Previous studies have shown that alanine neighbors increase the pPII propensity of alanine. For instance, the central alanine residue in AAA was found to have a pPII content of >80%. Figure 7.6 displays the pPII and β-strand content obtained for alanine residues with different neighboring residue in GAyG and GxAG peptides and compares it to GAG. Apparently, the intrinsic pPII content of alanine (as reflected by GAG) is reduced once a non-alanine neighbor is introduced. On the contrary, a comparison of the alanine residues within G*AG, G*AAG , GA*AG, and A*AA (where the star here comes before the alanine residue of interest) reveals that alanine as an N-terminal neighbor (i.e. GA*AG) stabilizes pPII content of nearby alanines more so than alanine as a C-terminal neighbor (ie. G*AAG). In addition, we find that the conformational ensembles of GAG and GGAG are statistically similar, with GGAG losing only a minor portion of β-strand content in favor of turn-like conformations. This increase in turn content is likely the result of extra flexibility lent to the GGAG peptide by the additional glycine residue. The similarity of GAG and GGAG lends another piece of evidence to the notion that GxG peptides reflect intrinsic propensities of the x-residue and are equally adequate model systems in comparison to the longer glycine based

systems (i.e., GGxGG) for the unfolded state. Insertion of non-alanine or non-glycine residues significantly decreases the pPII content of alanine in favor of β-strand conformations. As shown in Figure 7.6, in general, as the pPII content of a peptide decreases the β-strand content systematically increases regardless of the identity of the interacting neighbors. This results in an extended state population that fluctuates only mildly (Figure 7.7). Aspartic acid and particularly serine as upfield neighbors significantly reduce the pPII content of alanine from 72% in GAG to 62% and 50%, in GDAG and GSAG, respectively. Interestingly, while the loss of pPII here is accompanied by an increase in β-strand content, an above average portion is re-distributed to turn-like conformations, which nearly double from 10% in GAG to 17% and 19% in GDAG and GSAG, respectively. This is noteworthy in light of the fact that aspartic acid and serine themselves were previously reported by us (see *Ch 1.3.2.1 of this thesis*) to have high intrinsic propensities for turn-like conformations. It seems that this preference for turns is communicated downfield to alanine in these peptides. For visualization, Figure 7.7 displays a plot that correlates changes of sub-distributions within each peptide's conformational ensemble. In contrast, if one examines the effect of hydrophilic residues valine and leucine on the conformation of alanine upfield, the loss in pPII results in an increase β-strand conformations with more minor variations in turn-like conformations. For GAVG the turn population increases from 10% to 12% (i.e., stays constant in the limit of uncertainty) whereas in GALG the turn population of alanine actually decreases to 6% in favor of a more pronounced pPII content (65%) and overall dominance of

extended conformations (94%). Interestingly, while aspartic acid and serine have the largest effect on stabilizing turn conformations, valine has the largest effect overall on the intrinsic conformational propensity of alanine, greatly reducing the pPII content to only 40% in favor of a dominance in β-strand conformations (48%). In fact, the ensemble of alanine in GAVG is the only one in which pPII is no longer the dominant conformation. This is likely due to the branched valine side-chain de-stabilizing the optimal hydration shell around alanine which allows for pPII dominance.Valine has been shown by us to have a high β-strand population in a purely glycine environment.

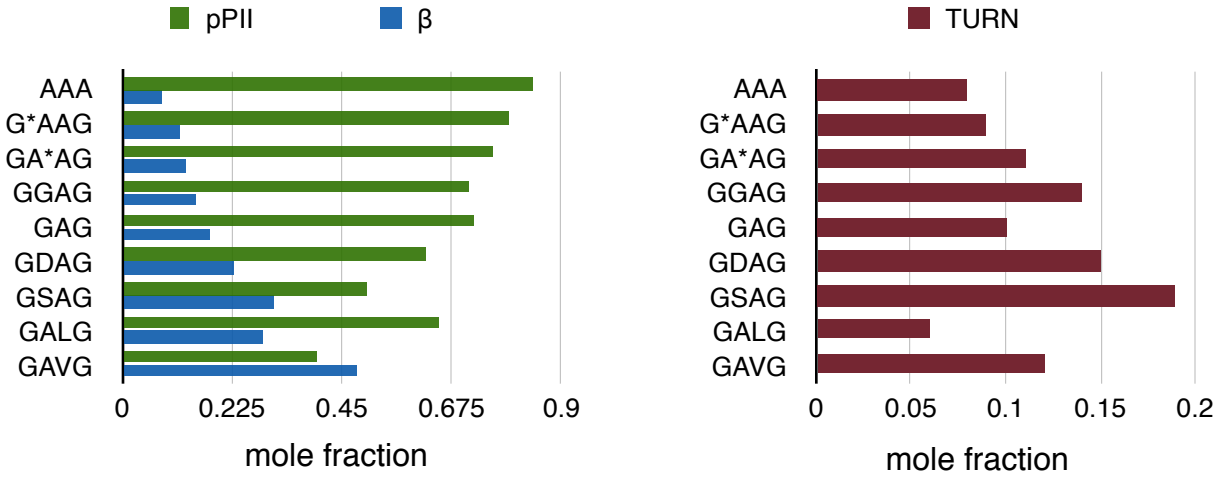

Figure 7.6: pPII (green) and β-strand (blue) and turn (red) content obtained for alanine residues within the denoted peptide systems. G*AAG denotes the first alanine residue of GAAG, whereas GA*AG denotes the second alanine residue in GAAG.

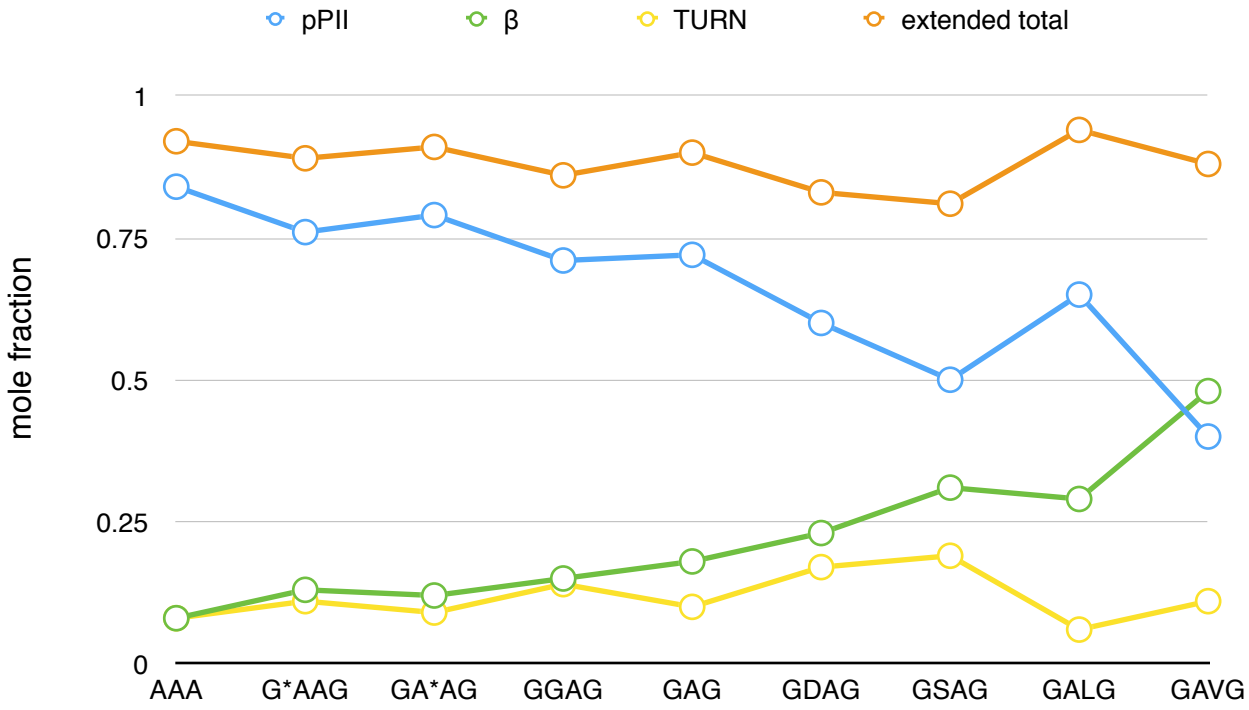

Figure 7.6: pPII (green) and β-strand (blue), turn (yellow), and total extended conformations (pPII plus β) (orange) obtained for alanine residues within he denoted peptide systems. G*AAG denotes the first alanine residue of GAAG, whereas GA*AG denotes the second alanine residue in GAAG.

7.2.2.2 NN Effects on Turn Preferences

Aspartic acid and serine are among a sub-class of amino acid residues that we have found to have above-average preferences for turn-like conformations. Aspartic acid has the largest intrinsic turn content (32%) followed by serine (25%). In addition to large propensity for a variety of turns, these two amino acid residues contain a particular preference for the so called asx-turn region located in the upper right quadrant of the Ramachandran plot, which we have shown in aspartic acid to be stabilized by intra-peptide hydrogen bonding (55). The question arises to what extent these preferences for turn-like conformations are affected by nearest neighbors. We chose to examine the effect of A, K, L, V amino acids on the conformational ensemble of D and S as these amino acids showed the largest effects in restricted coil libraries (153). Figures 7.8 and 7.9 display the change in extended and turn populations for aspartic acid and serene respectively upon insertion of an aliphatic (A, L, V) and a charged side chain (K). The respective propensities of D and S are also compared to the intrinsic preferences of these residues as ascertained from GDG and GSG peptides as well as the homo-peptide DDD. The turn content of the central residue of DDD was found to be quite large, encompassing 38% of the total conformational ensemble, indicating that aspartic acid as a neighbor stabilizes the intrinsic turn population of nearby aspartic acid residues. Noteworthy is the complete lack of any pPII sub-population in DDD. As evident from Figure 7.8, the turn-forming capability of both aspartic acid and serine is significantly reduced upon insertion of any aliphatic residue. Effects on aspartic acid are more

dramatic with respect to loss in turn propensity than those on serine. In general, as the bulkiness of the residue increases, i.e., A<L<V there is a concomitant decrease in turn preference. Again, valine seems to be the most efficient in eliminating this propensity for both aspartic acid and serine, dropping total turn content from 32% to 15% in D, and 25% to 12% in S, respectively. For aspartic acid, most of this drop is due to the loss in ass-turns, which decreases from 23% to 10% of the total conformational ensemble (Table 7.2A). In combination with the obtained influence of valine on alanine discussed above this observation suggests that valine functions as an eraser of conformational propensities. In addition, valine was shown earlier (see *Chapter 6.1 of this thesis*) to be special with regard to its large enthalpic and entropic differences between pPII and β strand, which we hypothesized to be a result of the β-branched side chain limiting rotational degree of freedom in the pPII compared to the β-strand. It is possible that the β-branching of valine disrupts the intra-peptide H-bonding capability of aspartic acid and serine which effectively eliminates turn preference in favor of a more balanced ensemble.

With the reduction in turns, the pPII content of both aspartic acid and serine generally increases. This trend is most dramatic for aspartic acid, for which in a purely glycine environment has a relatively low pPII content of only 20%. When alanine is inserted this rises only modestly to 24%, but upon insertion of large residues, such as lysine, leucine, and valine, the pPII content increases to 40-45%. In contrast, the β-strand population remains 40% or larger for all investigated peptide motifs. As shown in Figure 7.9, NN interactions seem to implement changes within the ensemble, such that increases

in pPII are somewhat balanced by decreases in β-strand and turn populations. However, for some peptides (GDKG, GDLG, GDVG, GSLG, GSVG) the decrease in turns and β-strand populations does not outweigh the increase in pPII, resulting in a larger extended state population as a whole.

