Experimental Characterization of Polyalanine Helices in Short and Long Contexts

by

Robert J. Kennedy III

B.A. Chemistry, College of the Holy Cross, 1996

Submitted to the Department of Chemistry in partial fulfillment of the requirements for the degree of
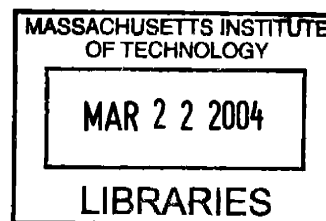
Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2004

Signature of Author _____

Department of Chemistry
November 24, 2003

Certified by _____

Daniel S. Kemp
Professor of Chemistry
Thesis Supervisor

Accepted by _____

Robert W. Field
Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor JoAnne Stubbe_____

Chair

Professor Daniel S. Kemp_____

Thesis Supervisor

Professor Timothy M. Swager_____

Experimental Characterization of Polyalanine Helices in Short and Long Contexts

by
Robert J. Kennedy III

Submitted to the Department of Chemistry at the
Massachusetts Institute of Technology on November 24, 2003 in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemistry

ABSTRACT

Three principal analytical methods: circular dichroism (CD), hydrogen exchange, and the t/c ratio of Ac-Hel, a helix stabilizing N-cap, are used in combination to quantify fractional helicity in order to determine the intrinsic helical propensity of an alanine residue, free from the hydrophobic packing within proteins. Constructed cross checks between these methods add rigor to propensity assignments and present the opportunity to resolve controversies regarding helical propensities and limiting ellipticities used in CD.

Unlike proteins that approximate a single folded conformation, simple helical polyalanine peptides exists as a manifold of partially helical conformers. The Ac-Hel template characterizes short helical conformations that exist within larger manifolds. Using an isolation and solubilization methodology, t/c measurements of Ac-Hel alanine conjugates up to 14 residues indicate an increase in the helical propensity with length that was previously undetected for 6 residue conjugates. A constructed cross check of t/c with circular dichroism ellipticities indicate that both methods provide consistent measurements of peptide helicity.

Protection factors, measured by hydrogen exchange, assign site helicities that provide incisive information about the manifold of helical conformers for a fifteen residue helical alanine region. Protection factor measurements are also used to corroborate the length dependence seen by t/c over an extended length series between 5 and 25 alanine residues.

A joint analysis of the experimental CD ellipticity and the fractional helicity as determined by hydrogen exchange is tested as a cross check to calibrate CD. Long helical peptides with moderately stabilizing N-caps were previously shown to exceed a fractional helicity of 1.0. The conformational averaging seen for intrinsic polyalanine helices can be dramatically reduced by highly stabilizing N- and C- terminal caps. An octaalanine core, stabilized by Ac-$^\beta$aspartyl-Hel as an N-cap and β-amino alanine as a C-cap, also exceeds current literature standards for factional helicity. As a proof of principle, hydrogen exchange measures factional helicity approaching 0.95 for all eight alanine residues. A complete series of this type is proposed to establish the perresidue molar ellipticity for a infinite length helix ($[\theta]_{\infty,222}$) and its length dependence.

Thesis Supervisor: Professor Daniel S. Kemp
Title: Professor of Chemistry

AUTHOR'S BIOGRAPHICAL NOTE

EDUCATION

2003   Ph.D.   Organic Chemistry
Massachusetts Institute of Technology, Cambridge, MA
Thesis Title: Experimental Characterization of Polyalanine
        Helices in Short and Long Contexts
Advisor: Professor Daniel S. Kemp

1996   B.A.   Honors Chemistry, Cum Laude
College of the Holy Cross, Worcester, MA
Thesis Title: Synthesis and Investigation of a γ-Turn Mimic
Advisor: Professor Timothy P. Curran

1992   High Honors
Conard High School, West Hartford, CT

1988   Sedgwick Middle School, West Hartford, CT

1986   Braeburn Elementary School, West Hartford, CT

RESEARCH EXPERIENCE

1996-2003   Graduate Student, Research Assistant
Massachusetts Institute of Technology
Advisor: Professor Daniel S. Kemp

1994-1996   Summer Research Assistant (1994, 1995)
     Howard Hughes Fellow: 1994
Honors Research (1994-1996, 4 semesters)
College of the Holy Cross
Advisor: Professor Timothy P. Curran

AWARDS

MIT Wyeth-Ayerst Scholar, 2001
American Institute of Chemists Award, 1996
Excellence in Analytical Chemistry, 1995

PROFESSIONAL AFFILIATIONS

American Chemical Society
American Peptide Society

PUBLICATION LIST

Primary:

Consistent Helicities from CD and Template *t/c* Data for N-Templated Polyalanines: Progress toward Resolution of the Alanine Helicity Problem: Kennedy, R. J.; Kwok, K. Y.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 934-944.

Short, Solubilized Polyalanines are Conformational Chameleons: Exceptionally Helical if N- and C-Capped with Helix Stabilizers, Weakly to Moderately Helical if Capped with Rigid Spacers: Miller, J. S.; Kennedy R. J.; Kemp, D. S. *Biochemistry* **2001**, *40*, 305-309.

An Amino Acid that Controls Polypeptide Helicity: β-Amino Alanine, the First Strongly Stabilizing C-Terminal Helix Stop Signal: Deechongkit, S.; Kennedy, R. J.; Tsang, K. Y.; Renold, P.; Kemp, D. S. *Tetrahedron Lett.* **2000**, *41*, 9679-9683.

Large Circular Dichroism Ellipticities for N-Templated Helical Polypeptides are Inconsistent with Currently Accepted Helicity Algorithms: Wallimann, P.; Kennedy, R. J.; Kemp, D. S. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1290-1292.

N-α-Benzoyl-*cis*-4-Amino-L-Proline: A γ-Turn Mimetic: Curran, T. P.; Chandler, N. M.; Kennedy, R. J.; Keaney, M. T. *Tetrahedron Lett.* **1996**, *37*, 1933-1936.

Collaborative:

Dual Wavelength Parametric Test of Two-State Models for Circular Dichroism Spectra of Helical Polypeptides: Anomalous Dichoric Properties of Alanine-Rich Peptides: Wallimann, P; Kennedy, R. J.; Miller, J. S.; Shalongo, W.; Kemp, D. S. *J. Am. Chem. Soc.* **2003**, *125*, 1203-1220.

Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities: Miller, J. S.; Kennedy, R. J.; Kemp, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 945-962.

Optimal N-Caps for N-Terminal Helical Templates: Effects of Changes in H-Bonding Efficiency and Charge: Maison, W.; Arce, E.; Renold, P.; Kennedy, R. J.; Kemp, D. S. *J. Am. Chem. Soc.* **2001**, *123*, 10245-10254.

C-Terminal Helix Capping Propensities in a Polyalanine Context for Amino Acids Bearing Nonpolar Aliphatic Side Chains: Maison, W.; Kennedy, R. J.; Miller, J. S.; Kemp, D. S. *Tetrahedron Lett.* **2001**, *42*, 4975-4977.

Chaotropic Anions Strongly Stabilize Short, N-Capped Uncharged Peptide Helices; A New Look at the Perchlorate Effect: Maison, W.; Kennedy, R. J.; Kemp. D. S. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 3819-3821.

Acknowledgements

Education and science share a common passion: the discovery of knowledge and understanding. I would like to especially thank Prof. Daniel Kemp for the opportunity to work and learn in his laboratory. Thank you for sharing the discovery experience, thought provoking discussions and synergetic results. My tenure at MIT was made possible with support by the MIT Chemistry Department (1996-7), the National Science Foundation and the National Institutes of Health (1997-2003).

The work presented here represents the culmination of many ideas and experimental results from projects that have been shared with other members of the Kemp group. Team efforts have supported science and improvements in laboratory instrumentation. The following coworkers were involved: Janet MacLaughlin, Bjoern Heitmann, Sharon Walker, Gabe Job, Bob Moreau, Wolfgang Maison, Eva Arce, Justin Miller, Evan Powers, William Shalongo, Peter Wallimann, Sam Tsang, Peter Rudolf, Lawrence Williams, Kristian Kather, Songpon Deechongkit, Linda Shimizu and Janette Lee.

NMR experiments were preformed at the Department of Chemistry Instrumentation Facility; Jeff Simpson, Mark Wall, and David Bray provided excellent support. Collaborations with Prof. Harald Schwalbe along with his students Elke Duchardt and Julia Wirmer presented the opportunity to explore multidimensional NMR experiments in water some of which are reported in Chapter 1.

Beyond the basement laboratories, I would like to thank my classmates Shana Sturla and Rod Andrade, along with Mike Hewitt, Kevin Shea, Jason Diffendal and Dave Amos for their camaraderie.

Over the past 7 years MIT and the Department of Chemistry planned and executed the renovation of all labs and services in Building 18. The people responsible for choosing to renovate an occupied building include: Marc Jones, Eleonora Serotek, Joe Collins, architects Goody Clancy & Associates and contractors Linbeck, Kennedy and Rossi. These parties deserve to be cited for creating a unique working environment.

To those people who have been lights in dark places, the night life in Boston has never been the same: John Ambrosino, Ron Bilodeau, Rob Botelho, Alan Caplan, Craig Gagnon, Rob Mannke, Corey Reeves, Joe Vieira and Kirstin Zilinskas. Thanks for the laughs and good times.

Discovery has been a reoccurring theme throughout my education. I would like to thank my professors and teachers for their inspiration: All of my teachers at Conard High School, Sedgwick Middle School and Braeburn elementary in West Hartford, CT; also my professors at Holy Cross, especially Prof. Tim Curran, Prof. Margaret Freije and the Department of Chemistry.

I would like to thank three very important educators in my life: Mom, Dad and my brother Ryan. None of this would have been possible with out their guidance and the support of family. Graduate school has been an education for all.

Dedicated to Nancy and Robert Kennedy Jr.
for their love and support

PREFACE

Relying primarily on studies of polyalanines, the simplest of helical peptides, the work presented in this thesis takes key steps towards achieving a long-range Kemp group aim: a chemical understanding of the factors that govern helicity, embodied in an algorithm that accurately predicts the stability of peptide helices from their amino acid compositions and sequences. Resolution of this aim will solve the controversies that have developed in the helicity field utilizing results of carefully planned new experiments and reinterpreting previously published data.

This thesis falsifies two key hypotheses. The first hypothesis regards the temperature independence and stability of short alanine helices that was the cornerstone of nearly all work carried out in the Kemp Group for the past 12 years. Chapter 2 presents new experiments indicating that these data must be reinterpreted. Chapter 3 calls into question a more general hypothesis that underlies the computational modeling of all peptide helices carried out during the past 40 years. These falsification experiments may require a modification of the assumptions used in a new generation of algorithms.

As an illustration, a well known falsification experiment appears in Figure 1. Introduced by the philosopher Karl Popper as an essential component of the scientific method, falsification experiments[1] rarely play a major role in modern mature scientific fields, except in cases where fundamental disagreements concerning methodology or data interpretation exist. The disagreements in the helicity field concern the selection of appropriate helical model peptides and appropriateness of the experimental methods used to quantify their helicities. The studies in this thesis make a case for using new N- and C-

---

[1] Popper, K, R. In *Popper Selections,* Miller, D., Ed.; Princeton University Press: Princeton, NJ 1985.

capped solubilized polyalanines as a host for helicity studies and demonstrates that within this context three key quantitative tools yield mutually consistent results.
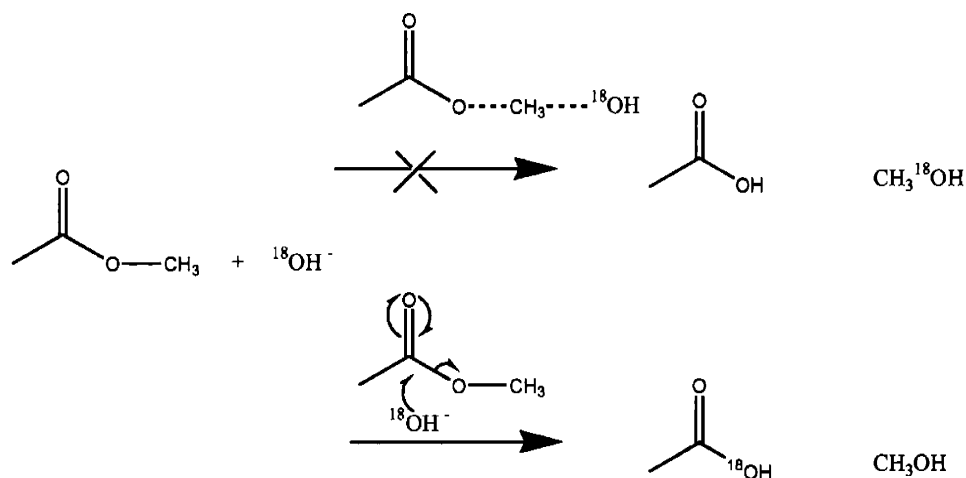

Figure 1: Rate analysis alone is unable to differentiate two proposed mechanisms for ester hydrolysis. A second experiment using isotopically labeled oxygen and monitoring the final products falsifies the hypothesis of alkyl oxygen cleavage, leaving the acyl oxygen cleavage mechanism.

Over the past seven years, a variety of experiments have been carried out in the effort to explore the structural variables that affect the helicity of polyalanine peptides. Many exploratory leads were abandoned as more suitable helical hosts were developed; NMR experiments were limited by instrument time and the lack sufficient of departmental support. Although these experiments have provided publishable results, further work is necessary to understand their full scope. This thesis presents a selected series of fundamental experiments that have impacted the direction of research in this area and in the Kemp group. The work presented here represents the hypotheses and conclusions that can be supported at this time. Until the controversial aspects of peptide helices are resolved, using multiple analytical methods provide a rigorous cross check for potential falsification results.

"I wish the ring had never come to me...I wish none of this had happened."

"So do all who live to see such times, but that is not for them to decide.
All we have to decide is what to do with the time that is given to us."

The Lord of the Rings: The Fellowship of the Ring, 2001

TABLE OF CONTENTS

TABLE OF ABBREVIATIONS

Ala (A)	Alanine
Lys (K)	Lysine
Orn (O)	Ornithine
Arg (R)	Arginine
β	β-amino-alanine
$^{\beta}$Asp ($^{\beta}$D)	Aspartate, β-linked by the side chain
Inp	Isonipecotic acid
^{t}L	*tert*-leucine
PDB	Protein data base
LR	Lifson-Roig
w	Helical residue propensity
n	Peptide length
k	Length of helical conformer $3 \leq k \leq n$
w_k	Length dependent w value applied to conformers of length k
v^2	Lifson-Roig helical initiation constant
SS_n	State Sum (total weight of all conformations in equilibrium for peptide of length n)
$\chi_{\alpha Ci,H}$	Mole fraction of helical (H) α-carbons at residue site i
FH	Fractional Helicity (total number of helical α-carbons/ total α-carbons)
CD	Circular Dichroism
$[\theta]$	Molar ellipticity (deg cm^2 dmol^{-1})
$[\theta]_{n,222,obs}$	Experimental perresidue molar ellipticity (deg cm^2 dmol^{-1} res^{-1}) at 222 nm
$[\theta]_{n,222}$	Limiting perresidue molar ellipticity (deg cm^2 dmol^{-1} res^{-1}) for a helix of length n at 222 nm
$[\theta]_{\infty,222}$	Perresidue molar ellipticity for a helix of infinite length (deg cm^2 dmol^{-1} res^{-1}) at 222 nm
x	Length correction for the limiting ellipticity $[\theta]_{n,222} = [\theta]_{\infty,222} (1\text{-}x/n)$
X	Residue mutant site

TABLE OF ABBREVIATIONS (CONT.)