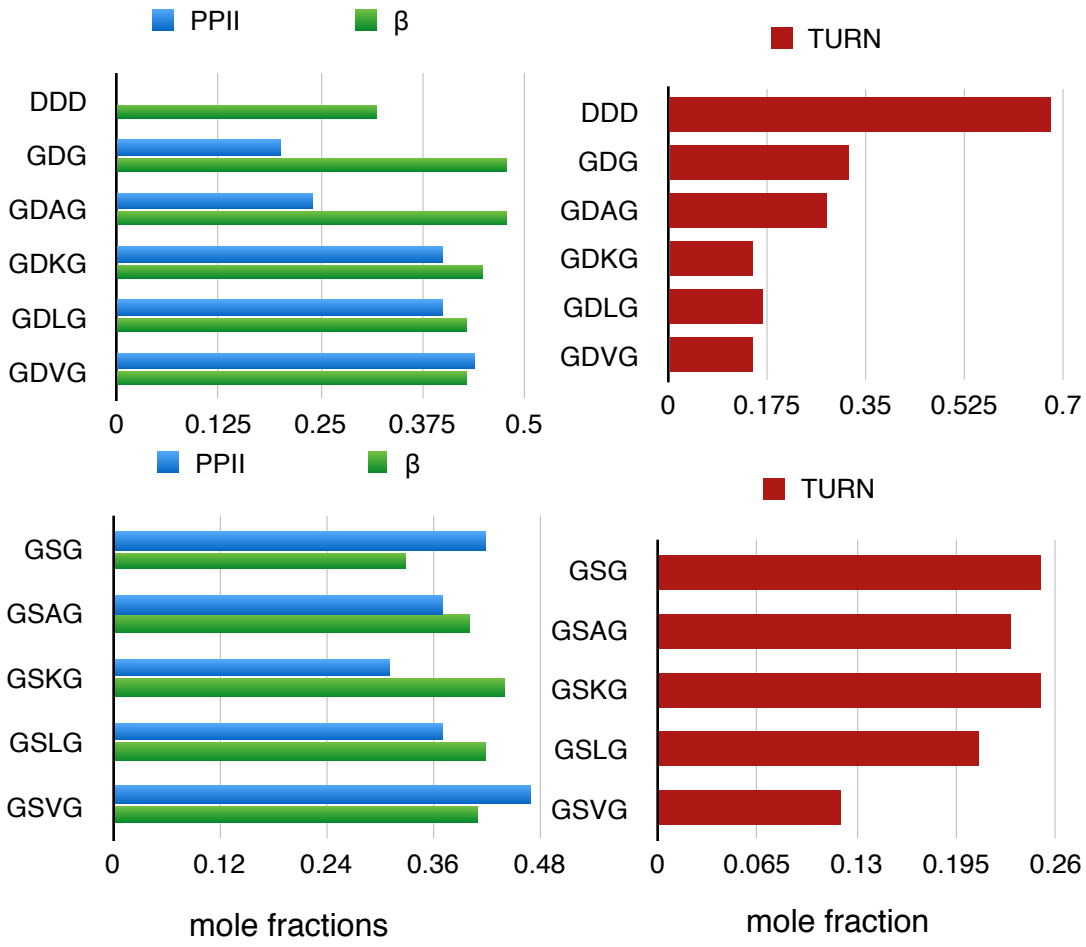

Figure 7.7: pPII (green) and β-strand (blue) and turn (red) content obtained for aspartic acid (upper panels) and serine (lower panels) within he denoted peptide systems.

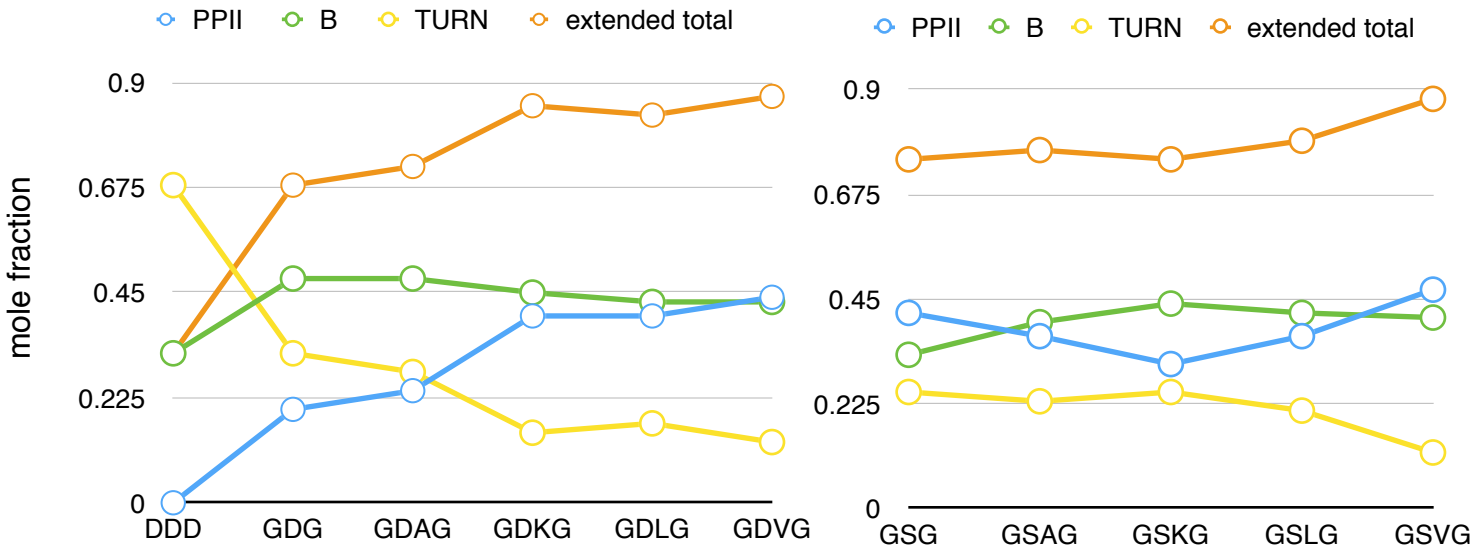

Figure 7.9: pPII (green) and β-strand (blue), turn (yellow), and total extended conformations (orange) obtained for aspartic acid (left panel) and serine (right panel) residues within the denoted peptide systems.

7.2.2.3 NN Effects on Lysine, Leucine, and Valine

Lysine and leucine are two amino acid residues that have intrinsically more balanced conformational ensembles than the other residues addressed thus far. The leucine residue in GLG contains 45% and 42% pPII and β-strand content respectively. Lysine has slightly more pPII bias with 50% and 41% pPII and β-strand content respectively. Compared to the high preferences for particular regions in the Ramachandran plot observed in many amino acid residues, these ensembles are much closer to what Scheraga et al. would define as a statistical coil (84). Figure 7.9A and 7.9B display the change in conformational content with respect to each investigated K and L containing peptide. Compared to what was observed and discussed above for more conformationally biased residues such as alanine, aspartic acid, serine, there is relatively little change in the conformational ensembles of K and L. Both pPII and β-strand conformations fluctuate about their mean values. In contrast, NN effects on valine are slightly more pronounced (Figure 7.10C). This is likely due to the fact that valine has a slight intrinsic preference for the β-strand region of the ramachandran plot (i.e., in a purely glycine environment). Valine has been shown by us to have an extended population dominated by β-strand (46%) (vs. 32% pPII) with a somewhat higher turn preference (22%). As seen in Figure 7.10C aspartic acid and serine increase this disparity in population among extended structures whereas leucine and lysine decrease this difference. The turn forming residues seem to slightly increase the already present β-

strand preference of valine, whereas aliphatic residues destabilize this preference resulting in a more balanced distribution.

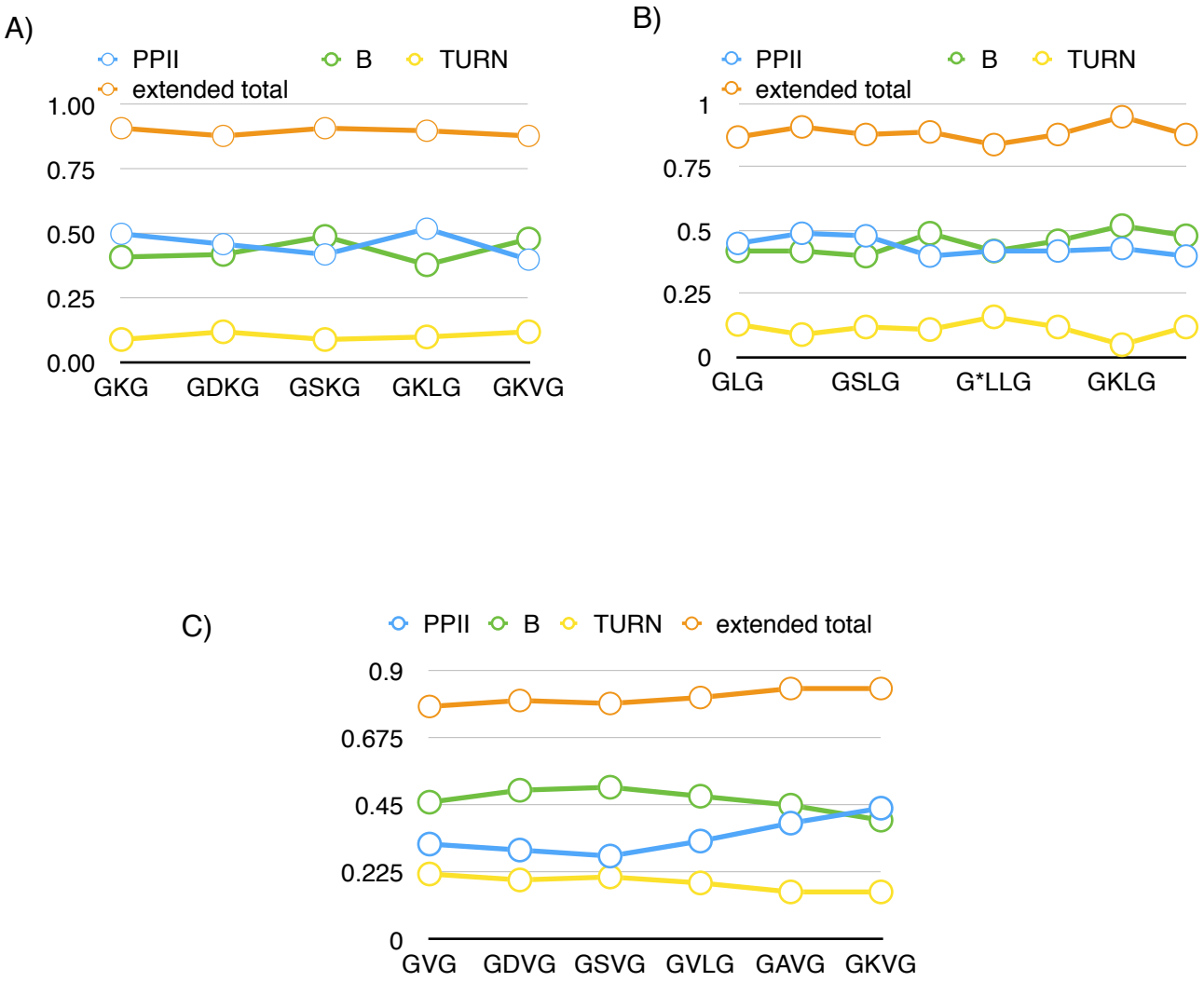

Figure 7.9: pPII (blue) and β-strand (green), turn (yellow), and total extended conformations (orange) obtained for A) lysine, B) leucine, and C) valine.

7.2.2.4 Thermodynamics

It is useful to examine the factors underlying the Gibbs free energy landscape in order to gain a better understanding of why certain conformations are (de)stabilized. To extract thermodynamic information, the $^{3}J(H^{N}H^{\alpha})$ coupling constants for each residue within each peptide were measured as a function of temperature. As previously discussed, these constants reflect the average ϕ value of the correlated residue according to the Karplus relationship. In line with our strategy adopted for the thermodynamic analysis of GxG peptides, we utilized the afore-described pseudo-two state thermodynamic model that considers redistributions among extended structures (pPII-β) as temperature dependent whereas turn populations are assumed temperature independent in accordance with recent results from Rybka et al. (141). This model is described in detail in *Ch 4 of this thesis*. The experimental $^{3}J(H^{N}H^{\alpha})$ (T) data were fitted with equation 4.7, using again solely ΔH as a free paramter. Using the pseudo-two state model, we obtained very good fits to the experimental data $^{3}J(H^{N}H^{\alpha})$ as a function of temperature for all residues (Figure 7.11). For comparison, within each amino acid series the $^{3}J(H^{N}H^{\alpha})$ (T) data for the corresponding GxG peptide model are also presented.

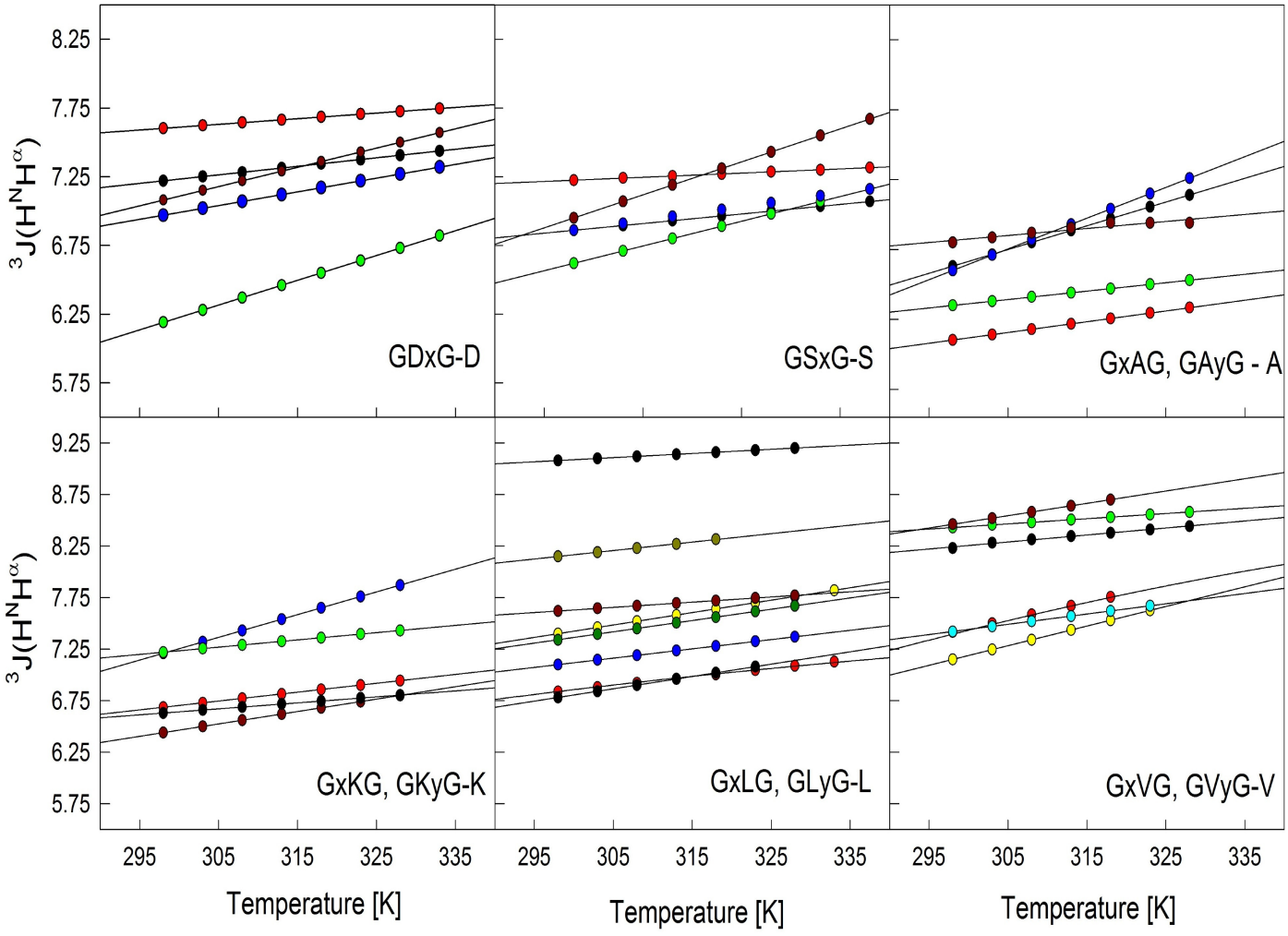

Figure 7.10: $^3J(H^NH^\alpha)$ as a function of temperature for GxyG peptides. The amino acid of interest in each plot is noted by -X. For instance, GDxG-D shows the ^3J coupling constants as a function of temperature for the amino acid residue D in all GDxG peptides.

In general, the $^3J(H^NH^\alpha)$ coupling constants for all residues increase as a function of temperature indicating a conformational redistribution from predominantly pPII to β-strand conformations with increasing temperature. Amino acid residues with high populations of turns have weaker temperature dependencies of $^3J(H^NH^\alpha)$. In contrast, residues with more balanced pPII-β distributions exhibit large changes in ^3J values with temperature. This interplay can be seen when comparing NN effects in, for instance, the GSxG peptide series (Figure 7.11). As discussed above, valine as neighbor practically eliminates the turn forming propensity of aspartic acid. In line with this, the temperature dependence of $^3J(H^NH^\alpha)$ for serine in GSVG becomes quite large in comparison to GSG. The $^3J(H^NH^\alpha)$ coupling constants for D and S at room temperature drop significantly upon insertion of any NN residue. This is due to the large increases in pPII (and decreases in turn content) for these residues as discussed above (Figure 7.9). In contrast, the decrease in pPII content observed for all alanine containing tetrapeptides due to NN interactions effectively increases the $^3J(H^NH^\alpha)$ compared to that of GAG (Figure 7.11). The thermodynamic values ΔH and ΔS obtained from our fits of $^3J(H^NH^\alpha)$ (T), and the corresponding room temperature ΔG values calculated based on the pPII and β mole fractions derived from vibrational analysis are visualized in Figures 7.12-7.18.

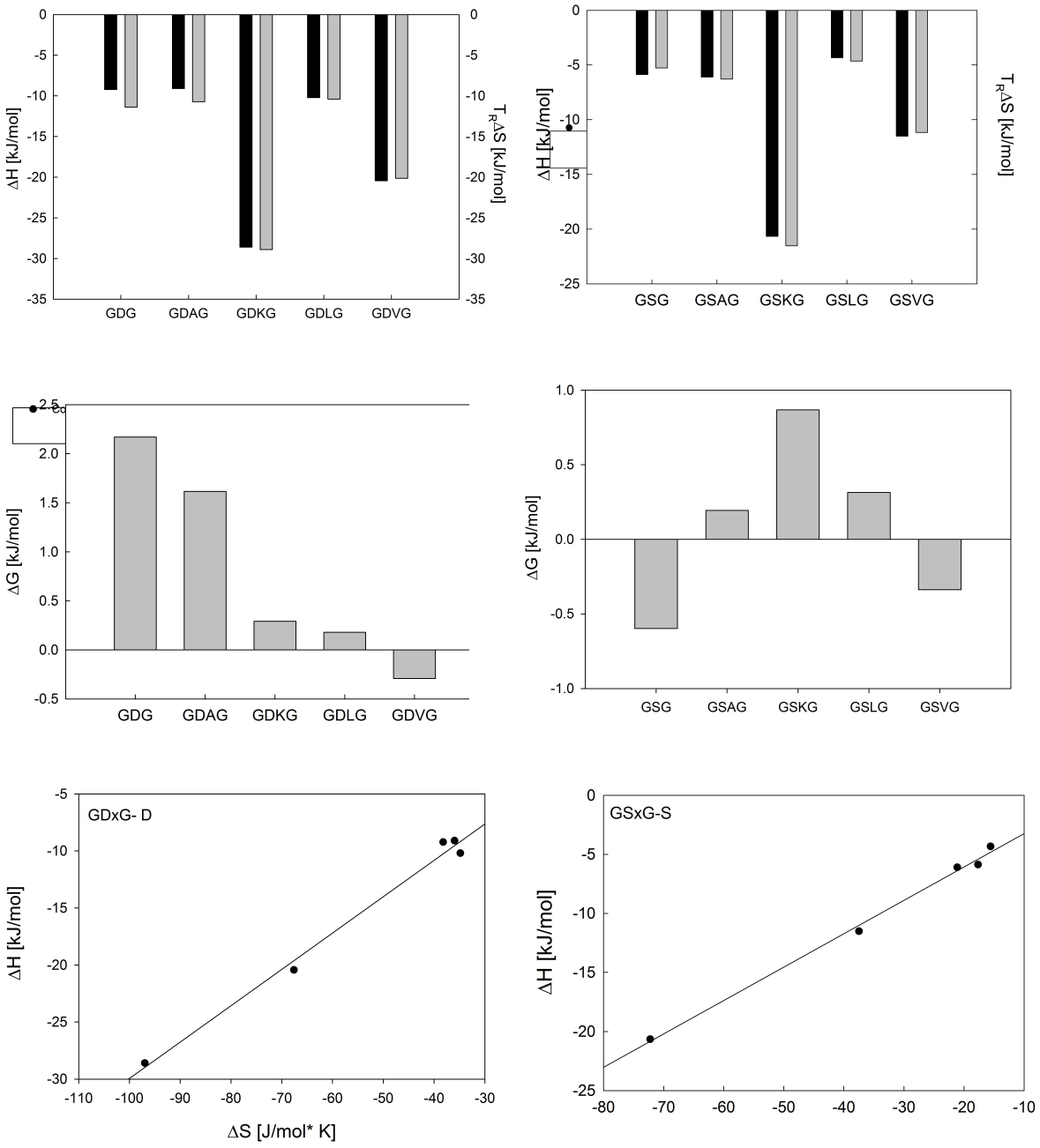

Figure 7.11: (A) ΔH and TΔS values obtained from fitting of ^{3}J coupling constants with pseudo-two state model, (B) ΔG values calculated from respective pPII and β mole fractions obtained from vibrational and 2D NMR analysis and (C) relationship between ΔS and ΔH for aspartic acid and serene residues within the GDxG and GSxG peptide series.

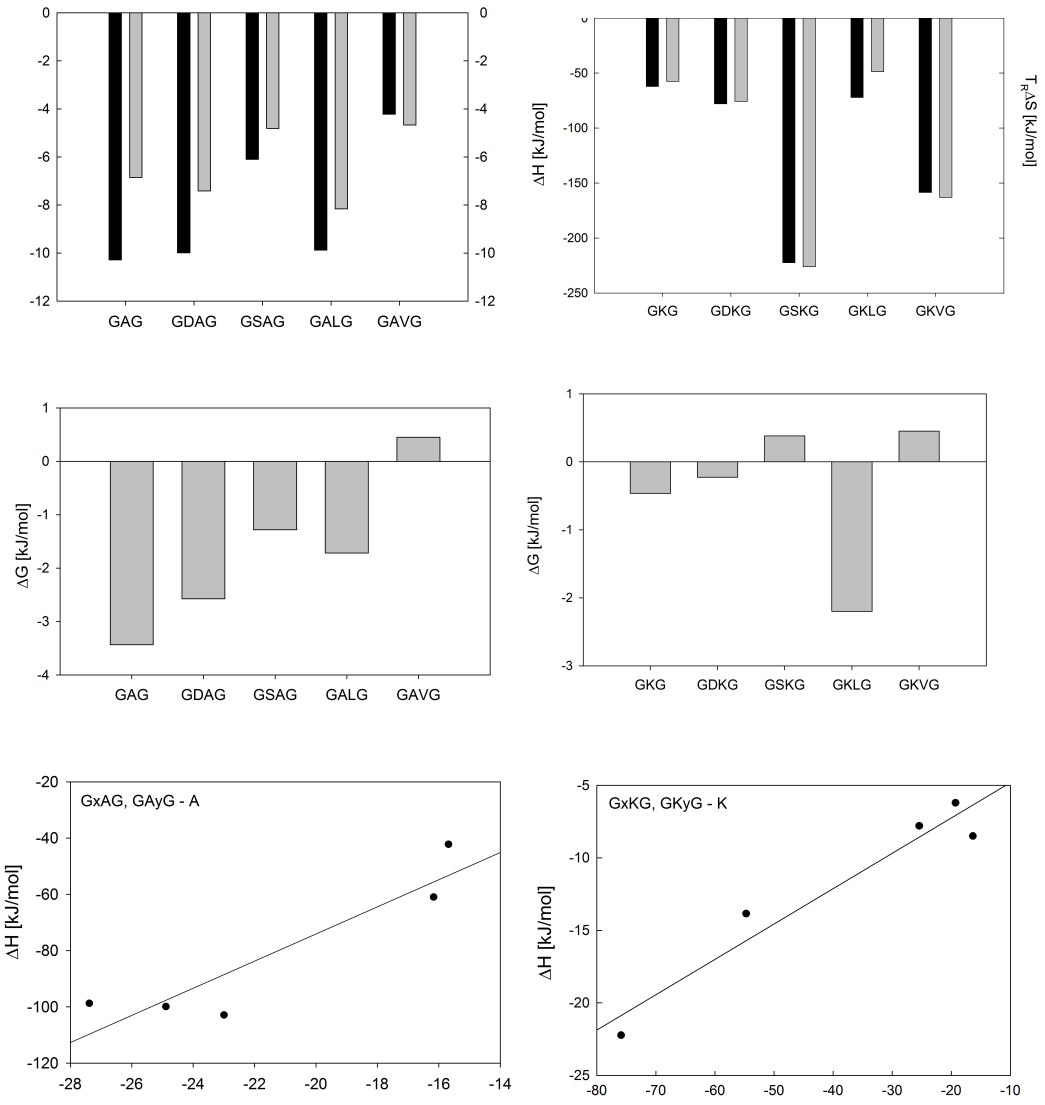

Figure 7.12: (A) ΔH and TΔS values obtained from fitting of ^{3}J coupling constants with pseudo-two state model, (B) ΔG values calculated from respective pPII and β mole fractions obtained from vibrational and 2D NMR analysis and (C)relationship between ΔS and ΔH for lysine residues within the GAyG and GxAG as well s GxKG and GKyG peptide series.