PF_i	Protection Factor at residue site i
k_{se}	Rate constant for a completely solvent exposed amide NH
k_{obs}	Rate constant for a potentially helical amide NH
$\chi_{NHi,se}$	Mole fraction of solvent exposed NHs at residue site i
$\chi_{NHi,hb}$	Mole fraction of hydrogen bonded NHs at residue site i
k_f	Proposed fast rate constant for the first three helical residues with solvent exposed amide NHs at residue site i
χ_f	Mole fraction of the first three helical residues with solvent exposed amide NHs at residues site i
k_n	Normal rate constant for solvent exposed residues
χ_n	Mole fraction of non-helical residues of solvent exposed residues
Ac-Hel	Constrained Ac-Proyl-Proline stabilizing N- Cap:(2S, 5S, 8S, 11S,)-1-Acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricylco[2.8.2.0^{4,8}]-tridecane (Ac-Hel-OH)
t/c	Ratio of *trans* and *cis* conformers of Ac-Hel. t/c = [ts + te] /[cs]
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
HSQC	Heteronuclear Single Quantum Coherence
MAS	Magic Angle Spinning
AUC	Analytical ultracentrifugation

CHAPTER 1: INTRODUCTION AND BACKGROUND

INTRODUCTION

This thesis focuses on the energetics of helices formed by peptides. For 40 years

this field has witnessed an intense effort driven by the desire to incorporate this energetic

information into a complete understanding of the analogous protein folding problem, the

folding of an amino acid sequence into native structures in which helical regions appear.

Although globular protein structures also contain other secondary structural elements,

e.g. sheets, loops, and turns, approximately 30% of the amino acids in the globular

protein data base lie within helices.[1] Helices have been postulated as early intermediates[2]

in protein folding pathways, and some biologically important protein classes contain only

helices.[3] An algorithm that reliably predicts the amount of helical character (helicity) and

its helical stability would provide a fundamental tool to elucidate protein folding.

Two distinct types of algorithms, qualitative and quantitative (energetic), have

been constructed through the analysis of protein helices. The first algorithms were

qualitative, based on the known structure of crystalline proteins in the protein database.

These identified the amino acids seen most often in sheets, turns and helices ranking their

relative preferences for each.[4] Widely used to predict regions of secondary structure

from a primary amino acids sequence, these algorithms are good at predicting helical

regions, but only moderately successful in predicting the overall structures of globular

---

[1] Barlow, D. J.; Thornton, J. M. "Helix Geometry in Proteins" *J. Mol. Biol.* **1988**, *201*, 601-619.

[2] Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. "The Complete Folding Pathway of a Protein from Nanoseconds to Microseconds" *Nature* **2003**, *421*, 863-867; Laurents, D. V.; Baldwin, R. L. Protein Folding: Matching Theory and Experiment" *Biophys. J.* **1998**, *75*, 428-434.

[3] Examples of such proteins include: tropomyosins, keratins and leucine zippers.

[4] Chou, P. Y.; Fasman, G. D. "Prediction of Protein Conformation" *Biochemistry* **1974**, *13*, 222-244.

proteins.[5] Although unable to predict helical energetics, they can identify amino acids that frequently and rarely appear in helices.

A second class of protein studies assigns to each amino acid a helical propensity (w), the intrinsic tendency to stabilize or destabilize a helix. These are the fundamental variables for any quantitative, energetic helicity algorithm. Summed for all of the amino acids in the protein sequence, the w values and related energetic terms describe the energetic stability of the native structure vs. the unstructured state. A spectroscopic technique such as circular dichroism (CD) that measures conformational change can be applied to calculate the change in free energy (ΔG) for a protein that undergoes a reversible two-state denaturation as seen in Figure 1. The relative changes in free energy ($\Delta\Delta G$) of denaturation for a series of site mutations, where one amino acid site within a helical region has been varied, can be used to calculate a scale of relative w values.

Protein Native Structure   A   Unstructured Coil Conformations
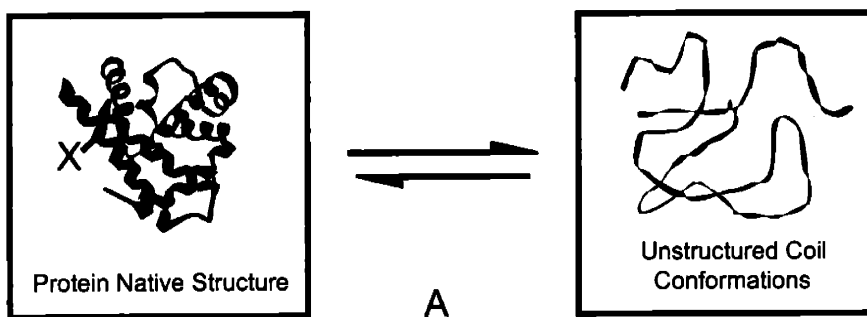
Figure 1: Thermal or chemical protein denaturation results in the loss of native structure, forming a manifold of molecules in unstructured conformations. Owing to folding cooperatively, the native structure can be approximated as a single energetic state and the diverse unstructured conformations are grouped as a single reference state defining the two-state transition. *X* marks a potential solvent exposed mutant site, within a helical region of the native protein.

---

[5] Schulz, G. E.; Schirmer, R. H. In *Principles of Protein Structure* Cantor, C. R. Ed., Springer-Verlag, New York, **1979**, 108-130.

Blaber and Matthews[6] carried out such a study on T4 lysozyme. Solvent exposed mutant sites in helical regions were selected to minimize sensitivity of *w* to other stabilization effects such as hydrophobic packing of secondary structures or side chain crowding by neighboring residues. X-ray crystallography was used to verify that the native helical structure was not perturbed by neighboring amino acids side chains (site dependence) at each of the two independent sites where substitutions were made. The results were then analyzed to assign a set of helical propensities for the 20 natural amino acids. Consistent *w* values for the 8 non-polar amino acids were assigned at both sites.

Can this set of helical propensities be used to predict the effect of mutants on the stability of other proteins? This important question was addressed by Myers, Pace and Scholtz,[7] who applied the same analysis to define helical propensities within a 17 residue helix of RNase T1. A linear regression of the *w* values of Blaber and Scholtz showed a slope of 0.92 and a correlation coefficient 0.93. The two data sets span the same range and correlate linearly. This agreement indicates that if mutant sites are selected with care, helical propensities can indeed be generally applicable to other proteins. The Scholtz study also investigated whether or not this scale predicts the effects of mutant studies on the isolated helical protein fragment, free of folded protein interactions. A series of peptide mutants were synthesized; their helical propensities were assigned by CD and were found to agree with those assigned for the native protein host. Similar

---

[6] Blaber, M.; Zhang, X.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W. A.; Matthews, B. W. "Determination of α-Helix Propensities within the Context of a Folded Protein" *J. Mol. Biol.* **1994**, *235*, 600-634.
[7] Myers, J. K.; Pace, C. N.; Scholtz, J. M. "A Direct Comparison of Helix Propensity in Proteins and Peptides" *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2833-2837.

agreements of this type have been reported by others.[8,9] Although peptides and proteins have intrinsically different folding properties, factors that increase the stability of an isolated helical peptide can also increase the stability of the overall protein in which it is a component to an equivalent degree. Both peptides and proteins thus can be used as hosts for carrying out site mutation studies to define helical propensity scales.

Protein site mutations clearly provide an essential calibration of the parameters used to construct helicity algorithms, but during the past forty years, studies of the simpler, experimentally tractable helical peptides have largely been used to generate those parameters. Ease of peptide synthesis allows the formation of large databases in which site mutants within many types of peptide hosts are studied. These yield information about helical structure and its response to interpretable chemical variables such as amino acid sequence, length, temperature and solvent, that are difficult or impossible to study in a protein context. As seen in Table 4, many sets of helical parameters measured from within different hosts are currently available, and in some cases quantitative parameters that define the energetics of local interactions between side chains of pairs of neighboring amino acids are also available.[10] More recently, a number of workers have analyzed large databases constructed from studies of both peptide and

---

[8] O'Neill, K. T.; DeGrado, W. F.; "A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids" *Science* **1990**, *250*, 646-651.

[9] Horovitz, A.; Matthews, J. M.; Fersht, A. R. "Alpha-helix stability in proteins. II. Factors that Influence Stability at an Internal Position" *J. Mol. Biol.* **1992**, *227*, 560-568.

[10] Chakrabartty, A.; Baldwin, R. L. "Stability of α-Helices" *Adv. in Protein Chem.* **1995**, *46*, 141-176 and references therein; Padmanabhan, S.; Baldwin, R. L. "Helix-Stabilizing Interaction between Tyrosine and Leucine or Valine When the Spacing is *i, i* + 4." *J. Mol. Biol.* **1994**, *241*, 706-713.

protein helices, to generate algorithms[11] that can predict with reasonable accuracy the

helical properties of most peptide sequences that occur in globular proteins.

Crystal Database	Polymer	Peptide	Protein
Dirkx[12] PDB 1975	Scheraga[13] HBLG & HPLG	Baldwin[14] -AAKA-X-AAKA-	Ferscht[9] Barnase -KSAQ-X-LG-
Fasman[4] PDB 1974	Blout[15] Poly-Glu, Poly-Lys	Kallenbach[16] -KKK-XXX-EEE-	Matthews[6] T4 lysozyme -QAAK-X-ELDK-
Scheraga[17] PDB 1974		Stellwagen[18] -AKEA-X-AKEA-	Degrado[8] Coiled-Coils -AALE-X-KLQA-
		Scholtz[7] RNase T$_1$ -STAQ-X-AAYK-	

Table 1: List of investigators (non-inclusive) and the hosts used to measure the helical propensities for the 20 natural amino acids. Protein and peptides used site mutants at site X.

[11] Munoz,V.; Serrano, L. "Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters" *Struct. Biol.* **1994**, *1*, 399-409.

[12] Beghin, F.; Dirkx, J. "Une Methode Statistique Simple de Prediction des Conformations Proteiques" *Arch. Int. Physiol. Biochim.* **1972**, *80*, 185-187.

[13] Wojick, J.; Altman, K. H.; Scheraga, H. A. "Helix-coil Stability Constants for the Naturally Occurring Amino Acids in Water. XXIV. Half-cystine Parameters from Random Poly(hyroxybutylglutamine-co-S-methylthio-L-cysteine" *Biopolymers* **1990**, *20*, 121-134.

[14] Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. "Helix Propensities of the Amino Acids Measured in Alanine-Based Peptides without Helix-Stabilizing Side-Chain Interactions" *Protein Sci.* **1994**, *3*, 834-853.

[15] Blout, E. R.; Loze, C.; Bloom, S. M.; Fasman, G. D. "The Dependence of the Conformations of Synthetic Polypeptides on Amino Acid Composition" *J. Am. Chem. Soc.* **1960**, 82, 3787-3889.

[16] Kallenbach, N. R.; Lyu, P.; Zhou, H. In Circular Dichroism and the Conformational Analysis of Biopolymers ed. Fasman G. D. (Plenum, New York) **1996**, 343-352.

[17] Kotelchuck, D.; Scheraga, H. A. "The Influence of Short-Range Interactions on Protein Conformation. I. Side Chain-backbone Interactions within a Single Peptide Unit" *Proc. Natl. Acad. Sci. USA* **1968**, *61*, 1163-1170.

[18] Park, S. H.; Shalongo, W.; Stellwagen, E. "Modulation of the Helical Stability of a Model Peptide by Ionic Residues" *Biochemistry* **1993**, *32*, 121-134.

Three important points must be made about existing parameter sets and the energetic algorithms that have been constructed from them. 1) No single algorithm predicts the helicity of all classes of peptides with reliability or accuracy. 2) Only moderate consistency is observed between different helical propensity scales. 3) The existence of these scales has not resolved fundamental disagreements concerning the degree to which host peptides are free from context effects, the reliability of the analytical techniques used to monitor helicity, or the intrinsic helical properties of the simplest helical peptides. New experimentation is needed to solve these issues in order to create an accurate and complete algorithm that predicts structure from amino acid sequence. This thesis will make a contribution toward establishing reliable analytical techniques and investigate simple helical peptides.

To frame these studies, it is necessary to review key features of previous work focusing on simple polyalanine peptides. Peptide helicity is fundamentally a problem of conformational analysis, which has two aspects, structural and energetic. The conformational analysis of a molecule requires a series of structural and energetic steps where: 1) A molecule can be assigned to a series of defined conformational states. 2) Each conformational state is structurally characterized. 3) A temperature range is defined where the conformational manifold is at equilibrium. 4) A spectroscopic method or model can be applied to the manifold to quantify the amount of each conformation, expressed as condition dependent mole fractions χ_i. The remainder of this chapter will focus on how these requirements are fulfilled for helical peptides in solution and how the problems with this fulfillment have generated the controversies regarding methods and interpretation.

The first section that follows presents key structural features of peptide helices; the second and third analyze experimental features of the formation of helical conformations by homo-oligopeptides for which convenient preparative routes were discovered[19] in the 1940s. The energetic properties of their conformations and the "helicity algorithms" developed specifically to model and predict their stability and structural properties have provided the foundation for all subsequent studies of helical peptides in solution. One of those helicity algorithms,[20] introduced in 1962 by Lifson and Roig,[21] (LR) with minor modifications is nearly universally used for modeling peptide helicity. Following a brief summary of major, subsequent developments in the measurements of w values, with their attendant controversies, the chapter closes with a focused restatement of the aims of this thesis.

BACKGROUND

HELICAL STRUCTURE

The α-helical structure of polypeptides was first assigned by Pauling and Corey[22] in 1952 by utilizing bond distances and angles measured from X-ray structures of small molecules. Pauling verified a key prediction of his theory of resonance[23] when the amide

---

[19] Woodward, R. B.; Schramm, C. H "Synthesis of Protein Analogs" *J. Am. Chem. Soc.* **1947**, *69*, 1151-1152.

[20] Future use of the word algorithm will refer specifically to the Lifson-Roig helicity algorithm used to assign w values for peptide helices.

[21] Lifson, S.; Roig, A. "On the Theory of Helix-Coil Transition in Polypeptides" *J. Chem. Phys.* **1961**, *34*, 1963-1974.

[22] Pauling, L.; Corey, R. B. "Two Hydrogen-bonded Spiral Configurations of the Polypeptide Chain" *J. Am. Chem. Soc.* **1950**, *72*, 5349-5349; Pauling, L.; Corey, R. B." The Structure of Synthetic Polypeptides" *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 241-250.