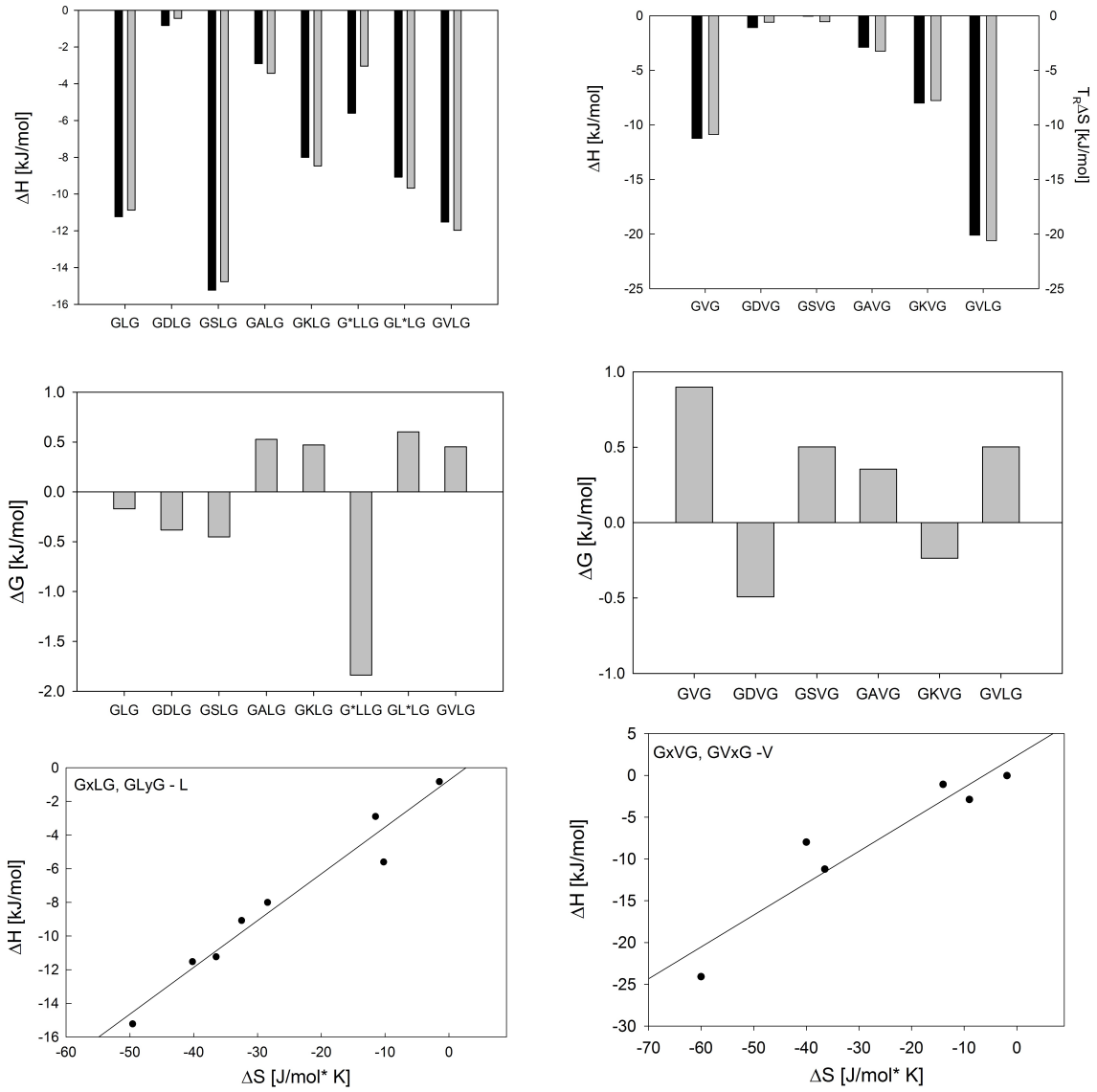

Figure 7.13: (A) ΔH and TΔS values obtained from fitting of ^{3}J coupling contestants with pseudo-two state model, (B) dG values calculated from respective pPII and β mole fractions obtained from vibrational and 2D NMR analysis and (C)relationship between ΔS and ΔH for lysine residues within the GLyG and GxLG as well s GxVg and GVyG peptide series.

Figures 7.11-7.13 display the thermodynamic results (ΔH and ΔS) obtained from fitting of $^{3}J(T)$ data in Figure 7.10 with the pseudo two-state model. The enthalpic and entropic differences between pPII and β are noticeably different within each amino acid series. The same can be said for the ΔG values that were obtained independently from mole fractions of pPII and β derived from the combined vibrational and 2D-NMR analysis. While the variant ΔG values within each series reflect the above described NN effects on the conformational ensembles of the investigated amino acid residues each peptide, the large variations of ΔH and ΔS within each peptide series were not. For instance, as noted above the conformational ensembles obtained for lysine within the GxKG and GyKG series did not drastically change upon different NN insertions, however the respective thermodynamics underlying the pPII/β transitions range by 15kJ/mol for ΔH and 75kJ/mol for TΔS (at room temperature) (Figure 7.12). In particular, serine as a neighbor drastically increases both ΔH and ΔS for lysine as compared to that in a purely glycine environment. In addition, the balance between enthalpy and entropy is reversed, with the entropic contribution to Gibbs free energy outweighing the enthalpic stabilization of pPII. Similar effects are seen with valine as neighbor. The largest range in ΔH and ΔS as compared to intrinsic values are seen for the GDxG series, with enthalpy increasing by approximately 22kJ/mol GDKG relative to GDG, underscoring again the large effect that lysine apparently has on affecting the underlying pPII/B thermodynamics. For illustration, Figure 7.12 displays the obtained thermodynamics behind the pPII/B

transition for aspartic acid residues within the GDyG peptide series. In contrast, alanine, which was shown earlier to be affected dramatically by NN interactions (in terms of both decreases in pPII content and increases in turn content), shows comparatively weaker changes in pPII/B thermodynamics due to NN interactions. The range in changes of ΔH within the GxAG and GAyG series (Figure 7.12) varies by only 6kJ/mol, which is markedly lower than all other investigated residues within their respective NN series. This likely reflects the ability of alanine to preserve its hydration shell regardless of neighboring residue, due to its comparatively small side chain relative to other amino acid residues.

Figures 7.12C-7.14C display the relationships between ΔH and ΔS for all peptide GxyG series. Contrary to what was obtained earlier for GxG peptides, we generally obtain a lower degree of correlation between these two parameters due to NN effects. This is reflected by the larger deviation from linearity. The respective R^2 values along with the theoretical 'compensation' temperatures obtained from fitting each GxyG series with a linear relationship ($\Delta H = \alpha - T_c \Delta S$) are listed below in Table 7.3. As previously discussed, we assign ideal compensation to reflecting a common mechanism behind the effect, namely solvation. The aspartic acid and serine series both have the largest degree of ΔH/ΔS correlation, meaning that the comprising peptides in each series have very similar compensation temperatures. It is interesting that the NN effects for these amino acids exhibit a compensation effect, whereas we previously found that these amino acids in the respective GxG series do not share an isoequilbrium temperature with other amino

acid residues (*Chapter 6.1 of this thesis*). We attributed the lack of compensation for aspartic acid and serine as due to the increased degree of self solvation of these residues (i.e. intra-peptide H-bonding) competing with solvation by water. In contrast, the NN effect for alanine exhibits the lowest degree of compensation. As shown earlier, alanine in a purely glycine environment also did not exist in isoequilbrium with any other amino acid residue.

Table 7.3: R^2 and T_c values obtained from linear fitting of each peptide series GDyG, GSyG, GxAG and GAyG, GxLG and GLyG, GxKG and GKyG, GxVG and GVyG.

Peptide Series	GDyG	GSyG	GxAG, GAyG	GxLG, GLyG	GxKG, GKyG	GxVG, GVyG
R	0.96	0.96	0.79	0.94	0.9	0.86
Tc	318 K	282 K	482 K	278 K	243 K	381 K

7.3 Conclusion

To explore nearest-neighbor effects in the unfolded state we conducted a combined 2D-NMR and vibrational study on selected "GxyG" host guest peptides: GDyG, GSyG, GxLG, GxVG, where x/y={A, K, L, V}. This analysis allowed for the extraction of the conformational ensembles of both x and y residues within each GxyG series. Comparison of these ensembles with the respective amino acid ensembles obtained for GxG peptides shows large changes in sub-populations due to NN interactions. This is in direct contrast to the isolated pair hypothesis, and on a broader

scale, with the random coil model. Interestingly we showed that residues that have large intrinsic biases towards specific sub-populations tend to lose these preferences upon interaction with a given neighbor. For instance, residues that prefer turn-like conformations (namely aspartic acid and serine) lose this turn preference in favor of increased pPII populations, which ultimately increases the total extended state sampling. In a similar light, the large intrinsic preference of pPII for alanine is diminished upon insertion of any non-alanine neighboring residues. Large pPII propensities (above that for GAG) are only obtained with alanine as a neighbor. The loss of pPII for alanine generally increases with the bulkiness of the NN side chain, indicating that sterically more demanding side chains disrupt the optimal hydration of the alanine and hence its pPII propensity. This effect is seen most clearly for GAVG in which the pPII propensity of alanine is eliminated in favor of more balanced extended state sub-populations (pPII/ β). In contrast to alanine, aspartic acid and serine, the other investigated residues (leucine, lysine, and valine) show comparatively smaller effects in their conformational ensembles upon insertion of non-glycine neighbors.

Results of the associated thermodynamic analysis reveals that small changes in apparent conformational ensembles due to NN interactions may mask comparatively larger changes in underlying Gibbs free energy contributions. In addition, the lower degree of correlation between ΔH and ΔS values derived for the GxyG peptides series indicates that the compensation temperatures between pPII and β are substantially modified by NN interactions. This is particularly true for alanine, for which we obtain

very little compensation between ΔH and ΔS at room temperature. This implies that at temperatures sufficiently below the compensation temperature conformational preferences can become very pronounced. This non-ideal compensation effect obtained for GxyG peptides additionally underscores the reliability and significance of the exact ΔH/ΔS compensation obtained previously for GxG peptides.

Taken together, these results suggest that NN interactions act as conformational randomizers, eliminating intrinsic biases in favor of largely balanced pPII/ β dominated ensembles. With regard to unfolded or highly disordered states, these results suggest that local order along the chain will depend on amino acid sequence. However, the loss of conformational preference in favor of more extended conformations when one takes into account NN interactions implies that the unfolded chain becomes largely a pPII/β coil at or near physiological temperatures. This is particularly true for sequences with high degrees of heterogeneity in terms of amino acid type. IDPs in general are characterized by low content of bulky aliphatic amino acid residues and high content of polar amino acid residues. In light of our results attributing branched aliphatic residues as conformational randomizers, it is possible that intrinsically disordered regions in proteins avoid conformational randomization by minimizing the occurrence of these residues, hence promoting local order within the disordered region. In addition, many disordered proteins contain low complexity of sequence, i.e., have sequences with overrepresentation of residues. Given the finding that NN of the same type reinforce intrinsic conformational preferences, it may be that ID regions in proteins are designed to

impart local order through homologous sequences while maintaining disorder dominated by pPII/β coils on a larger scale.

CH. 8 SUMMARY AND OUTLOOK

The discovery of IDPs and bio-functionally important intrinsically disordered regions within proteins has motivated numerous studies aimed at describing residue-level conformational ensembles in the unfolded state. Multiple lines of evidence gathered over the last 15 years strongly suggest that amino acids residues display unique and restricted conformational preferences in the unfolded state. These preferences produce local order in the unfolded state, which could be relevant for understanding the structure-function relationship of disordered proteins. Unfolded state conformational preferences are different from those observed in folded regions of proteins, and their existence in the unfolded state is in direct contrast to the classic random coil model. To fully understand residue level order/disorder one has to formulate a quantitative, experimentally based picture of conformational ensembles and to determine the physical basis underlying these biases. To this end, this thesis has described (mainly) spectroscopic studies on short peptides that were designed to elucidate conformational preferences of amino acid

residues, their underlying thermodynamics, and how these are governed by nearest neighbor interactions and mediated by solvation.

With regard to conformational studies, a variety of short peptide model systems may be chosen, ranging from dipeptides to blocked and unblocked tripeptides. In the work conducted for this thesis, generally unblocked tripeptides or tetrapeptides were chosen as they offer the best spectroscopic resolution due to increased separation of individual amide I band. The choice of model system was validated by our results on trialanine which show that there is no substantial difference in conformational ensembles upon protonation of end groups. This suggests that groups beyond neighboring Cα moieties do not significantly influence intrinsic conformational propensities. Further, we showed that blocked forms of amino acids show remarkable similarity to the unblocked GxG peptides, which indicated that both are suitable model systems for investigating unfolded states, conformational preference.