[23] Pauling, L.; Corey, R. B. "The Planarity of the Amide Group in Polypeptides *J. Am. Chem. Soc.* **1952**, *74*, 3964-3964.

bond that links dipeptides was found to be planar. Incorporating these structural constraints for the repeating peptide unit, Pauling noted that the atomic coordinates of the helix are described by only two dihedral angles at the rotatable bonds ϕ, ψ. Utilizing known distances for hydrogen bonds and assuming the atoms of the amide NH, the hydrogen bond donor, and the amide carbonyl CO of the hydrogen bond acceptor are nearly colinear, as seen from small molecule X-ray data, Pauling proposed two helical structures. The right handed α-helix for peptides with L-amino acids was proposed with $\phi = -47°$, $\psi = -57°$, and its structure was shown to be consistent with the periodicity previously defined by others from fiber diffraction X-ray scattering data for α-keratins.[24] Although the dihedral angle definition is universally used, it is worth noting that this structure is independently defined by a repeating pattern of hydrogen bonds between the ith residue's carbonyl and the NH of the $(i + 4)$ residue as seen in Fig 4. These define a series of 13-membered rings.

A   B
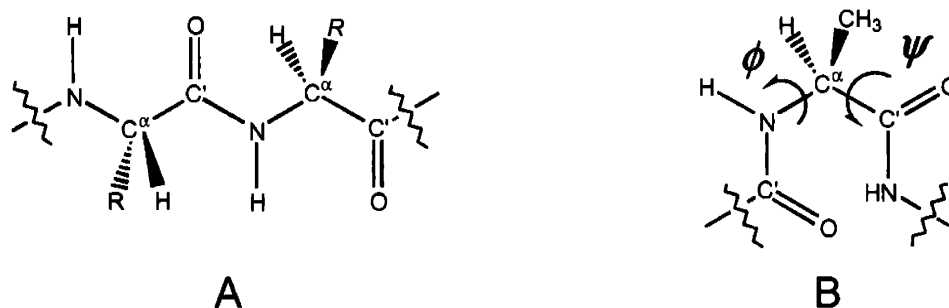
Figure 2: A) The more stable *trans*-conformation of the amide bond places atoms between two adjacent C^{α}'s in a plane; $\phi = \psi = 180°$. B) The rotatable bonds ϕ, ψ of the repeating N-C^{α}-C' peptide unit define the conformation of the polypeptide chain; $\phi = \psi = 0°$ is pictured. The 20 natural amino acids have unique R groups, the methyl group for alanine is shown.

---

[24] Pauling, L.; Corey, R. B.; Branson, H. R. "The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain" *Proc. Natl. Acad. Sci. U.S.A* **1951**, 37 205-11.

Figure 3: Three successive helical pairs of ϕ, ψ dihedral angles for Ac-Ala$_3$-NH$_2$ create the first turn of an α-helix with the formation of the first hydrogen bond, and defines a hydrogen bonded ring of 13 atoms. As illustrated, amino acids residues must be capped by amides in order to adopt helical ϕ, ψ values. Without amide caps a peptide like H$_3$N$^+$-Ala$_3$-CO$_2^-$ is unable to form helical hydrogen bonds.

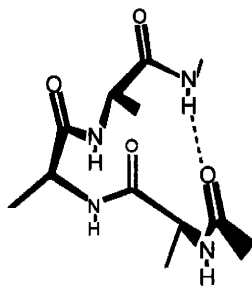A   B
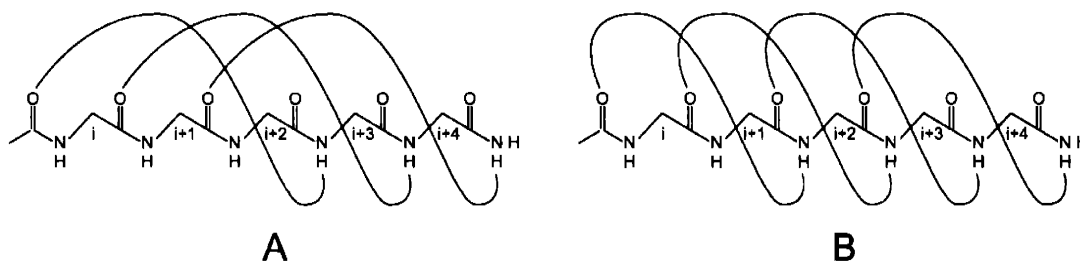
Figure 4: A) α-helical hydrogen bonding pattern between the carbonyl CO of the i and $i + 4$ amide NH sites incorporates sites $i + 1$, $i + 2$, $i + 3$ in the helical turn. B) 3_{10}-helix hydrogen bonding pattern between the i and $i + 3$ sites incorporates sites $i + 1$, $i + 2$ in the helical turn defined by a series of hydrogen bonded ring of 10 atoms.

The relevance of Pauling's α-helical model to the structures of globular proteins was definitively confirmed in 1958 when Kendrew and coworkers[25] solved the first high resolution X-ray structure for the oxygen-carrying muscle protein myoglobin. This protein was shown to have 85% of its residues within structures that fit Pauling's parameters for the α-helix. Analysis of the extensive protein database now available indicates that 30% of residues[1] typically appear in helices with the α-helical hydrogen

[25] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis" *Nature* **1958**, *181*, 662-666.

bonding pattern. A minor alternative, the 3_{10}-helix, with an i to $i + 3$ hydrogen bonding pattern, may appear in short loops or at the ends of α-helices.

Within protein helices with the α-helical hydrogen bonding pattern, the ϕ, ψ angles were frequently found to vary from the values proposed by Pauling. From analyses of the X-ray structural database Barlow and Thornton[1] have found average values of ϕ = -62 ± 7°, ψ = -41 ± 7°. However, values for ϕ and ψ are strongly correlated, the α-helical hydrogen-bonding pattern is maintained by an approximate relationship of $\phi + \psi \approx -103°$.[26] An increase in $-\phi$ results in an increase in the tilt of the amide plane with respect to the helical axis, negating Pauling's assumption of the necessity of collinear hydrogen-bonds. This plausible ϕ, ψ variation is also seen when gas phase molecular modeling is carried out on polyalanine helices.[27]

Figure 5 embodies the structural characteristics of an α-helix. The *trans*-amides are aligned with respect to the helical axis and point in the same direction, the CO toward the C-terminus and the NH toward the N-terminus, thus creating three chains of hydrogen bonds that run along the helix barrel. The chirality of L-amino acids dictates the right handedness of the helix, with all side chains pointing toward the N-terminus. In a helix three types of residue sites can be distinguished: a core region, and a pair of N- and C-terminal regions. Each amide in the core region forms two intra-helical hydrogen bonds. Moreover, the repeat of a 3.6 residue per helical turn seen in Figure 5B allows residue side chains at site i to contact residue side chains at sites: $(i - 3)$, $(i - 4)$, $(i + 3)$ and $(i + 4)$;

[26] Besley, N. A.; Hirst, J. D. "Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations" *J. Am. Chem. Soc.* **1999**, *121*, 9636-9644.

[27] Mahadevan, J.; Lee, K. H.; Kuczera, K. "Conformational Free Energy Surfaces of Ala_{10} and Aib_{10} Peptide Helices in Solution" *J. Phys. Chem B.* **2001**, *105*, 1863-1876.

within a loop of three residues, the side chains of its residues could have up to 12 contacts with residues in the proceeding loop. Structurally speaking an α-helix has exceptional integrity, since distortion of its regular cylindrical shape would involve the loss of potentially stabilizing side chain contacts and breaking of hydrogen bonds.

Residues that appear at the N- and C-termini form only one intra-helical hydrogen bond per amide residue, leaving an unsatisfied hydrogen bonding valence. Structural analysis by Rose et al.[28] within the protein database shows that these vacant amide hydrogen bonding sites are frequently filled by hydrogen bond donors and acceptors supplied by non-helical amino acids with polar side chains. In addition, positive charges from side chains and substrates have been found close to helical C-termini and negative charges have been found situated near N-termini.[29] These have been rationalized as creating favorable helix stabilizing dipole-charge interactions with the three aligned amides with vacant bonding valences, Figure 5A.

---

[28] Aurora, R.; Rose, G. D. "Helix Capping" *Protein Sci.* **1998**, *7*, 21-38.

[29] Forood, B.; Feliciano, E. J.; Nambiar, K. P. "Stabilization of α-Helical Structures in Short Peptides via End Capping" *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 838-842. Schneider, J. P.; DeGrado, W. F. "The Design of Efficient α-Helical C-Capping Auxiliaries" *J. Am. Chem. Soc.* **1998**, *120*, 2764-2767.
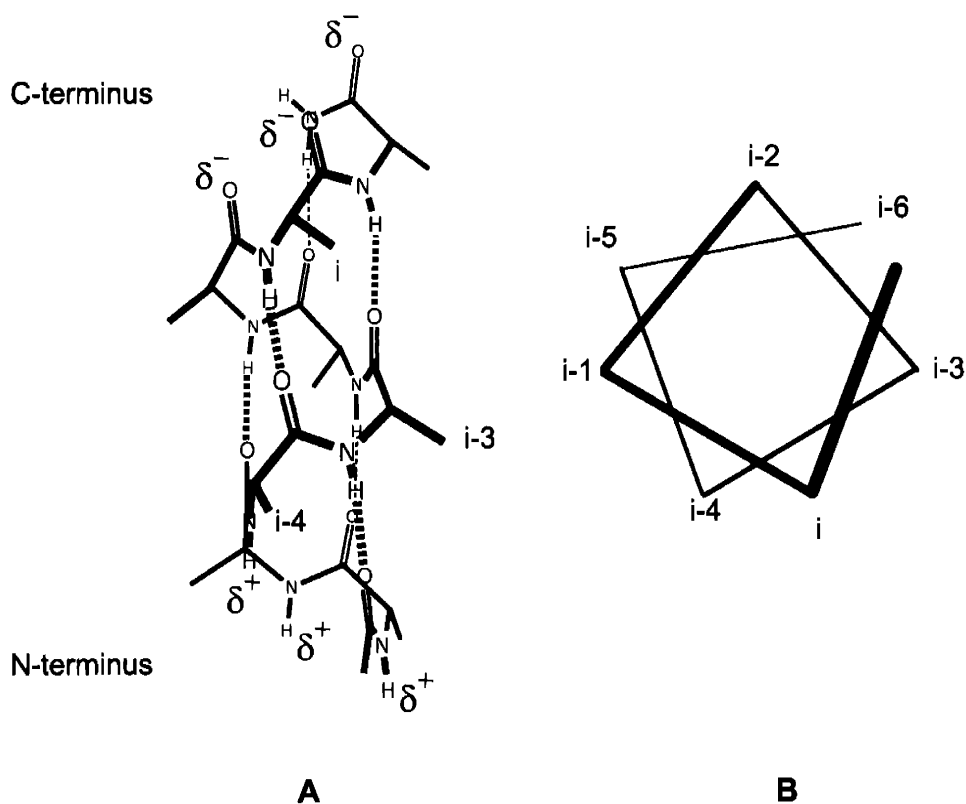
Figure 5: Two α-helical representations A) Perpendicular to the helical axis with *i*, (*i* - 3), (*i* - 4) labeled side chains and N- and C-terminal dipole designations B) Along the helical axis the helical wheel indicates the spatial distribution of the amino acid side chains. With particular amino acids side chains, pairs of amino acids at *i*, (*i* ± 3) or *i*, (*i* ± 4) separations can be in van der Waals contact.

STRUCTURAL ENERGETICS

A Ramachandran diagram[30] provides a first step toward understanding the energetics of helix formation. It plots isoenergetic contours that are calculated for the changes in energetics of a single residue α-carbon flanked by two amides as a function of the values for its ϕ, ψ torsional angles. Figure 6 shows the diagrams for alanine and glycine, with contours that enclose energetically accessible conformations under normal

[30] Ramachandran, G. N.; Sasisekharan, V. "Conformation of Polypeptides and Proteins." *Adv. Protein Chem.* **1968**, *23*, 283-438.

conditions. Ramachandran diagrams for most other amino acids, proline excluded, are similar to the alanine diagram, reflecting the large difference between the small glycine hydrogen side chain and the varying C^β bulk of the other amino acids.
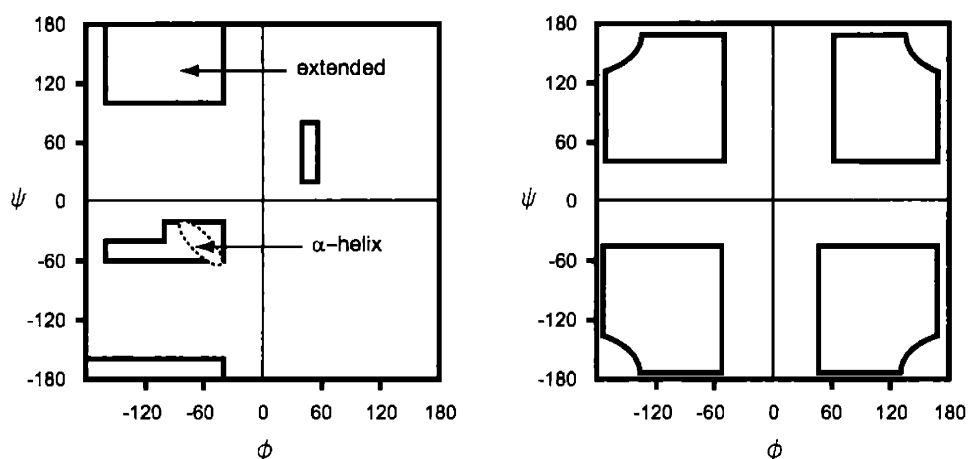
Figure 6: Ramachandran diagrams with approximate contours. A) Alanine has a large area of conformational freedom in the upper left quadrant. The helical region, marked with the dashed ellipse is of a slightly higher energy. B) Glycine has large accessible areas in all four quadrants.