To understand residue-level order/disorder, the physical bases underlying conformational preference must be investigated. In this context alanine and its high intrinsic pPII preference has garnered much attention. Theories on the mechanism of pPII stabilization (particularly for alanine) generally point to the role of optimal solvation, although unambiguous experimental evidence for this is lacking. To experimentally investigate whether solvation governs conformational preference and pPII biases specifically, the thermodynamics underlying preferences of trialanine in different binary solvents, as well as GxG peptides in aqueous solution, were determined through

temperature dependent NMR and UVCD studies. It was found that residues substantially differ in terms of respective enthalpy and entropic contributions to their pPII/β equilibrium. Regardless of individual differences, however, a majority of amino acid residues within model GxG peptides have nearly identical compensation temperatures and differences in their Gibbs free energy between pPII and β-strand conformations near physiological temperatures. This type of iso-equilbrium suggests a common mechanism behind conformational preference, namely peptide solvation. Residues that avoid iso-equilbrium, namely, A, D and T, are those which were found by us to have have large degrees of conformational preference; Alanine's small side chain allows for optimal backbone solvation, hence disproportionally stabilizing pPII, whereas aspartic acid and threonine have a tendency to self-solvate, which is then reflected in uniquely high turn-like preferences.

To further examine whether solvation is the key to alanine's uniquely high preferences for pPII, a separate thermodynamic investigation was performed on trialanine in various co-solvent systems. If optimal hydration around the residue is indeed pivotal for pPII preference, then perturbing the solvation system with co-solvents should destabilize pPII. It was found that small admixtures of alcohol co-solvents (5% v/v) considerably alters the enthalpic and entropic differences between pPII and β- strand sub-states. The changes of thermodynamic parameters and pPII content are not monotonous with respect to increasing co-solvent fractions, however, the combined manifold of ΔH and ΔS values obtained for the investigated binary mixtures exhibits an enthalpy-entropy

compensation. This indicates that trialanine in various co-solvent systems have a common transitions temperature, and in general, that conformational ensembles are governed by hydration of the peptide.

The information garnered regarding intrinsic conformational propensities and how these are governed by solvation must be complimented by an understanding of how conformational preferences are modified by nearest-neighbor interactions. To this end we conducted an extensive investigation of neighbor effects on selected GxyG tetrapeptide motifs: GDyG, GSyG, GxLG, GxVG, where x/y={A,K,LV}. Using a combined vibrational and 2D-NMR approach we were able to determine the conformational ensemble of each x and y residue within the GxyG tetrapeptides. Comparison of these ensembles with the respective amino acid ensembles obtained for GxG peptides shows large changes in sub-populations due to NN interactions. This is in direct contrast to the isolated pair hypothesis, and on a broader scale, with the random coil model. Interestingly, residues that have large intrinsic biases towards specific sub-populations tend to lose these preferences upon interaction with a given neighbor. In addition, the more bulky the side chain the more efficient it is at destabilizing conformational preferences. This indeed suggests that the mechanism behind neighbor interactions is solvent mediated, as disruption of neighboring residues' hydration shell is more probable with branched and/or bulky side chains. Thermodynamic results on the same GxyG peptides substantiates this finding, as steric and electrostatic effects should cause solely

enthalpic changes to the Gibbs free energy, whereas we obtain large changes in both enthalpic and entropic contributions.

Taken together, our results establish that not only are intrinsic conformational preferences (i.e., no NN interactions) in unfolded and disordered states governed by solvation, but that these preferences are then modified by solvent mediated NN interactions. The situation for highly disordered regions or IDPs then, is complex and largely dependent on amino acid composition and sequence. Randomization of conformational preference through NN interactions suggests that disordered regions exist largely as pPII/β coils. The high degree of population in the extended region of the Ramachandran plot decreases the entropy of the chain and distinguishes it from a random coil. Short segments containing order within a larger disordered chain may be obtained depending on the local sequence of the proteins.

BIBLIOGRAPHY

1. Venter JC (2001) Science 291:1304.

2. Pauling L, Corey, R.B., (1953) Proc. R. Soc. Lond. B. 141(10).

3. Lehninger A, Nelson, D.L, Cox, M. M. (1969) Principles of Biochemistry. 5th Edition ed. 2008.

4. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29(31): 7133-7155.

5. Ramachandran GN, Ramachandran C, & Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7:95-99.

6. Anfinsen CB, Haber, E. (1961) Journal of biological chemistry 236:422.

7. Anfinsen CB (1973) Science 181:223.

8. Levinthal C (1968) Are there pathways for protein folding? J. de Chim. Phys.et de Phys.-Chim. Biol. 65:44-45.

9. Bryngelson JD & Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci U S A 84(21):7524-7528.

10. Dill KC, HS (1997) From Levinthal to pathways to funnels. Nature Structural Biology 4(1):10-19.

11. Onuchic JNL-S, Z.; Wolynes, P.G.; (1997) Theory of protein folding; The energy landscape persepective. Ann. Rev. Phys. Chem. 48:545-600.

12. Wright PE, Dyson HJ, & Lerner RA (1988) Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. Biochemistry 27:7167-7175.

13. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, & Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272(20):5129-5148.

14. Uversky VN (2008) Natively Unfolded Proteins. Unfolded Proteins. From Denaturated to Intrinsically Disordered, ed Creamer TP (Nova, Nauppauge, NY).

15. Uversky V & Dunker AK (2012) Why Are We Interested in the Unfolded Peptides and Proteins? Protein and Peptide Folding, Misfolding, and Non-

Folding, Wiles Series in Protein and Peptide Science, ed Schweitzer-Stenner R (Wiley & Sons, Hoboken), pp 3-54.

16. Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Current Opinion in Structural Biology 8(1):101-106.

17. Dobson CM (1999) Protein misfolding, evolution and disease. Trends Biochem Sci 24(9):329-332.

18. Gsponer J, Haberthür U, & Calfisch A (2003) The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from yeat prion Sup35. Proc. Natl. Acad. Sci. USA 100:5154-5159.

19. Calamai M, et al. (2005) Reversal of protein aggregation provides evidence for multiple aggregated States. J Mol Biol 346(2):603-616.

20. Cho MK, et al. (2009) Structural characterization of alpha-synuclein in an aggregation prone state. Protein Sci 18(9):1840-1846.

21. Brant DA & Flory PJJ (1965) The Configuration of Random Polypeptide Chains. II. theory. J. Am. Chem. Soc. 87:2791-2800.

22. Flory PJ (1969) Statistical Mechanics of Chain Molecules (Wiley & Sons, New York) pp 30-31.

23. Tanford C (1968) Protein Denaturation. Adv. Prot. Chem. 23:121-282.

24. Toal S, Schweitzer-Stenner, R., (2014) Local Order in the Unfolded State: Conformational Biases and Nearest Neighbor Interactions. Biomolecules 4:725-773.

25. Alexandrescu AT, Abeygunawardana C, & Shortle D (1994) Structure and Dynamics of a Denaturared 131-Residue Fragment of Staphylococcal Nuclease: A heteronuclear NMR Study. Biochemistry 33:1063-1072.

26. Shortle D (2002) The Expanded Denaturated State: An Ensemble of Conformations Trapped in a Locally Encoded Topological Space. Adv. Protein Chem. 62:1-23.

27. Ohnishi S, Lee AL, Egdell MH, & Shortle D (2004) Direct Demonstration of Structural Similarity between native and Denatured Eglin C. Biochemistry 43:4064-4070.

28. Bernado P, et al. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci U S A 102(47):17002-17007.

29. Bernado. P, Bertoncini CW, Griesinger C, Zweckstetter M, & Blackledge M (2005) Defining Long-Range Order and Local Disorder in Nativea-Synuclein Using Residua Dipolar Couplings. J. Am. Chem. Soc. 127:17968-17969.

30. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, & Svergun DI (2007) Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. J. Am. Chem. Soc. 129:5656-5664.

31. Mukrasch MD, et al. (2007) Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. J. Am. Chem. Soc. 129:5235-5243.

32. Dunker KA & Obradovic Z (2001) The protein trinity-linking function and disorder. Nature Biotechnol 19:805-806.

33. Uversky VN, Oldfield CJ, & Dunker KA (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signalling. J. Mol. Recognit. 18:343-384.

34. Sandhu KS (2009) Journal of Molecular Recognition 22:1-8.

35. Dill KA & Shortle D (1991) Denatured State of Proteins. Annu. Rev. Biochem. 60:795-825.

36. Fiebig KM, Schwalbe H, Buck M, Smith LJ, & Dobson CM (1996) Toward a description of the Conformations of Denatured States of proteins. Comparison of a Random Coil Model with NMR measurements. J. Phys. Chem. 100:2661-2666.

37. Smith LJ, et al. (1996) Analysis of Main Chain Torsion Angles in proteins: Predicition of NMR coupling Constants for native and Random Coil Conformations. J. Mol. Biol. 255:494-506.

38. Schwalbe H, et al. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. Biochemistry 36:8977-8991.

39. Chandrasekar K, Profy AT, & Dyson HJ (1991) Solution Conformational Preferences of Immunogenic Peptides Derived from the Principal

Neutralizing Determinant of the HIV-1 Envelope Glycoprotein gp 120. Biochemistry 30.

40. Dyson HJ & Wright PE (1993) Peptide Conformation and Protein Folding. Cur. Op. Struct. Biol. 3:60-65.

41. Bai Y, Chung J, Dyson HJ, & Wright PE (2001) Structural and dynamic characterization of an unfolded state of polar apo-plastocyanin formed under denaturing conditions. Protein Sci. 10:1056-1066.

42. Dyson HJ & Wright PE (2004) Unfolded Proteins and Protein Folding Studied by NMR. Chem. Rev. 104:3607-3622.

43. Bruschweiler R, Blackledge M, & Ernst RR (1991) Multi-conformational peptide dynamics derived from NMR data: a new search algorithm and its application to antamanide. J Biomol NMR 1(1):3-11.

44. Meier S, Grzesiek S, & Blackledge M (2007) Mapping the Conformational Landscape of Urea-Denaturated Ubiquitin Using Residual Dipolar Couplings. J. Am. Chem. Soc. 129:9799-9807.

45. Gerum C, Silvers R, Wirmer-Bartoschek J, & Schwalbe H (2009) Unfolded-state structure and dynamic influence the fibril formation in human prion protein. Angew. Chem. Int. Ed. Engl. 48:9452-9456.

46. Jensen MR, Salmon L, Nodet G, & Blackledge M (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. J Am Chem Soc 132(4):1270-1272.

47. Silvers R, et al. (2012) Modulation of Structure and Dynamics by Disulfide Bond Formation in Unfolded States. J. Am. Chem. Soc 134:6846-6854.

48. Sziegat F, et al. (2012) Disentangling the Coil: Modulation of conformationa and Dynamic Properties by Site-directed Mutation in the Non-Native State of Hen Egg White Lysozyme. Biochemistry 51:3361-3372.

49. Shi Z, Olson CA, Rose GD, Baldwin RL, & Kallenbach NR (2002) Polyproline II structure in a sequence of seven alanine residues. Proc. Natl. Acad. Sci. USA 99:9190-9195.

50. Ding L, Chen K, Santini PA, Shi Z, & Kallenbach NR (2003) The Pentapeptide GGAGG Has PII Conformation. J. Am. Chem. Soc. 125:8092-8093.

51. Shi Z, et al. (2005) Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β-sheet scales. Proc. Natl. Acad. Sci. USA 102:17964 - 17968.

52. Hagarman A, et al. (2006) Conformational analysis of XA and AX dipeptides in water by electronic circular dichroism and 1H NMR spectroscopy. J Phys Chem B 110(13):6979-6986.

53. Shi ZS, Chen K, Liu ZG, Sosnick TR, & Kallenbach NR (2006) PII structure in the model peptides for unfolded proteins: Studies on ubiquitin fragments and several alanine-rich peptides containing QQQ, SSS, FFF, and VVV. Proteins-Structure Function and Bioinformatics 63:312-321.

54. Hagarman A, Measey TJ, Mathieu D, Schwalbe H, & Schweitzer-Stenner R (2010) Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I' band profiles and NMR scalar coupling constants. J Am Chem Soc 132(2):540-551.

55. Hagarman A, et al. (2011) Amino acids with hydrogen-bonding side chains have an intrinsic tendency to sample various turn conformations in aqueous solution. Chemistry 17(24):6789-6797.

56. Toal S, Meral D, Verbaro D, Urbanc B, & Schweitzer-Stenner R (2013) pH-Independence of trialanine and the effects of termini blocking in short peptides: a combined vibrational, NMR, UVCD, and molecular dynamics study. J Phys Chem B 117(14):3689-3706.

57. Shi Z, Shen K, Liu Z, & Kallenbach NR (2006) Conformation in the Backbone in Unfolded Proteins. Chem. Rev. 106:1877-1897.