Newman projections of the ϕ and ψ angles provide a useful visualization of the sterics that govern the Ramachandran diagram. Keeping the C^α carbon fixed, one can explore the energetically accessible space by rotating the adjacent amides. As seen in Figure 7C, the Newman projection illustrates how the average helical $\phi = -47°$ staggers the larger CO of amide 1 between the $C^\alpha H$ and CO of amide 2 and eclipses the smaller H with the potentially bulky side chain R. The wide ϕ range, between $-60°$ and $-130°$, of the helical slot apparent in the Ramachandran diagram reflects that a large variation in ϕ does not result in a significant increase in steric crowding.
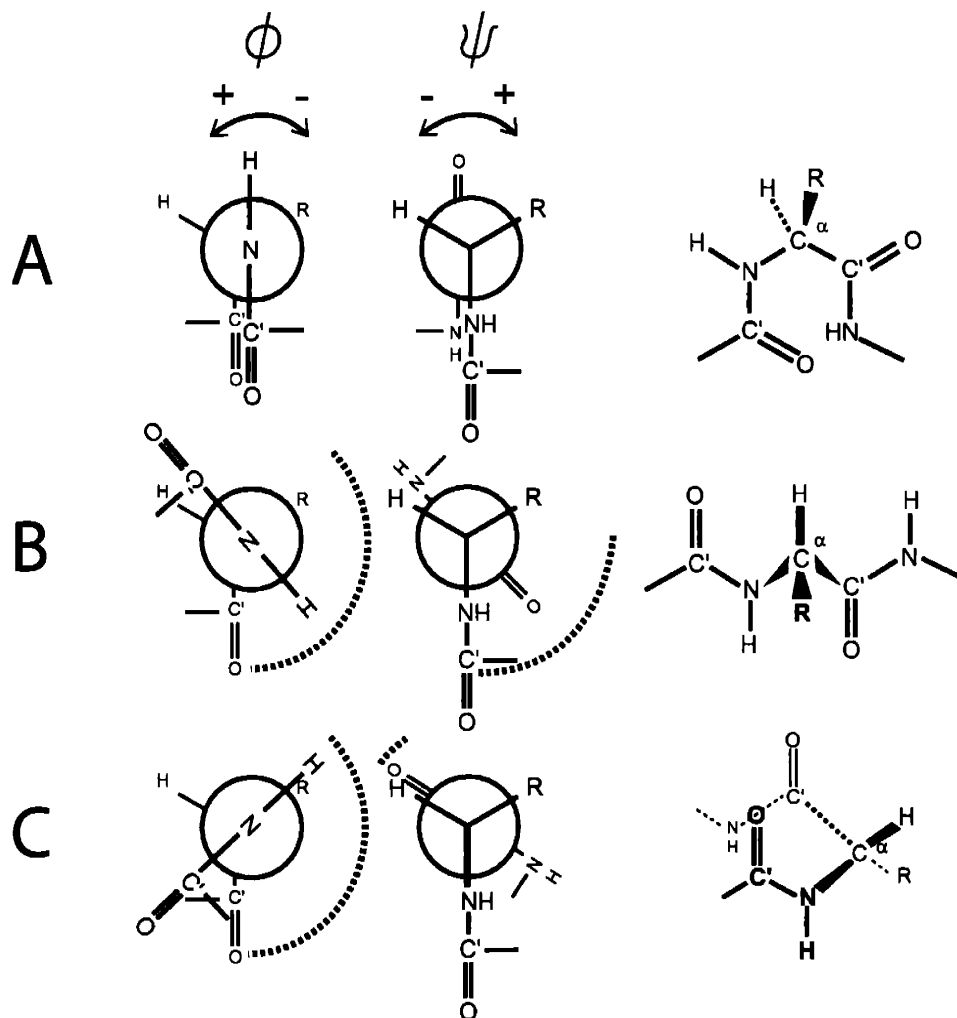
Figure 7: Newman projections and the resulting 3D backbone conformation. The dotted arcs represent the energetically accessible space for the NH of ϕ and the CO of ψ. A) Reference for $\phi = \psi = 0°$. B) Random conformation with $\phi = -139°$ and $\psi = +135°$ shown. C) Pauling's helix $\phi = -47°$, $\psi = -57°$ is pictured.

As seen in the Newman projection the helical $\psi = -57°$ angle orients the relatively

large amide NH between two bulky groups, the side chain R and the NH of amide and

eclipses the small CO with the small $C^\alpha H$. The narrow range of the ψ dimension of this

helical slot reflects the constraint of this angle and the large increase in crowding that

results when ψ is varied. Earlier it was noted that the analysis of the PDB correlates ϕ +

$\psi = -103°$. The ellipse defined by this analysis is also shown in Figure 6A. It overlaps

with the accessible Ramachandran prediction, but its restricted area represents the required hydrogen-bonding patterns for an α-helix.

The large upper left region of the Ramachandran diagram contains a variety of conformations, including all that have been assigned by molecular modeling as likely contributors to residues representing a disordered or unstructured peptide backbone. For the average conformation in this region, relative to helices, the distance between α-carbons at sites i and $(i + 2)$ along the peptide backbone is 25-30% larger. Typical local conformations that result in this increase are analyzed in the Newman projection in Figure 7B. These conformations place adjacent side chains on opposite sides of the back bone and minimize intra-molecular hydrogen bonding, leaving backbone amides solvated in protic solvents.

Relying on the Ramachandran analysis and based on polymer models, Flory[31] deduced that the conformational energetics of a polypeptide can be modeled as a sum of the energetics of each of its residues. Flory assumed that based on its ϕ, ψ dihedral angles each amino acid can be assigned a plausible energetic weight that is independent of conformations adopted by its adjacent residues.[32] Lifson and Roig (LR) constructed a related two-state helicity model in which a residue is classified as either helical or nonhelical. If a pair of ϕ, ψ angles lie in the non-helical Ramachandran space (abbreviated by c), the residue is assigned a reference weight of 1. If the ϕ, ψ pair lies in the more restricted, higher energy helical space (abbreviated by h) the residue is assigned the variable v which can be assigned by rigorous modeling of the Ramachandran

[31] Flory, P. J. In *Statistical Mechanics of Chain Molecules* Wiley & Sons: New York, **1969** 250-274.
[32] The extended nature of the backbone prevents stabilizing contacts.

energetics, or empirically though study of quantitative helicity properties of peptides. The value of v is expected to be < 1 considering that the non-helical regions include a large possible range of ϕ, ψ angles and that the smaller helical region has a higher energy. The ratio of accessible helical and unstructured Ramachandran spaces provides an approximation of $v = 0.05$, consistent with modeling results that are detailed subsequently. The overall weight of the peptide conformation[33] is a product of the $c = 1$ and $h = v$ weights of each of its residues, as seen in Figure 8. Nonhelical conformations by definition can contain no more than two contiguous h residues. The number of conformations in the random coil or non-structured state thus equals the possible ways of making single and, or double c→h substitutions to the all c conformation. Conformations with multiple isolated h terms are less probable, and despite their numbers make a smaller contribution to the total of non-structured weights, for medium length peptides.

$$
\begin{array}{ccccc}
\text{c} \quad \text{c} \quad \text{c} \quad \text{c} \quad \text{c} & \rightarrow & \text{c} \quad \text{c} \quad \text{c} \quad \text{h} \quad \text{c} & \rightarrow & \text{c} \quad \text{h} \quad \text{c} \quad \text{h} \quad \text{c} \\
1\times1\times1\times1\times1 = 1 & & 1\times1\times1\times v\times1 = v & & 1\times v\times1\times v\times1 = v^2
\end{array}
$$

Figure 8: Energetic assignments of weights for three unstructured conformations of a pentapeptide. The overall weight of the entire conformation is a product of the $c = 1$ and $h = v$ weights of each of its residues.

The LR parameters define equilibria between paired conformations with h, c conformation designators that differ at a single α-carbon. For example conformations A and B of Figure 9 are inter-converted by a c→h change at the central α-carbon of the tripeptide. Because all amide functions of A and B are solvent exposed and the steric interactions of their side chains are assumed to be modeled accurately by Ramachandran

---

[33] With in the LR weighting scheme "conformation " is used to describe the helical and nonhelical portions of a peptide chain, rather than specific ϕ, ψ angles, h conformations can sample the helical Ramachandran space and c conformations can sample the random space.

energetics, the appropriate equilibrium constant is v. Because conformation C contains

an intra-molecular hydrogen bond and the interaction of the side chain cannot be modeled

fully by local interactions mirrored by the v parameter, LR introduced a new parameter w,

the helix propensity as the B ⇌ C equilibrium constant.
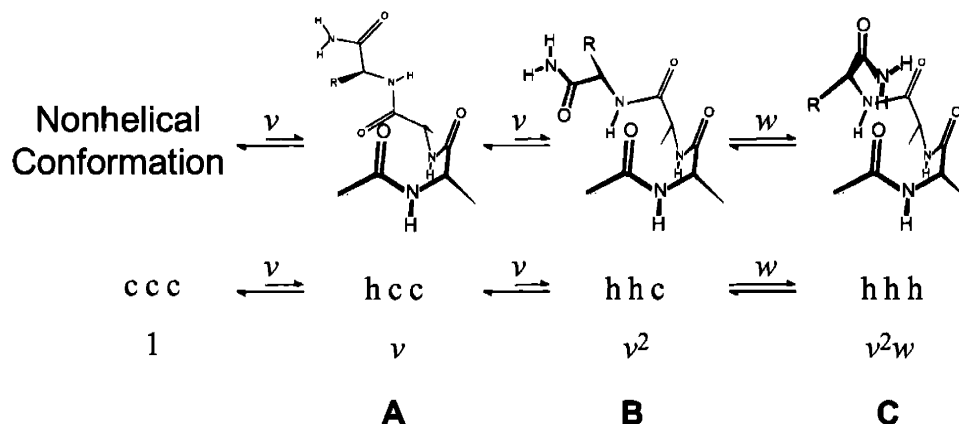
Figure 9: The helical manifold for a tripeptide illustrating the relationship between structure and the peptide conformer h, c identifiers. The LR algorithm models the formation of single and double helical ϕ, ψ pairs (A⇌B) from random conformations as v; w represents a nonhelical residue joining an existing helix (B⇌C). The weights assigned by LR relative to the nonhelical conformation are also listed. Note that the c-state structure is represented by the extended conformation $\phi = \psi = 180°$.

In the LR model for a homopeptide, w is employed as a universal helicity

parameter, used when a tripeptide helix is formed and when a helix of any length is

extended by one residue. The structural regularity of the core region of a helix justifies

the assumption that the helical propensity remains constant; the LR weight of a helical

conformation of residue length k is thus $v^2 w^{k-2}$ ($3 \leq k \leq n$). Amino acids with an

indifferent helical propensity have w values close to 1, indicating a 50% population of the

helical and non-helical conformations at that site. Helix breaking amino acids have w

values < 1 and helix stabilizing amino acids have w values > 1.

The capped pentapeptide ($n = 5$) example in Table 2 illustrates how this two-state approximation creates a manifold of 2^n or 32 total conformations and conformation weights. One can list the conformations with different helical lengths (k) and positions starting with the longest helical conformation. There is only one way to place 5 helical residues in a pentapeptide (hhhhh), hence there is only a single completely folded helical conformation. There are two tetrapeptide helices, chhhh and hhhhc, and three major and two minor (hhhch + hchhh) tripeptide helices.[34] Summing the weights for all of the conformations in Table 2 yields the state sum (SS) in Figure 10.

Helical res.	Helical Conformations	Helical Weight	#	Non-Helical Conformations	Random Weight	#
5	hhhhh	v^2w^3	1			
4	hhhhc + chhhh hhhch + hchhh	v^2w^2 v^2wv	2 2	hhchh	v^4	1
3	hhhcc + cchhh + chhhc	v^2w	3	hhcch + hhchc + hcchh + hchch + hchhc + chhch + chchh	v^3	7
2				hhccc + chhcc + cchhc + ccchh + cchch + chcch + chchc + hccch + hcchc + hchcc	v^2	10
1				hcccc + chccc + cchcc + cccchc + cccch	v	5
0				ccccc	1	1
Total # of helical conformations			8	Total # of nonhelical conformations		24

Table 2: List of helical and nonhelical conformations for a capped pentapeptide.

---

[34] Additional v terms are added for isolated single or paired h residues that appear within helical conformations. A final LR assumption concerns conformations that contain two independent helical regions. Consistent with the analysis of Figure 9, these are weighted as two independent helices [chhhcchhh = $v^2w \times v^2w = v^4w^2$ and chhhhhhcc = v^2w^4]

$$SS_5 = 1 + 5v + 10\,v^2 + 7\,v^3 + v^4 + 3\,v^2w + 2\,v^3w + 2\,v^2w^2 + v^2w^3$$

Figure 10: State sum for a capped pentapeptide can be generated by totaling the weights in Table 2 or using LR matrix as seen in Equation 2.

The state sum contains a 5-term v-polynomial, the weight of the non-helical reference state, and the remaining polynomial in v and w is the weight of the helical state. If v and w are known, the state sum is easily converted into mole fractions of states of a particular length or individual conformations based on helical α-carbons ($^\alpha C_i$). For the pentapeptide example, the mole fraction of the non-helical state ($\chi_{\alpha C_i,\ k<3}$) and the fully helical conformation ($\chi_{\alpha C_i,\ k=5}$) are calculated in Equation 1.

$$A)\quad \chi_{\alpha C_i,k<3} = \frac{1 + 5v + 10v^2 + 7v^3 + v^4}{SS_5} \qquad B)\quad \chi_{\alpha C_i,k=5} = \frac{v^2w^3}{SS_5}$$

Equation 1: Calculation of mole fractions A) The nonhelical $\chi_{\alpha C_i,\ k<3}$ is the total weight of the pure polynomial in v within SS$_5$ divided by all of SS$_5$. B) $\chi_{\alpha C_i,\ k=3}$ is the weight of the single helical conformation that has length 5 divided by SS$_5$.

The simple example of Table 2 demonstrates the underlying reasoning that is always used to construct the LR state sum from a set of weights that are paired with h, c designators. In practice this state sum is rarely written out explicitly, but is calculated as a matrix product of two vectors and a series of n matrices, one for each amino acid residue in the potentially helical peptide.[21,35] Since w depends upon the nature of the peptide side chain, LR state sums for heteropeptides must be calculated from matrix

---

[35] Kemp, D. S. "Construction and Analysis of Lifson-Roig Models for the Helical Conformations of α-Peptides" *Helv. Chim. Acta.* **2002**, *85*, 4392-4423.

products in which the w value of the ith matrix corresponds to the ith residue in the peptide sequence. A LR state sum for a peptide made up of j different residues thus depends on $j + 1$ distinct variables, j w's and 1 v. The LR matrix algorithm is also able to accommodate other parameters that affect the helicity of short peptides. However, since the focus of this thesis will remain on the fundamental w value for alanine homopeptides, the matrices used are all identical, and the state sum depends on only two variables w_{Ala} and v.

$$SS_n = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Equation 2: Basic 4×4 L-R matrix that uses a two state approximation to weight all of the 2^n conformations in solution.

For practical reasons the parameter v is usually assumed to be independent of the amino acid residue, temperature and the presence of denaturants or helix inducing additives. The assumption of temperature independence implies that v has no enthalpic dependence and is entirely mirrored by the entropic factors implicit in the Ramachandran diagram. By contrast, the w parameter is residue, medium, and temperature dependent. As frequently used, the LR algorithm models the temperature dependent helix-coil transition entirely as the temperature dependence of w.

The assigned values of v and w will determine the population of helical conformers that determine the properties of the peptide in solution. If the helical propensities are only slightly >1 helical conformers of sufficient length are stabilized but

it will take a large number of residues to overcome the initiation parameter v^2; moreover, the small energy difference between similar length helices will create a manifold of entropically stabilized helices with lengths (k) from 3 to n. As the value of w is increased above 2, the totally helical conformer will become dominant. Since w, but not v, is temperature dependent, melting of the helical structure reflects a temperature dependent decrease in w that results in a less stable helical manifold that is dominated by shorter helices.

Are all 2^n conformations of a LR model energetically relevant? The potentially large numbers of helical conformers in solution are in fact dominated by molecules that have a single helical region. When comparing the major and minor populated conformations in small to medium sized peptides, the small value of v creates $(n - 1)(n - 2)/2$ major helical conformations, containing a single uninterrupted helical region, that do not pay the initiation price of v^4 to initiate two helices.[34] In addition the value of v reduces the random coil manifold to $[1 + n + n(n - 1)/2]$, dominant non-helical states, representing the all c conformation, and conformations with one h residue and two h residues. Conformations with more than three isolated h or hh designators have v^3, v^4 or higher terms and hence a minor population. Although the LR algorithm was originally used to model very long polymers where it was necessary to include sequences that had multiple helices broken by a few c residues, the simplification illustrated in Table 3 still represents 97% of the numerical weights within the helical manifold and will be used in Chapter 2 for data analysis. The large number of major states allows one to consider the effects of entropy on the manifold of helical conformations, if w is not large.

Length n	Total # of Conformations	# of Maj. Helical Conformations	# of Maj. Nonhelical Conformations
3	8	1	7
5	32	6	16
10	1,024	36	56
15	32,768	91	121
20	1,048,576	171	211

Table 3: Number of major conformations in the helix-coil manifold using the LR assumption and the value of v. The major helical conformations have a single helical region and no isolated h sites. The non-helical conformations consist of the all c conformation and the single and double h terms.

ANALYTICAL METHODS

How are the values of v and w assigned from experimental helicity data? For homo-oligopeptides that can be prepared in a wide range of lengths, analysis of melting transitions can be used to assign best fit values for w and v. Studies of this type are discussed in the next section. Other methods are needed for studies of short to medium sized peptides. Two experimental helicity measurements correlate with fractional helicity (FH), a residue property, which is easily calculated from a LR state sum using Equation 3. In all LR analysis, the mole fractions of helical conformations of distinct lengths k are calculated, and the length weighted sum of these from 3 to n provides the fractional helicity, the total number of helical α-carbons relative to the total number of α-carbons in solution. This section provides an introduction to these analytical techniques that will be followed by a detailed discussion within the chapter in which the technique is used.

$$FH = \sum_{k=3}^{n} \frac{k}{n} \chi_{\alpha_{C_i,k}}$$

Equation 3: Fractional helicity (FH) is calculated as the length weighted sum of the mole fractions of helical conformations with distinct lengths k from 3 to n. FH is easily calculated from an LR state sum. FH is the fraction of potentially helical α-carbons in the peptide that are actually helical. Its range is 0 to 1.

The most widely used experimental tool for FH assignment is circular dichroism[36] (CD), which is also the literature standard for protein and peptide structural characterization. Measured molar ellipticities $[\theta]$ play the role of extinction coefficients of UV spectroscopy and like them follow a Beer's Law concentration dependence, Equation 4. The primary CD chromophore is defined by the electronic transitions of the pair of amides that flank a residue α-carbon.[37] The signs and intensities of experimental $[\theta]_{obs}$ values depend on the corresponding pair of ϕ, ψ angles. At 222 nm, these are weak for extended peptide conformations, but exceptionally strong for helices. For a helical peptide of length n, the molar ellipticity is found experimentally to approximate $n \times [\theta]_{n,222}$ where $[\theta]_{n,222}$ is the molar ellipticity calculated on a perresidue basis.[38] In order to use Equation 5 to assign a FH value from a measured value of $[\theta]_{n,222,obs}$, $[\theta]_{n,222}$ must be known. As detailed later in this chapter and in Chapter 2 this assignment has

[36] Yang, J. T.; Wu, C. C.; Martinez, H. M.; "Calculation of Protein Conformation from Circular Dichroism" _Meth. Enzymol._ **1996**, _130_, 208-257; Rohl. C. A.; Baldwin, R. L. " Deciphering Rules of Helix Stabilities in Peptides" _Meth. Enzymol_ **1998**, _285_, 1-26.

[37] Manning, M. C.; Woody, R. W. "Theoretical CD Studies of Polypeptides Helices: Examination of Important Electronic and Geometric Factors" _Biopolymers_ **1991**, _31_, 569-586.

[38] The symbol $[\theta]$ is used in the literature interchangeably for both molar ellipticities (deg cm^2 $dmol^{-1}$) and perresidue molar ellipticities (deg cm^2 $dmol^{-1}$ res^{-1}) often without the unit identifier res^{-1}.

proven to be controversial. An experiment that opens up the possibility of a rigorous

assignment is described at the end of Chapter 3.

$$[\theta] = \frac{\theta}{c \; l} \qquad [\theta]_n = \frac{\theta}{c \; l \; n}$$

Equation 4: Molar ellipticity [θ] is proportional to the difference in molar extinction coefficients between left and right circularly polarized light, reported in degrees θ, divided by the peptide concentration and cell path length (deg cm^2 dmol^{-1}). Perresidue molar ellipticity [θ]$_n$ also corrects for the number of residues, n, within the helical region (deg cm^2 dmol^{-1} res^{-1}).

$$\mathrm{FH}_{helix} = \frac{[\theta]_{n,222,obs}}{[\theta]_{n,222}} \qquad [\theta]_{n,222} = [\theta]_{\infty,222}\left(1 - \frac{x}{n}\right) \qquad 2 < x < 5$$

Equation 5: CD derived experimental determination of fractional helicity as the observed perresidue molar ellipticity [θ]$_{n,222,obs}$ divided by the limiting perresidue molar ellipticity [θ]$_{n,222}$ for length n at 222 nm. The limiting perresidue molar ellipticity is calculated from a value of the perresidue molar ellipticity for a helix of infinite length [θ]$_{\infty,222}$ with moderate length correction (1 - x/n). Experimental fits for x have ranged from 2 < x < 5.[39]

The helicity of a peptide can also be determined by a quantitative study of

hydrogen exchange.[40] By dissolving a peptide in D$_2$O, the rate of exchange between a

site specific backbone amide NH $\rightarrow$ ND can be monitored by NMR. As illustrated in

Figure 11, the rate will be influenced by hydrogen bonding within the helical structure,

---

[39] Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W., Kallenbach, N. R. "The Helix-Coil Transition in Heterogeneous Peptides with Specific Side-Chain Interactions: Theory and Comparison with CD Spectral Data" *Biopolymers* **1991**, *31*, 1605-1614.

[40] A) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. "Protein Stability Parameters Measured by Hydrogen Exchange" Proteins: Struct., Funct., Genet. **1994**, *20*, B) 4-14; Rohl, C. A.; Baldwin, R. L. "Exchange Kinetics of Individual Amide Protons in ^{15}N-Labeled Helical Peptides Measured by Isotope-Edited NMR" *Biochemistry* **1994**, *33*, 7760-7767.

and is exceptionally slow for an NH that is fully hydrogen-bonded. The ratio of the experimentally observed exchange rate constant for a particular helical amide NH (k_{obs}) and the exchange rate constant for the same residue that is known to be fully solvent exposed (k_{se}) is defined as the protection factor (PF_i) at amide NH_i (Equation 6). The protection factor is by definition inversely proportional to the mole fraction of conformations that do not have hydrogen bonding at site i ($\chi_{NHi,se}$). As shown in Equation 6, one can also solve for the mole fractions of conformations that have helical hydrogen bonding at site i ($\chi_{NHi,hb}$). The subtle difference between $\chi_{NHi,hb}$ and $\chi_{\alpha Ci, H}$ will be devolved in Chapter 3. Both types of mole fractions can be calculated using an LR analysis.
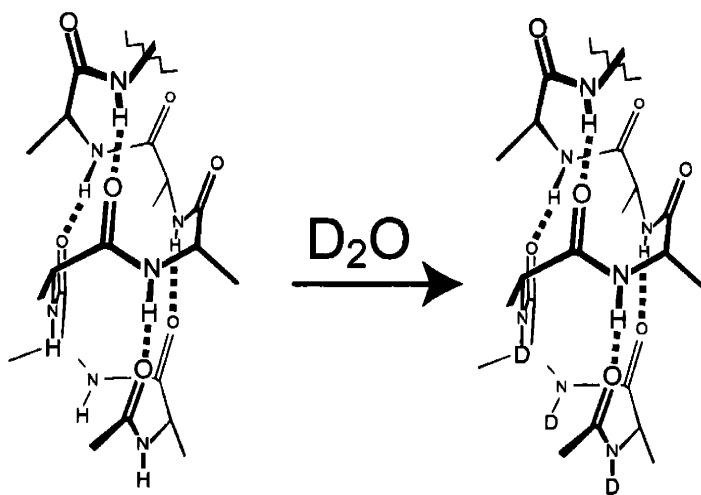
Figure 11: Structural representation of hydrogen exchange, showing the rapid exchange of non-hydrogen bonded amides and the slow exchange of amide NH's hydrogen bonded within the helix. A fully helical conformation that is assumed by a peptide of *n* residues that is N- and C-capped by amides is expected to have 3 solvent exposed NHs at its N-terminus and *n* - 2 hydrogen bonded amide NHs; one of these is formed by the C-terminal residue or NH_2 cap.

$$PF_i = \frac{k_{se}}{k_{obs}} = \frac{[NH_{i,T}]}{[NH_{i,se}]} = \frac{1}{\chi_{NHi,se}} = \frac{1}{1-(\chi_{NHi,hb})}$$

Equation 6: A protection factor (PF) is defined as the ratio of the experimental rate constants for the partially helical amide NH at site *i* amide (k_{obs}) and a model non-helical amide NH that is free to exchange with solvent (k_{se}). PF_i is equal to the inverse of the mole fraction of amide NH that are solvent exposed ($\chi_{NHi, se}$) or the inverse of the quantity unity minus the mole fraction that does have α-helical hydrogen bonding ($\chi_{NHi, hb}$).

An LR analysis can produce an overall fractional helicity, Equation 3, including all conformers with greater than three helical residues or individual site helicities $\chi_{\alpha Ci, H}$. For helical manifold that contains helical conformers of many different lengths, what does the LR algorithm predict for $\chi_{\alpha Ci, H}$? Figure 12 totals the helical residues at each site. The middle of the helical peptide to exhibits a larger site FH value than those at the ends; this is commonly referred to as fraying. The degree of fraying is determined by the value of *w*, a larger value will stabilize the completely helical conformation over the shorter conformations with a reduction in end fraying. The variation of site helicity as a function of *w* is illustrated in Figure 13. The physical properties of an LR manifold will reflect the weighted average conformation; in effect fraying causes this single hypothetical conformation to look as if it has less helical structure at the ends. This point is most easily seen by considering site helicity at each α-carbon ($\chi_{\alpha Ci, H}$) as seen in Figure 12 and Figure 13.

There is no rigorously established experimental measure of $\chi_{\alpha Ci, H}$, but as discussed in detail in Chapter 3, PF_i values provide closely related helicity information. FH derived from a single CD ellipticity represents a single average over all of the residue sites. Hydrogen exchange measures the ($\chi_{NHi,hb}$) at a specific residue sites, providing a

more incisive characterization of the distribution of helical lengths in the helical
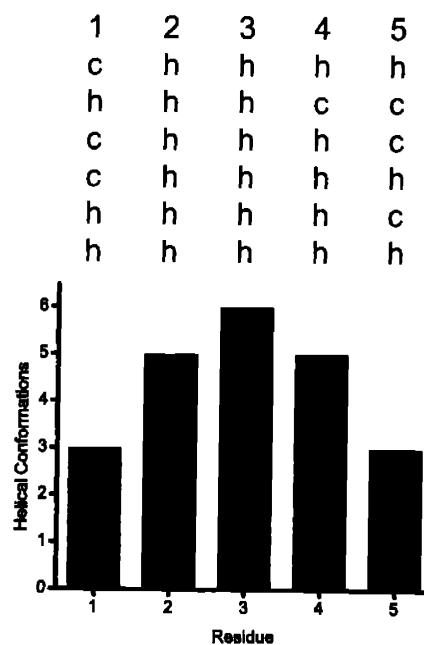
manifold.

Figure 12: The six major helical conformations of and N- and C-capped pentapeptide are represented above by the h,c sequences. Weighting each equally ($w = 1$) and counting over each vertical column yields the site helicity shown in the bar graph below. The resulting site helicities reflect the decrease in the average helicity at the ends of peptide.
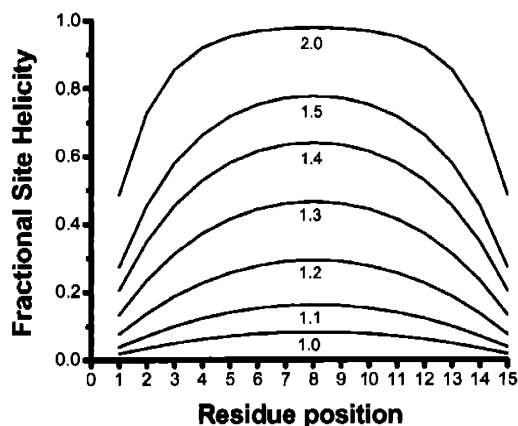
Figure 13: Site helicities are calculated for an N- and C- capped peptide ($n = 15$) by a LR algorithm with $v = 0.048$. Longer helices are stabilized as the value of w becomes larger increasing the average helicity for central residues more than those at the ends. Implausible large w values are required to eliminate N- and C-terminal fraying. FH of terminal residues is 0.90 for $w = 10$, 0.97 for $w = 30$ and 0.98 for $w = 50$.

Homopolymers as Helical Models

The use of peptide based models for the analysis of helical stability predates the protein mutant experiments described in the first part of this chapter. The rediscovery of α-amino-N-carboxyanhydrides in 1947 allowed the synthetic preparation of oligopeptides[41] that spanned a large molecular weight range (20-2000 residues) with a characterized length distribution. Most of these homo-polymers of the natural amino acids were found to be insoluble in aqueous solutions, but some were soluble in carefully chosen organic solvents and solvent mixtures. Experimental techniques were used to study conformations of these polymers in solution as a function of temperature, solvent,

---

[41]Woodward, R. B.; Schramm, C. H "Synthesis of Protein Analogs" *J. Am. Chem. Soc.* **1947**, *69*, 1151-1152; Frankel, M.; Katchalski, E. "Poly-condensation of α-Amino Acid Esters II. Poly-condenstaion of Alanine Ethyl Ester" *J. Am. Chem. Soc* **1942**, *64*, 2268-2271; Leuchs, H. *Chem. Ber.* "The Anhydrides of α-Amino-N-Carboxylic and α-Amino Acids' **1908**, *41*, 1721-1726.

and polymer length and composition.   Poly-γ-benzylglutamate was an extensively

studied example.   In aggressively hydrogen-bonding solvents, this polymer had

characteristics of a random coil, implying dominance of conformations typical of the

upper left quadrant of the Ramachandran diagram.   At low temperatures, in less

aggressive hydrogen-bonding solvents like formamide, the physical properties of the

solution reflected a dramatic conformational change consistent with assumption of a

compact rod-like shape by the polymer.   Using statistical polymer models, the calculated

width of the rod was found to match the model for the Pauling helix[42].   Moreover, in the

solid state by electron microscopy rods were observed and characterized as having the

Pauling helical dimensions.   The identification of helical structure of these polymers

provided the first non-protein models for the formation of helical secondary structure,

and the dependence of their helicities on length and temperature provided the first test of

the Zimm-Bragg[43] and Lifson-Roig[21] helicity models, which had been constructed for this

purpose.

The equilibrium transition from the oligopeptide's helical state to the random

state was studied by analytical methods such as optical rotation and solution viscosity.

Studies on poly-γ-benzylglutamate were initially confined to organic solvents, but later

extensions to aqueous solvents required water soluble oligopeptides that were obtained

from polymers of natural amino acids with charged side chains.