58. Tiffany ML & Krimm S (1968) New chain conformations of poly(glutamicacid) and polylysine. Biopolymers 6:1379-1382.

59. Tiffany MLK, S. (1972) Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. Biopolymers 11(11):2309 - 2316.

60. Cowan PM & Mc Gavin S (1955) Structure of Poly-L-Proline. Nature 176:501-503.

61. Woody RW (1992) Circular Dichroism of unordered polypeptides. Adv. Biophys. Chem. 2:37-79.

62. Gokce I, Woody RW, Anderluh G, & Lakey JH (2005) Single peptide bonds exhibit poly(pro)II ("random coil") circular dichroism spectra. J. Am. Chem. Soc. 127(27):9700-9701.

63. Woody R (2005) The Exciton model and the circular dichroism of polypeptides. MONATSHEFTE FUR CHEMIE 136(3):347-366.

64. Woody RW (2009) Circular Dichroism Spectrum of Peptides in the Poly(Pro)II Conformation. J. Am. Chem. Soc. 131:8234-8245.

65. Mattice WL (1974) The effect of temperature and salt concentration on the circular dichroism exhibited by unionized derivatives of L-alanine in aqueous solution. Biopolymers 13(1):169-183.

66. Dukor RK & Keiderling TA (1991) Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. Biopolymers 31(14):1747-1761.

67. Schweitzer-Stenner R (2013) Different Degrees of Disorder in Long Disordered Peptides Can Be Discriminated by Vibrational Spectroscopy. J. Phys. Chem. B 117:6927-6936.

68. Woutersen S & Hamm P (2000) Structure determination of trialanine in water using polarized sensitive two-dimensional vibrational spectroscopy. J. Phys. Chem. B 104:11316-11320.

69. Woutersen S & Hamm P (2001) Isotope-edited two-dimensional vibrational spectroscopy of trialanine in aqueous solution. J. Chem. Phys. 114:2727-2737.

70. Eker F, Cao X, Nafie L, & Schweitzer-Stenner R (2002) Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR, and VCD spectroscopy study. J Am Chem Soc 124(48):14330-14341.

71. Woutersen S, et al. (2002) Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. J. Chem. Phys. 117:6833-6840.

72. Eker F, et al. (2003) The Structure of Alanine Based Tripeptides in Water and Dimethyl Sulfoxide Probed by Vibrational Spectroscopy. J. Phys. Chem. B 107:358-365.

73. Eker F, Griebenow K, Cao X, Nafie LA, & Schweitzer-Stenner R (2004) Preferred peptide backbone conformations in the unfolded state revealed

by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. Proc Natl Acad Sci U S A 101(27):10054-10059.

74. Graf JN, P.H.; Stock, G.; Schwalbe, H. (2007) Structure and dynamics of a homologous series of alanine peptides: A joint molecular dynamics/nmr study. Jounral of the American Chemical Society 128:1179-1189.

75. Toal S, Omidi A, & Schweitzer-Stenner R (2011) Conformational Changes of Trialanine Induced by Direct Interactions between Alanine Residues and Alcohols in Binary Mixtures of Water with Glycerol and Ethanol. J. Am. Chem. Soc. 133:12728.

76. Han W-G, Jakanen KJ, Elstner M, & Suhai S (1998) Theoretical study of aqueous N-Acetyl-L-alanine N_-methylamide: structures and Raman, VCD, and ROA spectra. J. Phys. Chem. B 102:2587-2602.

77. Poon CD, Samulsi ET, Weise CF, & Weisshaar JC (2000) Do Bridging Water Molecules Dictate the Structure of a Model Dipeptide in Aqueous Solution? J. Am. Chem. Soc. 122:5612-5613.

78. Weise CF & Weisshaar JC (2003) Conformational Analysis of Alanine Dipeptide from Dipolar Couplings in a Water-Based Liquid Crystal. J. Phys. Chem. B 107:3265-3277.

79. Kim YS, Wang J, & Hochstrasser RM (2005) Two-dimensional infrared spectroscopy of the alanine dipeptide in aqueous solution. J Phys Chem B 109(15):7511-7521.

80. McColl IH, Blanch EW, Hecht L, Kallenbach NR, & Barron LD (2004) Vibrational Raman optical activity characterization of poly(l-proline) II helix in alanine oligopeptides. J Am Chem Soc 126(16):5076-5077.

81. Mikhonin AV, Ahmed Z, Ianoul A, & Asher SA (2004) Assignments and Conformational Dependencies of the Amide III Peptide Backbone UV Resonance Raman Bands. J. Phys. Chem. B 108:19020-19028.

82. Mikhonin AV & Asher SA (2005) Uncoupled Peptide Bond Vibrations in aa-Helical and Polyproline II Conformation of Polyalanine Peptides. J. Phys. Chem. B 109:3047-3052.

83. Schweitzer-Stenner R, Eker F, Griebenow K, Cao X, & Nafie L (2004) The conformation of tetraalanine in water determined by polarized Raman, FTIR and VCD spectroscopy. J. Am. Chem. Soc. 126:2768-2776.

84. Makowska J, et al. (2006) Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. Proc. Natl. Acad. Sci. USA 103:1744-1749.

85. Makowska J, et al. (2007) Further Evidence for the Absence of Polyrpoline II Stretch in the XAO Peptide. Biophys. J. 92:2904-2917.

86. Tanaka S & Scheraga HA (1976) Statistical Mechanical Treatment of Protein Conformation. I. Conformational Properties of Amino Acids in Proteins. Macromolecules 9:142-157.

87. Tanaka S & Scheraga HA (1976) Statistical mechanical treatment of protein conformation. II. A three-state model for specific-sequence copolymers of amino acids. Macromolecules 9:150-167.

88. Zagrovic B, et al. (2005) Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci U S A 102(33):11698-11703.

89. Schweitzer-Stenner R & Measey T (2007) The Alanine-Rich XAO Peptide Adopts a Heterogeneous Population, Including Turn-Like and PPII Conformations. Proc. Natl. Acad. Sci. USA 104:6649–6654.

90. Schweitzer-Stenner R (2009) Distribution of Conformations Sampled by the Central Amino Acid Residue in Tripeptides Inferred From Amide I Band Profiles and NMR Scalar Coupling Constants. J. Phys. Chem. B 113:2922-2932.

91. Kaminski GA, Friesner RA, Tirado-Rives, Julian, & Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B 105(28):6474-6487.

92. Hu H, Elstner M, & Hermans J (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. Proteins 50(3):451-463.

93. Drozdov AN, Grossfield A, & Pappu RV (2004) Role of solvent in determining conformational preferences of alanine dipeptide in water. J Am Chem Soc 126(8):2574-2581.

94. Gnanakaran S & Garcia AE (2005) Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. Proteins 59(4):773-782.

95. Feig M (2008) Is Alanine Dipeptide a Good Model for Representing the torsional Preferences of Protein Backbones. J. Chem. Theo. Comput. 4:1555-1564.

96. Kwac K, Lee KK, Han JB, Oh KI, & Cho M (2008) Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: comparisons with IR and vibrational circular dichroism spectra. J Chem Phys 128(10):105106.

97. Cruz V, Ramos j, & Martinez-Salazar J (2011) Water-Mediated Conformations of the Alanine Dipeptide as Revealed by Distributed Umbrella Sampling Simulations, Quantum Mechanics Based Calculations, and Experimental Data. J. Phys. Chem. B 115:4880-4886.

98. Garcia-Pietro FF, Galván IF, Aguliar MA, & Martin ME (2011) Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology. J. Chem. Phys, 135:194502.

99. Ioannou F, Archontis G, & Leontidis E (2011) Specific interactions of sodium salts with alanine dipeptide and tetrapeptide in water: insights from molecular dynamics. J Phys Chem B 115(45):13389-13400.

100. Ishizuka R, Huber GA, & McCammon JA (2011) Solvation Effect on the Conformation of Alanine Dipeptide: Integral Equation Approach. J. Phys. Chem. Letts. 1:2279-2283.

101. Best RB, Buchete NV, & Hummer G (2008) Are current molecular dynamics force fields too helical? Biophys J 95(1):L07-09.

102. Best RB & Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. J Phys Chem B 113(26): 9004-9015.

103. Beck DAC, Alonso DOV, Inoyama D, & Dagget V (2008) The intrinsic conformational propensities of the 20 naturally occuring amino acids and reflection of these propensities in proteins. Proc. Natl. Acad. Sci. USA 105:12259-12264.

104. Mu Y, Kosov DS, & Stock G (2003) Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. J. Phys. Chem. B 107:5064-5073.

105. Gnanakaran S & Garcia AE (2003) Validation of an All-Atom Protein Force Field: From Dipeptides to larger Peptides. J. Phys. Chem. B 107:12555-12557.

106. Kentsis A, Mezei M, Gindin T, & R. O (2004) Unfolded State of Polyalanine Is a Segmented Polyproline II Helix. Proteins: Structure, Function and Genetics 55:493-501.

107. Zaman MH, Shen M-Y, Berry RS, Freed KF, & Sosnick TR (2003) Investigations into Sequence and Conformational Dependence of Backbone Entropy, Inter-basin Dynamics and the Flory Isolated-pair Hypothesis for Peptides. J. Mol. Biol. 331:693-711.

108. Verbaro D, Ghosh I, Nau WM, & Schweitzer-Stenner R (2010) Discrepancies between Conformational Distributions of a Polyalanine Peptide in Solution Obtained from Molecular Dynamics Force Fields and Amide I' Band Profiles. J. Phys. Chem. B 114:17201-17208.

109. Duan Y, et al. (2003) A Point-Charge Force Field for Molecular Mechanics Simulations of Protein Based on Condensed-Phase Quantum Mechanical Calculations. J. Comp. Chem. 24:1999-2012.

110. Lanza GC, Maria A. (2013) Comprehensive and Accurate Ab Initio Energy Surface of Simple Alanine Peptides. Chemphyschem 14(14):3284 - 3293.

111. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79(2):926.

112. Berendsen HJC, Grigera JR, & Straatsma TP (1987) The missing term in effective pair potentials. J. Phys. Chem.. 91(24):6269-6271.

113. Mahoney MWJ, W. L.; (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J. Chem. Phys. 112:8910-8922.

114. Horn HW, et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120(20):9665-9678.

115. Nerenberg P & Head-Gordon T (2011) Optimizing Protein-Solvent Force Fields To Reproduce Conformational Preferences of Model Peptides. J. Chem. Theo. Comp.:in press.

116. Wickstrom L, Okur A, & Simmerling C (2009) Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling data. Biophys. J. 97:853-856.

117. Avbelj F & Baldwin RL (2003) Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. Proc Natl Acad Sci U S A 100(10):5742-5747.

118. Rucker AL, Pager CT, Campbell MN, Qualls JE, & Creamer TP (2003) Host-guest scale of left-handed polyproline II helix formation. Proteins: Structure, Function and Genetics 53:68-75.

119. Mezei M, Fleming PJ, Srinivasan R, & Rose GD (2004) Polyproline II Helix Is the Preferred Conformation for Unfolded Polyalanine in Water. Proteins: Structure, Function and Genetics 55:502-507.

120. Fleming PJ, Fitzkee NC, Mezei M, Srinivasan R, & Rose GD (2005) A novel method reveals that solvent water favors polyproline II over β-strand conformation on peptides and unfolded proteins: conditional hydrophobic accessible surfact areas (CHASA). Prot. Sci. 14:111-118.

121. Avbelj F (2012) Solvation and electrostatics as Determined of Local Structural Order in Unfolded Peptides and Proteins. Protein and Peptide Folding, Misfolding and Non-Folding, Wiley Series in Protein and Peptide Science, ed Schweitzer-Stenner R (John Wiley & Sons,, Hobocken), pp 131-158.

122. Kelly MA, et al. (2001) Host-guest study of left-handed polyproline II helix formation. Biochemistry 40(48):14376-14383.

123. Pappu RV & Rose GD (2002) A simple model for polyproline II structure in unfolded states of alanine-based peptides. Protein Sci. 11:2437-2455.

124. Garcia AE & Sanbonmatsu KY (2002) Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci U S A 99(5):2782-2787.

125. Law P & Dagett V (2010) The relationship between water bridges and the polyproline II conformation: a large scale analysis of molecular dynamics simulations and crystal structures. Prot. Eng. Design. Selection 23:27-33.

126. Liu Z, et al. (2004) Solvent dependence of PII conformation in model alanine peptides. J Am Chem Soc 126(46):15141-15150.

127. Grdadolnik J, Mohacek-Grosev V, Baldwin RL, & Avbelj F (2011) Populations of the three makor backbone conformations in 19 amino acid dipeptides. Proc. Natl. Acad. Sci. USA 108:1794-1798.

128. Mertuka G, Dyson HJ, & Wright PE (1995) 'Random coil'^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. J. Biomol. NMR 5:14-24.