---

[42] Tanford, C. *Physical Chemistry of Macromolecules*; Wiley & Sons: New York, **1967**.
[43] Zimm, B. H.; Bragg, J. K. "Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains" *J. Chem. Phys.* **1959**, *31*, 526-535.

Polylysines[44] and polyglutatmates[45] were studied in aqueous media and found to be helical, showing similar length and temperature properties. Neither polymer was helical when fully charged, and each tended to aggregate when uncharged, but a dramatic helix-coil transition was observed at low temperatures in a narrow pH range, where the polymer is only partially charged. This indicated that within the fully charged polymer charge-charge repulsion selectively destabilized the helix with respect to the more extended conformations typified by those of the upper left region of the Ramachandran diagram.

A very strong length dependence was observed for the helix-coil transitions. Large oligopeptide polymers (>>100 residues) exhibited high helicities and melted at high temperatures over short temperatures ranges. The medium sized polymers (30-80 residues) melted at lower temperatures and over a larger range, and helix-coil transitions were incomplete at both temperature extremes. Without assuming the independent contribution of each residue to the thermodynamic stability of long helical conformations it is difficult to understand the striking experimental differences between helical oligopeptides of average lengths of ca 50 residues and those with lengths at >100, but the LR model provides a simple explanation. The narrow temperature range, the completeness of the transition, and the increased mean melting temperature seen for lengths >100 are consistent with a proportionate length dependence for the enthalpy of the helix-coil transition. This length dependence is implied by the LR model in which the

---

[44] Epand, R. F.; Scheraga, H. A. "The Helix-Coil Transition of Poly-L-Lysine in Methanol-Water Solvent Mixtures" *Biopolymers* 1968, 6, 1383-1386.

[45] Zimm, B. H.; Doty, P.; Iso, K. "Determination of the Parameters for Helix Formation in Poly-γ-L-Glutamate" Proc. Natl. Acad. Sci. U.S.A. **1959**, *45*, 1601-1607.

helix enthalpy is the sum of a fixed enthalpy term embodied in the perresidue w value.

The two-state Lifson and Roig algorithms provided a good fit to the experimental data for

the length dependent melting of these homopolymers, yielding values of v consistent with

the Ramachandran diagram and w values slightly above 1.0.
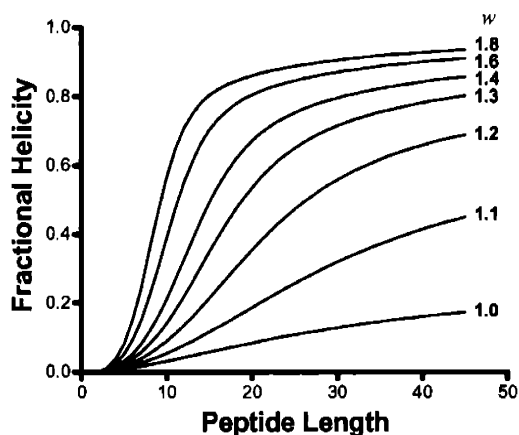
Figure 14: LR prediction of FH as function of length (n) and helical propensity varying (w) [v = 0.048]. The low values of w for short homopolymers restrict the intrinsic ability to form helices.

Figure 14 illustrates how the LR algorithm models FH as a function of w and n.

For the w value of 1.3, short peptides <10 residues are not helical and medium length

peptides dramatically increase in FH with length. Yet even peptides between 40 and 50

residues do not become fully helical, reflecting the entropic stabilization of shorter

helices within the experimentally determined values of w for homopolymers.

HELICAL HOSTS, ALGORITHMS & ANALYTICAL METHODS

It is important to understand the fundamental difference between the temperature

dependent equilibria of helical conformational states of these oligopeptides and the

typical native states of helical globular proteins. Whereas the proteins approximate a

two-state transition from a single folded conformation to a unstructured state that occurs

abruptly over a short temperature range, helical polymers of a biologically relevant length

(15-50 residues) melt over a broad range, usually without ever becoming completely

helical at low temperature.
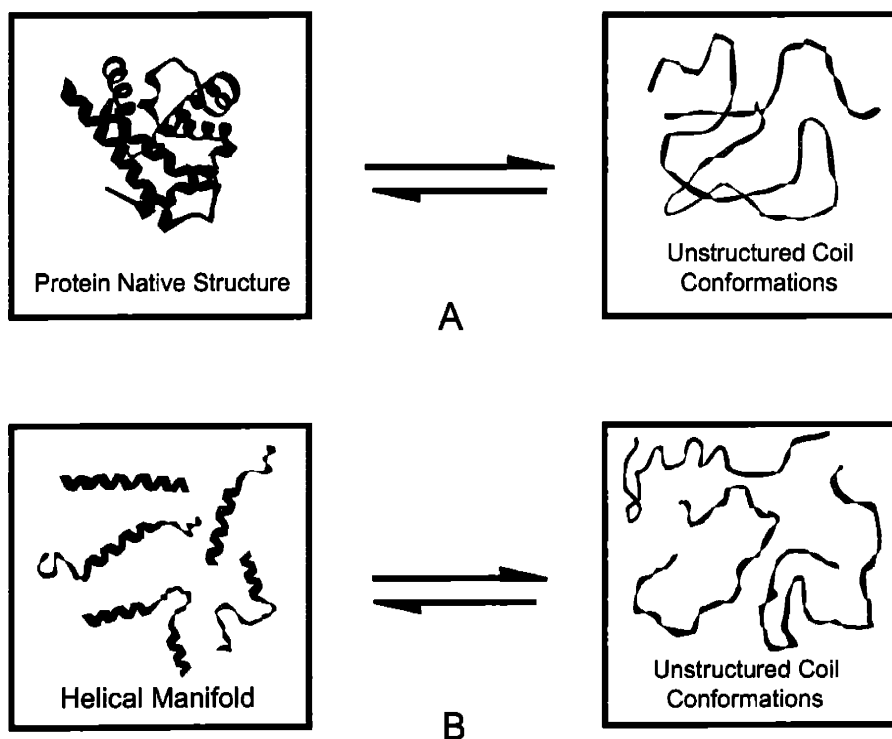
Figure 15: A) Approximate two-state protein unfolding from a single native structure to an unfolded manifold, energetically considered one state. B) Helical unfolding from a manifold of helical states to a manifold of unstructured states.

As mentioned before, the conformational analysis of a molecule requires the

following: 1) A molecule can be assigned to a series of defined conformational states. 2)

Each conformational state is structurally characterized. 3) A temperature range is defined

where the conformational manifold is at equilibrium. 4) A spectroscopic method or

model can be applied to the manifold to quantify the amount of each conformation,

expressed as condition dependent mole fractions χ_i. Only two of these conditions are met

for oligopeptide helices, the helical conformation is well characterized, and equilibrium between the helical conformations has been established.[46] Because no experimental method is able to determine the relative abundance of each helical conformation, the average molecular and site properties, observed by CD and hydrogen exchange, must be modeled using LR to quantify the population of each conformational state. Whereas the CD ellipticity for a folded protein represents to a good approximation the concentration, a molecular property, of a single conformation, the CD ellipticity of a helical peptide represents, fractional helicity, a site property, the average over all of the helical residues.

$$\text{A) } \chi_{native\ protein} = \frac{[\theta]_{observed} - [\theta]_{unstructured}}{[\theta]_{native} - [\theta]_{unstructured}} \qquad \text{B) } FH_{helix} = \frac{[\theta]_{observed} - [\theta]_{unstructured}}{[\theta]_{helical} - [\theta]_{unstructured}} = \sum_{k=3}^{n} \frac{k}{n} \chi_k$$

Equation 7: Whereas the molecular amount of native, folded protein is described by equation A, fractional helicity is describes the average helical properties of a manifold of helical conformations with varying lengths k, equation B.

In summary, the discovery of water-soluble helical homopolymers, such as polyglutamate and polylysine, provided synthetic hosts that could be studied in a solution like helical proteins. Experiments established the key variables for helix formation: 1) amino acid composition, 2) length, 3) temperature and 4) solvent. Chemical insight into how these polymers folded as measured by their physical properties was provided by LR and related models. These homopolymers also provided insight into the effects of charge on helices, as fully charged homopolymers were found to have extended conformations.

---

[46] Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. "Kinetics of Amide Proton Exchange in Helical Peptides of Varying Chain Lengths. Interpretation by the Lifson-Roig Equation" *Biochemistry* **1992**, *31*, 1263-1299.

Since the _w_ values calculated from these partially charged peptide molecules represent an average of the charged and uncharged residues, the complexity of their pH dependence excluded them as ideal hosts. Efforts to create new uncharged water-soluble hosts and determine the _w_ values for all 20 of the natural amino acids will be summarized in the next section. Although many investigators have made important contributions, it is convenient to focus on the work of the Scheraga and Baldwin groups for background information used within this thesis.

HISTORY OF HELICAL HOSTS: SCHERAGA

The research to create water soluble helical hosts and determine helical propensities generated controversy as different research groups compared the data and helicity parameters calculated from their own hosts with those of others in the field. In addition to summarizing previous efforts to quantify _w_ values, this section introduces key features of the existing controversies regarding polyalanines that are addressed in subsequent Chapters. This thesis provides a major experimental foundation for a new scheme by which these controversies can be resolved.

In an effort to assign the absolute values of helical propensities, starting in the 1970's and continuing into the 1990's, Scheraga and coworkers studied carefully designed uncharged oligopeptides. These are formed by polymerizing a mixture of host poly-γ-benzylglutamate monomer and a smaller amount of guest monomer and further modified by transforming the resulting poly-benzyl ester to glutamine-type amides by reaction with 4-aminobutanol. These copolymers are water-soluble and form helices that melt between 0 and 60 °C, allowing accurate experimental characterization of the helix-

coil transition. Rigorous review of polymer modeling showed that random copolymers could be used to determine w values, independent of abundance, length, or distribution of the guest regions.[47] After a thermodynamic melting analysis, as observed by optical rotatory dispersion (ORD), of the poly-(hydroxybutylglutamine) host,[48,49] host-guest melting studies varying the host/guest residue ratio were analyzed to assign temperature dependent w values for all 20 of the natural amino acids.[50] At low temperature, 2 °C, the range of propensities varied from those of helix breakers, 0.08 for Pro and 0.5 for Gly to those characteristic of helix formers, 1.2 for Met and Ile. Alanine, the simplest of the chiral helix forming amino acids, was assigned a w value of 1.1 that only modestly varied with temperature. Half of the 20 natural amino acids were found to be helix formers, $w >$ 1 and half were helix breakers, $w < 1$.[51] The magnitudes of these intrinsic w values predict that heteropeptides containing fewer than 30 residues would not form detectable helices.[52] In early, parallel work by Scheraga and Epand[53], protein fragments with 24, 57 and 79 residues, cleaved from the highly helical protein myoglobin, were studied

[47] Dreele, P. H.; Poland, D.; Scheraga, H. A. "Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. I. Properties of Copolymers and Approximate Theories" Macromolecules 1971, 4, 396-407.

[48] Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. "Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. II. Characterization of the Host Polymers and Application of the Host-Guest Technique to Random Poly(hydroxypropylglutamine-co-hydroxybutylglutamine)" Macromolecules 1971, 4, 408-407424.

[49] ORD is a UV technique similar to CD.

[50] Scheraga published σ and s values consistent with the Zimm Bragg model, the values quotes have been converted to LR w values using the following relation ships $s = w/(1+v)$ and $\sigma = v^2/(1+v)^4$ Qian, H.; Schellman, J. A. "Helix-Coil Theories: A Comparative Study for Finite Length Polypeptides" J. Phys. Chem. 1992, 96, 3987-3998.

[51] Wojcik, J.; Altmann, K. H.; Scheraga, H. A. "Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. XXIV. Half-Cystine Parameters from Random Poly(Hydroxybutylglutamine-co-S-Methylthio-L-Cysteine" Biopolymers 1990, 30, 121-134.

[52] Thirty helix forming amino acids with a propensity of 1.2 would be needed to get 20% of the residues helical $v^2w^{30}/(1+30v+ v^2w^{30})=0.6/(2.5+0.6)=0.2$ using a two state approximation.

[53] Epand, E. M.; Scheraga, H. A. "The Influence of Long-Range Interactions on the Structure of Myoglobin" Biochemistry 1968, 7, 2864-2872.

using several spectroscopic methods. Helices were not detected in water, consistent with the oligopeptides-derived w values.

Scheraga's work was intended to assign the intrinsic helical propensities of the amino acids, which in the absence of context or site dependent[54] helix stabilizing or destabilizing interactions should model peptide helicity properties accurately. The assignment of intrinsic propensities was the first step toward the goal of predicting secondary structure within heteropeptides. However other interactions are likely to occur, as noted by Scheraga: "In order to use the data to predict the probability of occurrence of an α-helix in a specific sequence [protein] at any temperature it is necessary to modify these intrinsic values by incorporating specific medium and long range interactions."[13] These could include charge, salt bridge and hydrophobic interactions that are sequence specific and would not be detectable using the nonsequence specific copolymer methods.

In the host planning stages for their helicity studies, the Scheraga group rejected all natural amino acids as host choices, due to the insolubility or melting temperature range. The structural versatility of the unnatural hydroxyalkyl side chain of the Glu-derived oligopeptides that were selected allowed for the essential tuning of the solubility and melting properties. Although these poly(hydroxybutylglutamines) hosts may classify the system as unnatural, with a larger number of hydrophobic methylenes that might interact with guest residues at $(i, i\pm3)$, $(i, i\pm4)$ positions, Scheraga argued that the resulting w values are supported in polymer theory under the assumption that the

---

[54] Context effects arise from interactions with specific neighboring residues. Site effects arise from the position within the helical conformation.

stabilizing or destabilizing interactions are small. The validity of this assumption was consistent with helicity studies of a solubilized polyalanine in Scheraga's coblock polymer study (Lys_m-Ala_n-Lys_m, n = 10- 1000 and m = 80 – 210)[55] where large blocks of polyalanine are unable to interact with the solubilizing lysine residues. This study assigned w values for alanine that are consistent with the host-guest studies.[56]

HISTORY OF HELICAL HOSTS: BALDWIN

An exception to the nonhelical myoglobin fragments was reported in 1969. Using CD, Brown and Klee found heteropeptides fragments of the S-peptide of ribonuclease A with 13, 15 and 20 residues[57] were weakly helical at low temperatures. A series of publications from the Baldwin group attempted to determine the stabilizing factors that could explain this apparent anomaly. Using solid phase peptide synthesis, Baldwin and coworkers investigated the cause of this unsuspected stabilization by mutating singly and multiply charged residues and monitoring the pH-dependent changes in helicity. This process identified stabilizing salt bridges[58] and the role of N- and C- caps as helix stabilizers. These N- and C- cap variables are unique features of small peptides and are not present in the Scheraga copolymers. The Baldwin S-peptide study was the first demonstration of an important finding that changes in cap identity have large effects on the overall helicity of short and medium sized peptides.

---

[55] Ingwall, R. T.; Scheraga, H. A. "Conformational Studies of Poly-L-Alanine in Water" *Biopolymers* **1968**, *6*, 331-368.

[56] Scheraga, H. A.; Vila, J. A.; Ripoll, D. R. "Helix-Coil Transitions Re-Visited" *Biophys. Chem.* **2002**, *101-102*, 255-265.