129. Plaxco KW, et al. (1997) The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. J. Biomol. NMR 10:221-230.

130. Schwarzinger S, et al. (2001) Sequence-Dependent Correction of random Coil NMR Chemical Shifts. J. Am. Chem. Soc. 123:2970-2978.

131. Van der Velde F, Pereira L, & Rollema HS (2004) The revised NMR chemical shift data of carrageenans. Carbohydr. Res. 29:813.

132. Avbelj F, Grdadolnik SG, Grdadolnik J, & Baldwin RL (2006) Intrinsic backbone preferences are fully present in blocked amino acids. Proc Natl Acad Sci U S A 103(5):1272-1277.

133. Grdadolnik J, Grdadolnik SG, & Avbelj F (2008) Determination of Conformational Preferences of Dipeptides Using Vibrational Spectroscopy. J. Phys. Chem. B 112:2712-2718.

134. Rose GD (1978) Nature 272:586.

135. Dyson HJ & Wright PE (2002) Insights into the Structure and Dynamics of Unfolded Proteins from Nuclear Magnetic Resonance. Adv. Protein Chem. 62:311-340.

136. Richardson JS (1981) Advances in Protein Chemistry 34:167.

137. Rees DCL, M.; Lipscomb, W. N.; (1983) Journal of Molecular Biology.

138. Wan WYM-W, E. J.; (1999) Journal of Molecular Biology 286.

139. Duddy WJ, Nissink JW, Allen FH, & Milner-White EJ (2004) Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins. Protein Sci 13(11):3051-3055.

140. Song B, Kibler P, Malde A, Kodukula K, & Galande AK (2010) Design of Short Linear peptides That Show Hydrogen Bonding Constraints in Water. J. Am. Chem. Soc. 132(4508-4509):4508.

141. Rybka K, et al. (2013) Disorder and order in unfolded and disordered peptides and proteins: a view derived from tripeptide conformational analysis. II. Tripeptides with short side chains populating asx and beta-type like turn conformations. Proteins 81(6):968-983.

142. West N & Smith LJ (1998) Side-chains in Native and Random Coil Protein Conformations. Analysis of NMR Coupling Constants and χ_1Torsion Angle Preferences. J. Mol. Biol. 180:867-877.

143. Hennig M, et al. (1999) Side-chain conformations in an unfolded protein: chi1 distributions in denatured hen lysozyme determined by heteronuclear 13C, 15N NMR spectroscopy. J Mol Biol 288(4):705-723.

144. Hahnke MJR, C.; Heinicke, F.; Schwalbe, H. (2010) THeNH(COCA)- HAHB NMR experiment for the stereospecific assignment of hb- protons in non-native states of proteins. Journal of American Chemistry Society 2010(132).

145. Sziegat FS, R.; Hahnke, M.J.; Jensen, M.R.; Blackledge, M.; Wirmer-Bartoschek, J.; Schwalbe, H. (2012) Disentangling the coil: modula- tion of conformationa and dynamic properties by site-directed mutation in the non-native state of hen egg white lysozyme. Bio- chemistry. Biochemistry 51.

146. Pachler KG (1963) Nuclear magnetic resonance study of some α-amino acids-I Coupling constants in alkaline and acidic medium. Spectrochim. Acta 19:2085-2092.

147. Kaur H, Sasidhar, Y. U. (2012) For the SEQUENCE YKGQ, the turn andextended conformational forms are separated by small barriers and the turn propensity persists even at high temperatures: implications for protein folding. Journal of physical chemistry B 116.

148. Serrano L (1995) Comparison between the ϕ-distribution of the amino acids in the Protein Data Base and NMR data indicates that amino acids have various ϕ propensities in the random coil conformation. J. Mol. Biol. 254:322-333.

149. Jha AK, et al. (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. Biochemistry 44(28): 9691-9702.

150. Swindells MB, MacArthur MW, & Thornton JM (1995) Intrinsic ϕ,ψ propensities of amino acids, derived from the coil regions of known structures. Nat. Struct. Biol. 2:596-603.

151. Penkett CJ, et al. (1997) NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. J Mol Biol 274(2): 152-159.

152. Avbelj F & Baldwin RL (2004) Origin of the neighboring residue effect on peptide backbone conformation. Proc Natl Acad Sci U S A 101(30): 10967-10972.

153. Sosnick TR (Sampling library. pp http://godzilla.uchicago.edu/cgi-bin/rama.cgi.

154. Jha AK, Colubri A, Freed KF, & Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci U S A 102(37):13099-13104.

155. Pappu RV, Srinivasan R, & Rose GD (2000) The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. Proc. Natl. Acad. Sci. USA 97:12565-12570.

156. Kelly SM, Jess, T.J., Price, N.C. (2005) How to Study Proteins by Circular Dichroism. Biochim. Biophys. Acta 1751.

157. Bulheller BM, Rodger, A., Hirst, J. D. (2007) Circular and linar dichroism of proteins. Physical Chemistry Chemical Phsyics 9.

158. Moffitt W (1956) Journal of chemical physics 25.

159. Jensen MR, Zwckstetter, M., Huang, J., Blackledge, M. (2014) Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. Chemical Reviews 114.

160. Karplus M (1959) Theoretical calculation links NMR coupling constant to molecular geometry. J. Chem. Phys. 30:11.

161. Wang AC & Bax A (1996) Determination of the Backbone Dihedral Angles phi in Human Ubiquitin from Reparametrized Empirical Karplus Equations. J. Am. Chem. Soc. 118:2483.

162. Schmidt JM (2007) Asymmetric Karplus curves for the protein side-chain ^{3}J couplings. J. Biomol. NMR 37:287-301.

163. Vuister GW & Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNCHa) coupling. J. Am. Chem. Soc. 115:7772-7777.

164. Hu J, Bax, A. (1997) Journal of American Chemistry Society 119.

165. Weisemann R, et al. (1994) Determination of H(N),H-alpha and H(N), c' coupling-constants in C-13,N-15-labeled proteins. Jounral of Biomolecular NMR 4(2):231-240.

166. Wirmer J & Schwalbe H (2002) Angular dependence of (1)J(N-i,C-alpha i) and (2)J(N-i,C alpha(i-1)) coupling constants measured in J-modulated HSQCs. Journal of Biomolecular NMR 23:47-55.

167. Krimm S & Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. Adv Protein Chem 38:181-364.

168. Mirkin NG & Krimm S (1991) Ab initiovibrational analysis of hydrogen-bonded trans- and cis-N-methylacetamide. J. Am. Chem. Soc. 113:9742-9747.

169. Chen XG, Schweitzer-Stenner R, Asher SA, Mirkin NG, & Krimm S (1995) Vibrational Assignments of trans-N-Methylacetamide and Some of Its Deuterated Isotopomers from Band Decomposition of IR, Visible, and Resonance Raman Spectra. J. Phys. Chem. 99:3074-3083.

170. Schweitzer-Stenner R (2001) Visible and UV-resonance Raman spectroscopy on model peptides. J. Raman Spectrosc. 32:711.

171. Wang Y, Purello R, Georgiou S, & Spiro TG (1991) UVRR spectroscopy of the peptide bond. 2. Carbonyl H-bond effects on the ground- and excited-state structures of N-methylacetamide. J. Am. Chem. Soc. 113.

172. Schweitzer-Stenner R, Sieler G, Mirkin N, & Krimm S (1998) Intermolecular Coupling in Liquid and Crystalline States oftrans-N-Methylacetamide

Investigated by Polarized Raman and FT-IR Spectroscopies. J. Phys. Chem. A. 102:118-127.

173. Torii H, Tatsumi T, Kanazawa T, & Tasumi M (1998) Effects of Intramolecular Hydrogen-Bonding Interactions on the Amide I Mode of N-Methylacetamide: matrix Isolation Infrared Studies and ab Initio Molecular Orbital calculations. J. Phys. Chem. B 102:309-314.

174. Measey T, Hagarman A, Eker F, Griebenow K, & Schweitzer-Stenner R (2005) Side chain dependence of intensity and wavenumber position of amide I' in IR and visible Raman spectra of XA and AX dipeptides. J Phys Chem B 109(16):8195-8205.

175. Choi J-H, Ham S, & Cho M (2002) Inter-peptide interaction and delocalization of amdie I vibrational excitons in myoglobin and flavodoxin. J. Chem. Phys. 117:6821-6832.

176. Torii H & Tasumi M (1998) Ab Initio Molecular Orbital Study of the Amide I Vibrational Interactions Between the Peptide Groups in Di- and Tripeptides and Considerations on the Conformation of the Extended Helix. J. Raman Spectrosc. 29:81-86.

177. Schweitzer-Stenner R (2004) Secondary structure analysis of polypeptides based on an excitonic coupling model to describe the band profile of amide I of IR, Raman and vibrational circular dichroism spectra. J. Phys. Chem. B. 108:16965-16975.

178. Gorbunov RD, Kosov DS, & Stock G (2005) Ab initio-based exciton model of amide I vibrations in peptides: Definition, conformational dependence and transferrability. J. Chem. Phys. 122:224904-224915.

179. Holzwarth G & Chabay I (1972) Optical Activity of Vibrational Transitions: A Coupled Oscillator Model. J. Chem. Phys. 57:1632-1638.

180. Schweitzer-Stenner R & Measey TJ (2010) Simulation of IR, Raman and VCD amide I band profiles of self-asembled peptides. Spectroscopy 24:25-36.

181. Shaka A, Lee, C., Pines, A. (1988) Iterative Schemese for Bilinear Operators - Application to Sprin Decoupling. Journal of Magnetic Resonance 77.

182. Lohr F & Ruterjans H (1995) (H)NCAHA and (H)CANNH experiments for the determination of the vicinal coupling constants related to the phi-torsion angle. J Biomol NMR 5(1):25-36.

183. Lohr F & Ruterjans H (1999) Alternative E.COSY techniques for the measurement of 3J(C (i) (') (-1),C (i) (beta)) and (3) J(H (i) (N) ,C (i) (beta)) coupling constants in proteins. J Biomol NMR 13(3):263-274.

184. Hennig M, Bermel W, Schwalbe H, & Griesinger C (2000) Determination of psi torsion angle restraints from (3)J(C-alpha,C-alpha) and 3J(C-alpha,H-N) coupling constants in proteins. Journal of American Chemical Society 122(26):6268-6277.

185. Jentzen W, et al. (1995) Conformational Properties of Nickel(II) Octaethylporphyrin in Solution. 1. Resonance Excitation Profiles and

Temperature Dependence of Structure-Sensitive Raman Lines. J. Phys. Chem. 100:14184-14191.

186. Oh K-I, Lee K-K, Park EK, Kwang GS, & Cho M (2010) Circular Dichroism Eigenspectra of Polyproline II and β-Strand Conformers of Trialanine in Water: Singular Value Decomposition Analysis. Chirality 22:E186-E201.

187. Toal SE, Verbaro DJ, & Schweitzer-Stenner R (2014) Role of Enthalpy –Entropy Compensation Interactions in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. J. Phys. Chem B in press.

188. Shi Z, Chen, K., Liu, Z., Kallenbach, N. R. (2006) Conformatin of the Backbone in Unfolded Proteins. Chemical Reiviews 106:1877-1897.

189. Oh K-I, et al. (2012) A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. Proteins Struct. Funct. Genet 80:977-990.

190. Jansen T & Knoester J (2006) Nonadiabatic effects in the two-dimensional infrared spectra of peptides: application to alanine dipeptide. J Phys Chem B 110(45):22910-22916.

191. Schweitzer-Stenner R, Eker F, Huang Q, & Griebenow K (2001) Dihedral angles of trialanine in D_2O Determined by Combining FTIR and Polarized Visible Raman Spectroscopy. J. Am. Chem. Soc. 123:9628-9633.

192. He L, Navarro AE, Shi Z, & Kallenbach NR (2011) End Effects Influence Short Model Peptide Conformation. J. Am. Chem. Soc 134:1571-1576.

193. Gaigeot M-P (2010) Infrared spectroscopy of the alanine dipeptide analog in liquid water with DFT-MD. Direct evidence for PII/β conformations. Phys. Chem. Phys. Chem, 12:10198-10209.

194. Garcia AE (2004) Characterization of non-alpha conformations in Ala peptides. Polymer 120:885-890.

195. Sieler G, Schweitzer-Stenner R, Holtz JSW, Pajcini V, & Asher SA (1999) Different Conformers and Protonation States of Dipeptides Probed by Polarized Raman,UV-Resonance Raman, and FTIR Spectroscopy. J. Phys. Chem. B 103:372-384.

196. Oh K-I, Jung Y-S, Hwang G-S, & Cho M (2012) Conformational distributions of denaturated and unstructured proteins are similar to those of 20x20 blocked dipeptides. J. Biomol. NMR 53:25-41.

197. Schweitzer-Stenner R, Gonzales W, Bourne JT, Feng JA, & Marshall GA (2007) Conformational Manifold of a-Aminoisobutyric Acid (Aib) Containing Alanine-Based Tripeptides in Aqueous Solution Explored by Vibrational Spectroscopy, Electronic Circular Dichroism Spectroscopy, and Molecular Dynamics Simulations. J. Am. Chem. Soc. 129:13095-13109.

198. Schweitzer-Stenner R, Hagarman A, Toal S, Mathieu D, & Schwalbe H (2013) Disorder and order in unfolded and disordered peptides and proteins: a view derived from tripeptide conformational analysis. I. Tripeptides with long and predominantly hydrophobic side chains. Proteins 81(6):955-967.

199. Pajcini V, et al. (1996) Glycylglycinep->p*and charge transfer transition moment orientations: near resonance Raman single crystal measurements. J. Am. Chem. Soc. 118:9716-9726.

200. Schweitzer-Stenner R, et al. (2007) Conformations of Alanine-Based Peptides in Water Probed by FTIR, Raman, Vibrational Circular Dichroism, Electronic Circular Dichroism, and NMR Spectroscopy. Biochemistry 46:1587-1596.

201. Kaur H & Sasidhar YU (2012) For the Sequence YKGQ, the Turn and Extended Conformational Forms Are Separated by Small Barriers and the Turn Propensity Persists Even at High Temperatures: Implications for Protein Folding. J. Phys. Chem. B 116:3850-3860.

202. Duitch L, Toal S, Measey TJ, & Schweitzer-Stenner R (2012) Triaspartate: a model system for conformationally flexible DDD motifs in proteins. J Phys Chem B 116(17):5160-5171.

203. Eker F, Griebenow K, & Schweitzer-Stenner R (2003) Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy. J Am Chem Soc 125(27):8178-8185.

204. Lumry R & Shyamala R (1970) Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small M olecules: A Ubiquitous Property of Water. Biopolymers 9:1125-1227.

205. Dunitz JD (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. Chem. Biol. 2:709-712.

206. Grunwald E & Steel C (1995) Solvent Reorganization and Thermodynamic Enthalpy-Entropy Compensation. J. Am. Chem. Soc 117:5687-5692.

207. Liu L & Guo QX (2001) Isokinetic relationship, isoequilibrium relationship, and enthalpy-entropy compensation. Chem Rev 101(3):673-695.

208. Movileanu L & Schiff EA (2013) Entropy-enthalpy compensation of biomolecular systems in aqueous phase: a dry perspective. Montash Chem 144:59-56.

209. Beasley JR, Doyle DF, Cohen DS, Fine BR, & Pielak GJ (2002) Searching for Quantitative Entropy-Enthalpy Compensation among Protein Variants. Proteins 49:398-402.

210. Krug RR, Hunter WG, & Grieger RA (1976) Enthalpy-entropy compensation. 1. Some fundamental statistical problems associated with the analysis of van't Hoff and Arrhenius data. J. Phys. Chem. 80:2335-2341.

211. Qian H & Hopfield JJ (1996) Entropy-enthalpy compensation: Perturbation and relazation in thermodynamic systems. J. Chem. Phys. 105:9292-9298.

212. Ben Naim A (2009) Molecular Theory of Water and Aquesous Solutions-Part I: Understanding Water (World Scientific, Singapur).

213. Bai Y, Milne JS, Mayne L, & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. Proteins 17(1):75-86.

214. Dunbrack RL, Karplus, M. (1993) Backbone Dependent Rotamer Library for Proteins. Journal of Molecular Biology 230.

215. Hopfield JJ (1974) Electron Transfer between Biological Molecules by Thermally Activated Tunneling. Proc. Natl. Acad. Sci. USa 9:3640-3644.

216. Toal SE, Verbaro DJ, & Schweitzer-Stenner R (2014) Role of Enthalpy –Entropy Compensation Interactions in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. J. Phys. Chem B 118:1309-1318.

217. Baxa MC, Haddadain EJ, Jha AK, Fred KF, & Sosnick TR (2012) Context and Firce Field Dependence of the Loss of Protein Backbone Entropy upon Folding Using Realistic Denatured and Native State Ensembles. J. Am. Chem. Soc 134:15929-15936.

218. Chakrabartty A, Kortemme T, & Baldwin RL (1994) Helix propensities of amino acids measured in alanine based peptides without helix stabilizing side chain interactions. Prot. Sci. 3:843-852.

219. Schweitzer-Stenner R, Measey TJ, Hagarman A, & Dragomir I (2010) The Structure of Unfolded Peptides and Proteins Explored by Vibrational Spectroscopy. Instrumental Analysis of Disordered Proteins: assessing Structure and Conformation., eds Uversky V & Longhi S (Wiley & Sons, Inc, New York), p in press.

220. Creamer TP & Rose GD (1992) Side chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. Proc. Natl. Acad. Sci. USA 89:5937-5941.

221. Creamer TP & Rose GD (1994) α-Helix-Forming Propensities in Peptides and Proteins. Proteins 19:85-97.

222. Blaber M, et al. (1994) Determination of α-Helix Propensity within the Context of a Folded Protein. Site 44 and 131 in Bacteriophage T4 Lysozyme. J. Mol. Biol. 235:600-624.

223. Gekko K & Timasheff SN (1981) Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. Biochemistry 20(16):4667-4676.

224. Timasheff SN (1998) Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of complicated. Adv. Protein. Chem. 51:355-432.

225. Gekko K & Timasheff SN (1981) Thermodynamic and kinetic examination of protein stabilization by glycerol. Biochemistry 20(16):4677-4686.

226. Johnson ME, Malardier-Jugroot C, & Head-Gordon T (2010) Effects of co-solvents on peptide hydration water structure and dynamics. Phys Chem Chem Phys 12(2):393-405.

227. Malardier-Jugroot C, Bwron DT, Soper AK, Johnson ME, & Head-Gordon T (2011) Structure and water dynamics of aqueous peptide solutions in the presence of co-solvents. Phys. Chem. Chem. Phys. 12:382-392.

228. Vagenende V, Yap MGS, & Trout BL (2009) Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. Biochemsitry 48:11084-11096.

229. Mishra R, Seckler R, & Bhat R (2005) Efficient refolding of aggregation-prone citrate synthase by pylyol osmolytes: How well are protein folding and stability aspects coupled? J. Biol. Chem. 280:15553-15560.

230. Sedwick H, Cameron JE, Poon WCK, & Egelhaaf SU (2007) Protein phase behavior and crystallization: Effect of glycerol. J. Chem. Phys. 127:125102.

231. Sahu RK & Prakash V (2008) Mechanism of prevention of aggregation of proteins: A case study of aggregation of α-globulin in glycerol. Int. J. Food. Prop. 11:613-623.

232. Graf J, Nguyen PH, Stock G, & Schwalbe H (2007) Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/ NMR study. J Am Chem Soc 129(5):1179-1189.

233. Schweitzer-Stenner R, et al. (2002) Structure Analysis of Dipeptides in Water by Exploring and Utilizing the Structural Sensitivity of Amide III by Polarized Visible Raman, FTIR-Spectroscopy and DFT Based Normal Coordinate Analysis. J. Phys. Chem. B 106:4294-4304.

234. Chen XG, Schweitzer-Stenner R, Mirkin NG, Krimm S, & Asher SA (1994) N-Methylacetamide and Its Hydrogen-Bonded Water Molecules Are Vibrationally Coupled. J. Am. Chem. Soc. 116:11141-11142.

235. Chen K, Liu Z, Zhou C, Shi Z, & Kallenbach NR (2005) Neighbor effect on PPII conformation in alanine peptides. J Am Chem Soc 127(29): 10146-10147.

236. DeBartolo J, Jha A, Freed K, F., & Sosnick TR (2012) Local Backbone Preferences and Nearest Neighbor Effects in the Unfolded and Native States. Proteins and Peptides. Folding, Misfolding and Unfolding., ed Schweitzer-Stenner R (Wiley & Sons, Chichester), pp 79-98.

237. Pizzanelli S, et al. (2010) Conformation of phenylalanine in the trpeptides AFA and GFG probed by Combining MD Simulations with NMR, FTIR, Polarized Raman and VCD Spectroscopy. J. Phys. Chem. B 114:3965-3978.

238. Verbaro D, Mathieu D, Toal SE, Schwalbe H, & Schweitzer-Stenner R (2012) Ionized Trilysine: A Model System for Understanding the Nonrandom Structure of Poly-L-lysine and Lysine-Containing Motifs in Proteins. J. Phys. Chem. B. 116.

239. Pajcini V, et al. (1996) Glycylglycine p->p* and charge transfer transition moment orientations: near resonance Raman single crystal measurements. J. Am. Chem. Soc. 118:9716-9726.

VITA

Siobhan Eileen Toal was born in Brooklyn, NY, USA on March 27, 1986 to Thomas J. Toal and Fiona McNichol. She attended Carnegie Mellon University and received a Bachelor of Science in Chemistry. Upon completion of her undergraduate degree, she moved to Drexel University to study under the guidance of Dr. Reinhard Schweitzer-Stenner in the Chemistry department.

She has been a contributing author on ten peer reviewed journal articles, with more in preparation, and one book article throughout her research career. Of those peer reviewed journal articles, she was a first author on four. She has helped train and mentor undergraduates throughout her time at Drexel with projects that focused on solvation as well as peptide aggregation. As part of her research she was able to travel to Frankfurt, Germany to learn multidimensional NMR techniques at Johann Wolfgang Goethe University, under the guidance of, and in collaboration with, Dr. Harold Schwalbe. She has presented her work at many local, national, and international conferences. During her time in the graduate program in the Chemistry department at Drexel she has also taught many physical chemistry labs as well as co-lectured Atomic and Molecular Spectroscopy.

PUBLICATIONS CONTRIBUTING TO THIS THESIS

1. S. Toal, D. Mathieu, K. Rybka, C. Richter, H. Schwalbe and R. Schweitzer-Stenner. Nearest-Neighbor Interactions Quantified in Tetrapeptides: A Joint Vibrational and NMR Study. In preparation

2. R. Schweitzer-Stenner and S. Toal. Entropy Reduction in Unfolded Peptides due to Conformational Preferences of Amino Acid Residues. Submitted

3. S. Toal and R. Schweitzer-Stenner. Local Order in the Unfolded State: Conformational Biases and Nearest Neighbor Interactions. Biomolecules, 4, 725-773, 2014

4. S. Toal, D. Verbaro, and R. Schweitzer-Stenner. Role of Enthalpy-Entropy Compensation Interactions in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. J. Phys. Chem. B., 118, 1309-1318, 2014

5. R. Schweitzer-Stenner, A. Hagarman, S. Toal, D. Mathieu, H. Schwalbe. Disorder and Order in Unfolded Proteins: A View Derived from Tripeptide Conformational Analysis. I. Tripeptides with long and predominantly hydrophobic side chains. Proteins, 81, 955-67, 2013

6. K. Rybka, S. Toal, D. Verbaro, D. Mathieu, H. Schwalbe, R. Schweitzer-Stenner. Disorder and Order in Unfolded and Disordered Peptides and Proteins: A View Derived from Tripeptide Conformational Analysis. II. Tripeptides with short side chains populating asx and β-type like turn conformations Proteins, 81, 968-83, 2013

7. S. Toal, D. Verbaro, D. Myral, and B. Urbanc, and R. Schweitzer-Stenner. The pH-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides. A Combined Vibrational, NMR, UVCD and Molecular Dynamics Study. J. Phys. Chem. B., 117, 3689-3706, 2013.

8. L. Duitch, S. Toal, T. Measey, R. Schweitzer-Stenner. Triaspartate: A Model System for Conformationally Flexible DDD Motifs J. Phys. Chem. B. 116, 5160−5171, 2012

9. J. Soffer, S. Toal, D. Verbaro, R. Schweitzer-Stenner. Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools, Methods in Molecular Biology. Vol. 895, Eds.V. Usversky and K. Dunker, Humana Press, 271- 313, 2012.

10. D. Verbaro, D. Mathieu, S. Toal, H. Schwalbe, R. Schweitzer-Stenner. Ionized Trilysine: A Model System for Understanding the Non-Random Structure of Poly-L-

Lysine and Lysine Containing Motifs in Proteins. J. Phys. Chem. B. 116, 8084-8094, 2012.

11. S. Toal, A. Omidi, R. Schweitzer-Stenner. Conformational Changes of Trialanine Induced by Direct Interactions between Alanine Residues and Alcohols in Binary Mixtures. J. Am. Chem. Soc. 133, 12728- 39, 2011.

12. A. Hagarman, D. Mathieu, S. Toal, T. Measey, H. Schwalbe, R. Schweitzer-Stenner. Amino Acids with H-bonding Side Chains have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution. Chem. Eur. J. 17, 6789-97, 2011.