[57] Brown, J. E.; Klee, W. A. "Conformational Studies of a Series of Overlapping Peptides from Ribonuclease and Their Relationship to Protein Structure" *Biochemistry* **1969**, *8*, 2876-2879.

[58] Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin. R. L. "The Glu 2⁻---Arg 10⁺ Side-Chain Interaction on the C-peptide Helix of Ribonuclease" *BioPhys. Chem.* **1990**, 107-119.

To clarify the role of salt bridges in helix stabilization, Baldwin and Marqusee

studied a series of short polyalanines containing salt bridges of lysine and glutamate

spaced at $(i, i - 3)$ and $(i, i - 4)$ separations. In 1989, they discovered that the simpler N-

and C- capped peptides sequences that lacked salt bridges, $Ac-(A_4K)_m-NH_2$

$(m = 2 - 5)$,[59] are much more helical than the S-peptide and its analogs. Some of this

increased helicity can be attributed to stabilizing N- and C- terminal caps. For example,

at pH 9.6, if the deprotonated N-terminal amino acid $(H_2N-CHR-CO-)$ is replaced by an

N-acetyl cap $(CH_3-CON-CHR-CO-)$ for a series of 19 residue peptides, the helicity

increased by 70-120%. Large capping effects are also seen for positive C-cap residues

and negatively charged N-caps, owing to favorable helix stabilizing dipole-charge

interactions at the ends of helices. Amino acids with uncharged polar side chains, such

as Ser and Asn, also act as strongly helix stabilizing N-caps, although their w values are

small.[60] The Baldwin group proposed that the w value of an amino acid alone does not

define the helicity properties of an amino acid; in addition to the w value N- and C-

capping parameters were also assigned for all of the natural amino acids using a database

of peptides.

Under the assumption that the CD derived $[\theta]_{n,222}$ provides a precise[61] FH

quantitation tool, the Baldwin-Marqusee motif was used as the host system

$(-AAKA-X-AAKA-)$ for a host-guest study using a LR analysis to assign helical

[59] Marqusee, S.; Robbins, V. H.; Baldwin, R. L. "Unusually Stable Helix Formation in Short Alanine-Based Peptides" _Proc. Natl. Acad. Sci. U.S.A._ **1989**, _86_, 5286-5290.

[60] Doig, A. J.; Baldwin, R. L. "N- and C-Capping Preferences for All 20 Amino Acids in α-Helical Peptides" _Protein Sci._ **1995**, _4_, 1325-1336.

[61] Baldwin increased the value of $[\theta]_\infty$ by 26% over the value derived from the protein database. Chakrabartty, A.; Schellman, J. A.; Baldwin, R. L. "Large Differences in the Helix Propensities of Alanine and Glycine" _Nature_ **1991**, _351_, 586-588.

propensities along with N- and C-capping parameters at 2 °C for all of the amino acids.[10]

Baldwin assumed that the helical nature of the host peptide could be primarily attributed to the large percentage of alanine residues and that the Lys residues (20%), introduced to insure adequate solubility and freedom from aggregation, in fact destabilized the helix. This conclusion was supported by the results of introducing a guest X = Lys in the (-AAKA-X-AAKA-) sequence which substantially reduced the helicity of the peptide. Accordingly Lys was assigned as a helix breaker with a low w value. This LR host-guest study yielded an exceptionally large helical propensity for alanine of 1.6, significantly larger than all of the other amino acids. This led Baldwin to conclude that Scheraga's alanine value of 1.1 was wrong. He also cast doubt on all Scheraga w values, citing a possible bias attributable to stabilization by the hydroxybutylglutamine side chains of the Scheraga host. Suggestive experimental evidence in support of this conclusion was provided.[62]

In practical terms, alanine w values of 1.1 and 1.6 result in strikingly different helical predictions. A LR modeling shows that a polyalanine peptide with 11 residues would be 20% helical if w = 1.6, but a peptide with 30 residues would be required to yield the same helicity if the Scheraga value of 1.1 is correct. The effect of a small difference in w escalates with length; the Scheraga case weights the all helical conformation 1.1^{30}= 17 whereas in the Baldwin case a corresponding weight is 1.6^{30} = 1.3×10^6. The large disparity between weights and lengths of peptides that yield helices calls into dispute the assumptions underlying both assignments. Does this disparity

---

[62] Padmanabhan, S.; York, E. J.; Gera, L.; Stewart, J. M.; Baldwin, R. L. "Helix-Forming Tendencies of Amino Acids is Short (Hydroxybutyl)-L-glutamine Peptides: An Evaluation of the Contradictory Results from Host-Guest Studies and Short Alanine-Based Peptides" *Biochemistry*, **1994**, *33*, 8604-8609.

solely lie within the context of the hydroxybutylglutamine host, or are there unknown energetic effects within the Ac-(Ala$_4$Lys)$_n$-NH$_2$ context. Is the small value of w_{Lys} correctly assigned or should helicity be attributed in part to a stabilization by Lys? Could the long Lys side chain (-CH$_2$-CH$_2$-CH$_2$-CH$_2$-NH$_3^+$) make stabilizing helical contacts? Scheraga has calculated that (i - 3) and (i - 4) stabilizing contacts are possible among the many conformational[63] states of the lysine side chain within a helical conformation.[64] If these conformations are populated, the Baldwin-Marqusee host, like the Scheraga host, could contain uncharacterized context or site effects.[54] Baldwin's analysis assumes that w_{Lys} is site independent, despite evidence that the position of an amino acids from the N- and C- change the helicity of the peptide, explained by charge interactions with the helical dipole.[65] Baldwin also assumes that the multiple lysines within four residues (-AAKA-X-AAKA-), X = Lys is not influenced by the charge-charge repulsion observed for polylysine. This interaction could selectively destabilize the helix. Dismissing Scheraga's work without a careful analysis of the Baldwin hosts is premature. The uncertainty of the stabilizing mechanisms available to a Lys side chain makes the isolation of intrinsic effects from context and site effects difficult and of central importance when designing and characterizing helical hosts that contain lysine or similar residues.

---

[63] The conformational states of the side chain are substates of the h, c backbone conformations and are not addressed by the LR algorithm.

[64] Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. "Influence of Lysine Content and pH on the Stability of Alanine-Based Copolypeptides" *Biopolymers* **2001**, *58*, 235-246 and references therein.

[65] Armstrong, K. M.; Baldwin, R. L. "Charged Histidine Affects Alpha-Helix Stability at All Positions in the Helix by Interacting with the Backbone Charges" *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11337-40.

The methodology used to characterize helicity of the Baldwin-Marqusee peptides also poses interpretational problems. Both CD and hydrogen exchange have been used to prove that the Baldwin-Marqusee hosts are very helical. Published PF data confirm the overall helicity but do not address the relative stabilizing or destabilizing effects of alanine and lysine. Moreover, recent work by the Kemp group has called into question the values of $[\theta]_{\infty,222}$ and $[\theta]_{n,222}$ used in the CD-derived helicities from which the Baldwin group assigned their w values.

HISTORY OF HELICAL HOSTS: KEMP

In 1991, using a new approach with a molecule that provided a helical initiation site, the Kemp group assigned a new value for the helical propensity of alanine. Culminating his efforts to create exceptionally helix stabilizing N-caps, Tim Curran synthesized a molecule Ac-Hel pictured in Figure 16.[66] This constrained acetyl-prolyl-proline analogue was found to have three properties: 1) Helices are initiated when it is attached to the N-terminus of peptides through an amide bond. 2) The ^{1}H NMR properties of the slowly rotating acetyl group can be used to report the stability of the linked peptide helix. 3) The reporting feature can also be used to prove conformational communication along a peptide helix. The details of each of these properties will be covered in the next chapter of this thesis.

---

[66] Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd. J.G.; Muendel, C. "Studies of the N-Terminal Templates for α-Helix Formation. Synthesis and Conformational Analysis of (2S, 5S, 8S, 11S,)-1-Acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricylco[2.8.2.0^{4,8}]-tridecane (Ac-Hel$_1$-OH)" *J. Org. Chem.* **1991**, *56*, 6672-6682.
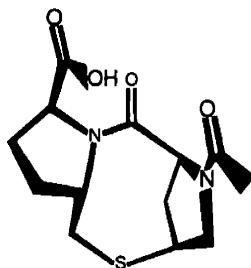
Figure 16: Ac-Hel-OH stabilizes helical peptides when amide linked to the N-terminus; ^{1}H NMR spectroscopy characterizes t/c, the relative abundances of the s-*cis*/s-*trans* rotamers of the acetyl function, which can be related to peptide helicity. The t/c ratio can be modeled as a function of the LR state sum.

Ac-Hel provided an opportunity to study short peptides that would not be helical without the influence of the strong cap. The first study focused on short polyalanine peptides with the intent of settling the dispute between Scheraga and Baldwin. The values of the t/c reporting function for the series Ac-Hel-A_n-OH and Ac-Hel-A_n-NH$_2$ ($n = 1$ - 6)[67] indicated that alanine is a weak helix former, with a temperature-independent w value of 1.0. This result is consistent with the value assigned by Scheraga.

The lengths of these series could not be extended, since six alanines is the limit for water solubility. Additional series were constructed by introducing internal Lys residues to solubilize longer polyalanine peptides extending to 12 residues. A large t/c database was constructed and the resulting analysis, which resulted in an excellent

---

[67] a) Kemp, D. S.; Boyd, J. G.; Muendel, C. C. " The Helical s Constant for Alanine in Water Derived from Template-Nucleated Helices" *Nature* **1991**, *352*, 451-454. b) Kemp, D. S.; Oslick, S. L.; Allen, T. J. "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. 3. Calculation of the Helical Propagation Constant *s* from the Template Stability Constants t/c for Ac-Hel$_1$-OH, *n* = 1-6" *J. Am. Chem. Soc.* **1996**, *118*, 4249-4255. b) Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. "For Short Alanine-Lysine Peptides the Helical Propensities of Lysine Residues (*s* Values) Are Strongly Temperature Dependent" *J. Am. Chem. Soc.* **1996**, *118*, 12234-12235.

correlation between the predicted and experimental t/c values, suggests that the low

helical propensity of Ala could be consistent throughout, if the Lys residue in a

polyalanine context is assigned a larger w value that is temperature and site dependent.

Examples of helical stabilization by Lys residues are well known in the literature,[68] and

the Kemp group has shown through an NOE NMR experiment that the side chain of the

Lys residue is in contact with the alanine helix barrel.[69]

A second approach was also explored by the Kemp group. Peptides in the data set

used by the Baldwin group to assign w values are characterized by only a small variation

in their ratio of Ala to Lys, making differentiation of the helical propensities of alanine

from those of lysine difficult. A series of inversely dependent w_{Ala} and w_{Lys} are expected

to give acceptable fits to such a data set. In 1998, Williams, Kather, and Kemp[70] created

a solubilized (-Ala$_{18}$-) core, by which the Ala/Lys ratio was varied from 18 to 3; analysis

of $[\theta]_{222}$ values yielded w values consistent with the earlier Kemp and Scheraga data.

The Kemp and Scheraga analyses of the propensity of alanine are generally

rejected by other investigators in the field. In 1999, major players published rebuttals

with rejection of these analyses, explicit in the titles of their papers. Baldwin et al.:

"Alanine is Helix-Stabilizing in both Template-Nucleated and Standard Peptide

[68] Esposito, G.; Dhanapal, B.; Dumy, P.; Varma, V.; Mutter, M.; Bodenhausen, G. "Lysine as Helix C-Capping Residue in a Synthetic Peptide" *Biopolymers* **1997**, *41*, 27-35.

[69] Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. "Template-Nucleated Alanine-Lysine Helices are Stabilized by Position-Dependent Interactions Between the Lysine Side Chain and the Helix Barrel" *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4025-4029.

[70] Williams, L.; Kather, K.; Kemp, D. S. "High Helicities of Lys-Containing, Ala-Rich Peptides are Primarily Attributable to a Large, Context-Dependent Lys Stabilization" *J. Am. Chem. Soc.* **1998**, *120*, 11033-11043.

Helices"[71] and Kallenbach et al.: "Alanine is an Intrinsic α-Helix Stabilizing Amino Acid."[72] In the Kemp data base consisting of Ac-Hel capped peptides less than 7 residues were considered, by other researchers, too short to be representative of longer helices, despite the presence of diagnostic α-helical NOE signatures. Assignments of w values for the longer peptides with internal Lys solubilizers were dismissed on the basis that w_{Lys} was treated as position dependent; furthermore it was suggested that the Lys side chain may influence the reporting function of the template rendering the t/c data un-interpretable.

The Baldwin and Kallenbach groups disputed conclusions drawn from the Ac-Hel and -Ala$_{18}$- Kemp hosts by introducing new data including the CD properties of short polyalanine cores N- and C-terminally capped and solubilized by lysine or ornithine residues. The Kallenbach peptide core (-O$_2$-A$_{13}$-O$_2$-) was similar to the Baldwin core (-K-A$_9$-K-). The relatively intense values of $[\theta]_{222,n}$ observed for these peptides were interpreted as reflecting the high intrinsic helicity of alanine itself. Plausible alternative explanations, such as large helix stabilization by the positively charged lysine side chain at the C-cap, or extension of the helical length beyond the Ala region and through the Lys residues were not considered. The results and interpretations of the Williams, Kather, Kemp paper were dismissed as resulting from initial aggregation problems despite presentation of data supporting lack of aggregation.

---

[71] Rohl, C. A.; Fiori, W.; Baldwin, R. L. "Alanine is Helix-Stabilizing in Both Template-Nucleated and Standard Peptide Helices" *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682-3687.
[72] Spek, E. J.; Olson, C. A.; Shi, Z.; Kallenbach, N. R. "Alanine is an Intrinsic α-Helix Stabilizing Amino Acid" *J. Am. Chem. Soc.* **1999**, *121*, 5571-5572.

For most bioscientists, the most important issues are centered on whether or not sufficient agreement exists between the published scales of absolute and relative helical propensities to predict helicities of regions within the primary sequences of natural proteins or as predictors for the construction of designer proteins. This question can be probed by three consistency criteria that are applied in Table 4 to the protein mutant derived relative w values of Blaber and Scholtz as well as the absolute w values of Scheraga and Baldwin. Also included are the t/c derived w values measured in the host AcHel-A_2-X-A_3K-NH_2 by Tom Allen of the Kemp group.[73] The ratio w_{ala}/w_{gly}, compares propensities of the stabilizing alanine with a strong helix breaker Gly,[74] this number defines the range of propensities. The large variation seen in Table 4, suggests that Baldwin has assigned extreme values to either alanine or glycine or to both. The exceptionally large w_{Ala} assigned by Baldwin is also apparent if the average $<w>$ for all of the amino acids is divided by the alanine value. Three of the four correlation coefficients establish acceptable consistency with the Scholtz dataset but the Scheraga values are clearly inconsistent. The excellent agreement seen between the data sets of Blaber, Scholtz, and Allen suggests that w values derived in alanine-rich host can correlate well with protein mutant data. In addition, the correlation between the results of Balber and Scholtz using protein host and Allen and Scholtz using a LR analysis for peptides, helps to validate the LR algorithm and suggest that amino acid helical propensities are similar in proteins and peptides.

---

[73] Allen, T. J. Studies of Template-Initiated α-Helix Propagation in Short Peptides. Ph.D. Dissertation, M.I.T. 1993.

[74] Excluding Pro, as secondary amine that is unable to form a hydrogen-bond in a peptide sequence.

Investigator	w_{ala}/w_{gly}	$<w>/w_{ala}$	CC with Scholtz
Baldwin	31	0.3	0.94
Scheraga	2	0.9	0.60
Kemp	3	0.6	0.96
Blaber	5	0.6	0.90
Scholtz	5	0.6	(1.00)

Table 4: Three consistency criteria for the *w* values obtained from the protein hosts of Blaber and Scholtz, Scheraga's copolymers and Kemp's AcHel polyalanines. The lack consistency for all three criteria suggest that context or site effects may have influences the measurement of intrinsic helical propensities.

NEW HELICAL HOSTS: OBJECTIVES & CONSTRAINTS

In 1999, faced with a series of fundamental, unresolved disputes about appropriate hosts and measurement techniques, the Kemp group set out to resolve these problems by devising a new host that would allow the study of the helical properties of polyalanines as soluble and unaggregated peptides that are experimentally demonstrated to be free of interactions with charged residues. The basic blueprint for a new study has the following features: 1) A well defined helical region in a host should allow variation of the amino acid composition and length. A polyalanine core region should be used since this is the simplest helix forming amino acid, and a polyalanine sequence is free from context uncertainties.

The goal is to create a helical peptide host that is significantly versatile to allow experimental characterization of the list of medium and structural variables in Figure 17. Water is the solvent of choice for initial studies because of its biological relevance, but the potential for further experiments with salts, denaturants and cosolvents had to be considered. A temperature range of 2° to 60 °C allows comparison of the new host's helical properties with previous studies.

The first aim of these studies remains the complete characterization of the polyalanine host peptide. Mutant w values cannot be assigned unless the absolute values of w for the host are known, since rigorous w_{Ala} assignment is of central importance to resolving controversy a thorough characterization of the host is essential. 2) Helicity measurement techniques must be applied with caution and an awareness of their limitations. In the past, most investigators have primarily relied on a single analytical method, namely CD. In order to prevent errors in a single analytical method from skewing the assignment of helical propensities, CD spectroscopy, hydrogen exchange and the Ac-Hel t/c ratio should be used together to probe the helical characteristics of the host and assign host w values. This methodology is the tenet of the Kemp group scheme.

3) A model that predicts how helix formation is governed by energetics. The LR model is used as it has become the established model and has been adapted to study short peptides. Although recent literature criticism regarding the LR weighting of the unstructured state for heteropeptides[75] has been raised, these are unlikely to be valid for polyalanine peptides. LR algorithms are the standard for assigning helicity parameters from CD, PF and t/c data. In addition, polyalanines are modeled using a single w value in standard LR algorithms, thereby restricting the number of free variables. Recently the Kemp group has expanded the scope and versatility of the LR algorithms.[35] For the first stages of polyalanine helicity studies, these LR models should be satisfactory.

---

[75] Pappa, R. V.; Srinvasan, R.; Rose, G. D. "The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding", *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12565-12570.

```
                    ┌─────────────────┐
                    │    Helicity     │
                    │    Variables    │
                    └─────────────────┘
           ┌──────────────┴──────────────┐
    ┌──────────────┐              ┌──────────────────┐
    │    Medium    │              │ Peptide Structure│
    │  Conditions  │              └──────────────────┘
    └──────────────┘
    ┌──────┴──────┐          ┌─────────────┼─────────────┐
┌──────────┐┌──────────┐┌──────────────┐┌──────────┐┌──────────────┐
│ Solvent &││Temperature││Helical Length││Amino Acid││ N- & C- Caps │
│ Additives││          ││              ││ Sequence ││              │
└──────────┘└──────────┘└──────────────┘└──────────┘└──────────────┘
```
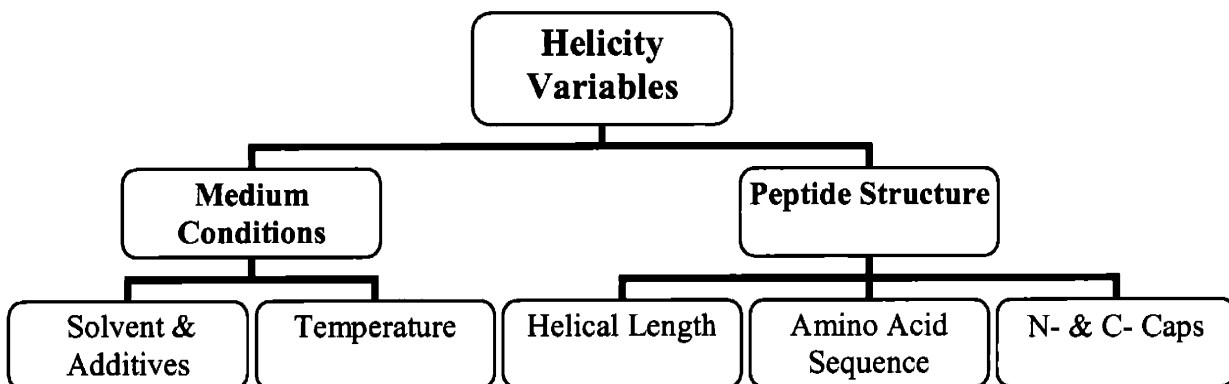
Figure 17: Major medium and structural variables known to influence the degree of helicity achieved by a peptide sequence. These variables can be changed to study causal relationships.

NEW HELICAL HOSTS: DEVELOPMENT

New helical hosts must meet a range of special conditions. The host must allow a large length variation within the helical polyalanine region in order to explore the previously established length dependence seen by Scheraga and Baldwin. The effects of the N- and C- terminal residues on the helical regions would also need to be investigated and the host should be designed to allow the introduction of helix stabilizing, neutral and destabilizing caps. The host amino acids sequence must play two distinct roles. It must solubilize a large range of polyalanines in water without significant aggregation, and it must shield the solubilized alanine region from any potential helix stabilizing or destabilizing effects of the solubilizers.

The solubilization problem is potentially solved by the construction of triblock peptide sequences in which polyalanine cores ($-Ala_n-$) are N- and C-capped by charged or strongly hydrophilic regions ($Cap_m-A_n-Cap_m$). As noted, above the Baldwin and

Kallenbach groups have recently provided short working examples in which Lys_m or

Orn_m regions are used as solubilizers. Demonstration of the scope of this strategy is

provided by earlier work by Scheraga and by Doty and Gratzer,[76] who demonstrated that

polymers with Lys_m or Glu_m regions of suitable length can achieve unaggregated

solubilization of alanine cores containing more than 100 residues. Given these

precedents, the solubilization role of charged end caps seemed answered.

During the host development phase, the Kemp group selected N- and C- regions

of lysine residues as solubilizers for the following reasons: their ease of synthesis, mass

spectrometric characterization, and the large pH window (0-9) in which the ionization

state, fully charged, of the tri-block peptide would be invariant. Baldwin and Kallenbach

reasoned that because the ends of the helix were frayed on average, that the presence of

the Lys residues did not effect or contribute to the helicity, hence the observed helicity

was attributed to solely alanine. Is this reasoning correct? Initial experiments were done

to test if the helical region was defined within the polyalanine region

The first experimental project presented in this thesis addresses this point and was

carried out in collaboration with Prof. Harald Schwalbe. The sequence Lys_3-A_{12}-Lys_3

with all of the alanine residues ^{15}N labeled was synthesized and structurally characterized

by NMR. Using NOE spectroscopy, a technique that is sensitive to the distance between

atoms, one can differentiate between amide NH's that are a part of the helix and those

that are not. Within the helix, the i to (i + 1), NH to NH distance is small in comparison

to extended conformations. A ^{15}N filtered NOESY experiment was used which only

---

[76] Gratzer, W. B.; Doty, P. "A Conformation Examination of Poly-L-alanine and Poly-D,L-alanine in Aqueous Solution" *J. Am. Chem. Soc.* **1963**, *85*, 1193-1197.

looks at the alanine NH's and results in cross peaks off of the diagonal between helical

residues. The connected chain of cross peaks indicates that, with the exception of Ala-1

the alanine region is helical. The key question is, does the helical region extend beyond
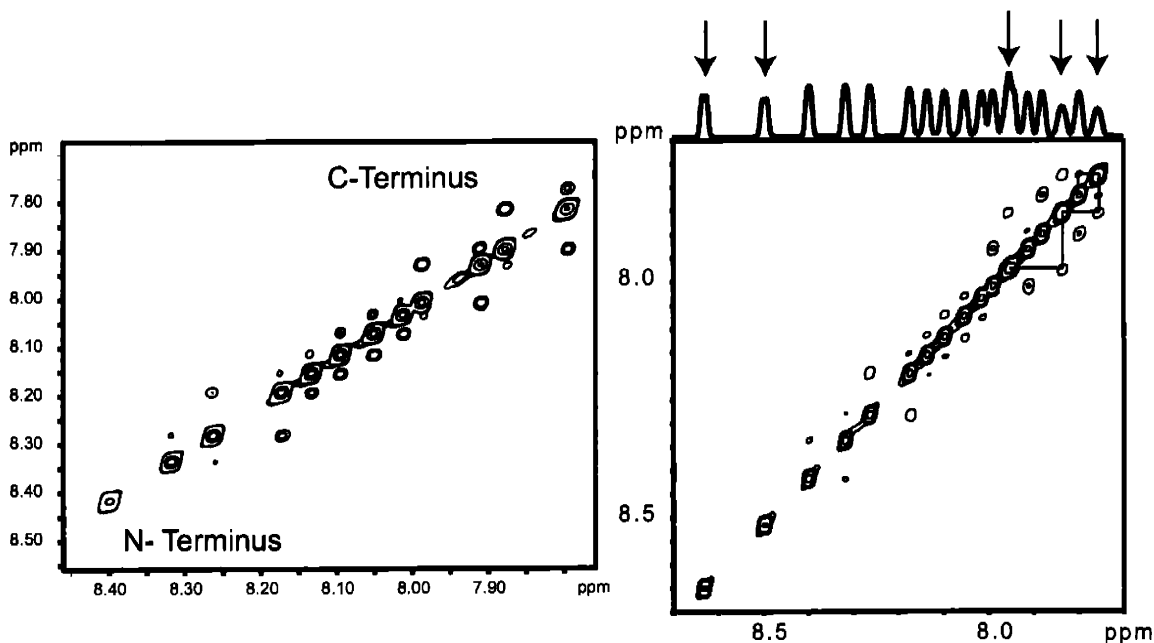
the alanine region?

Figure 18: K_3-A_{12}-K_3-W-NH$_2$ A) ^{15}N filtered NOESY spectrum isolating the alanine NH resonances and their cross peaks, indicates that the alanine region is helical B) ^{15}N decoupled NOESY spectrum with all of the Lys NHs, indicating that the helical structure extends into the Lys region.

A second NOESY NMR experiment was used to look at all of the amide NH's of

alanine and lysine. The lack of crosspeaks between the two Lys amide NH's at the N-

terminus, Figure 18B, strongly suggests that these residues are not in a helical

conformation, but the pattern of cross peaks observed within the helical alanine region

continues into the lysine residues at the C-terminus. This proves that helical structure is

not contained within the alanine region. The extension of the helical region into the Lys

residues forces the use of two w values to model the observed helicity, and the published reports of positive charge stabilization at the C-terminus create additional factors that need to be included in the LR analysis. Host peptides of this type are invalidated as choices for rigorous helicity studies.

NEW HELICAL HOSTS: RESULTS

In order to maintain the use of the solubilizing Lys residues, the N- and C-capping regions of the polyalanine core were engineered to isolate the helix stabilizing charge effects of Lys residues. Over a two year period, Justin Miller and I looked at the need to create a molecular spacer that would distance the charged residues from the alanine helix and restrict its length to the core region. The conceptual scheme of Figure 19 was used during the development and resulted in the selection of a pair of isolating spacers $Inp_2{}'L$, ${}'LInp_2$ that met the criterion established; freedom from aggregation was confirmed by analytical ultracentrifugation (AUC). Details of the development of the isolating spacer have been published and are not reported here.[77]

---

[77] a) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Short, Solubilized Polyalanines Are Conformational Chameleons: Exceptionally Helical If N- and C-Capped with Helix Stabilizers, Weakly to Moderately Helical If Capped with Rigid Spacers" *Biochemistry* **2001**, *40*, 305-309; b) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities" *J. Am. Chem. Soc.* **2002**, *124*, 945-962.

A	UV	SO	IS	A$_n$	IS	SO

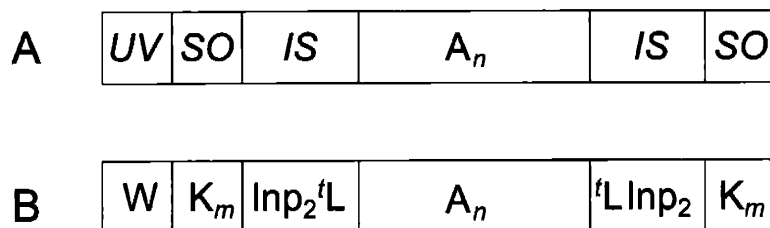| B	W	K$_m$	Inp$_2$^{t}L	A$_n$	^{t}LInp$_2$	K$_m$

Figure 19: A) Scheme for a new polyalanine host that incorporates a *UV* concentration reporter, *SO* solubilizing residues, *IS* isolating residues and a polyalanine core. B) The development phase proposed Trp as the UV reporter, a variable number of Lys residues for solubilization, and Inp$_2$^{t}L, ^{t}LInp$_2$ as isolating spacers.
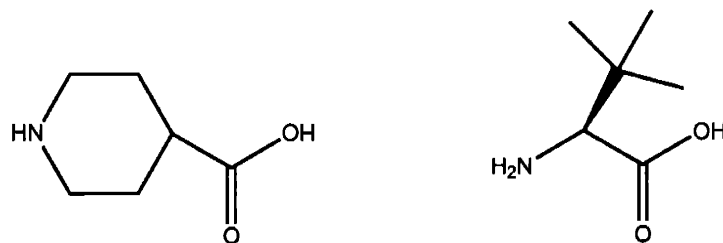
Figure 20: Amino acids used in the *IS* region. A) Inp, isonipecotic acids or 4-carboxypiperdine. B) ^{t}L, *tert*-Leucine.

The set of spacers accomplished the following: As pictured in Figure 20 the ring structure of Inp residues provides linear extension away from the helix. At the N-terminus the ^{t}L provided a secondary amide that allowed Ala-1 to join the helical region, while the bulky side chain prevents ^{t}L from adopting helical ϕ, ψ angles. At the C-terminus the ^{t}L provides the terminal amide that allows the last alanine to be helical. This spacing unit was tested to see if changes in the number of solubilizing residues in the *SO* region influence the alanine helicity as measured by CD. Consistent values of $[\theta]_{n,222.obs}$

as the length[78b] and identity[78] of the *SO* region were varied indicate that the *IS* region is effective. Justin Miller then created the length series, W-K_m-Inp_2'L -A_n- $^tLInp_2$ -K_m with (n = 4 - 45) 4 to 45 and CD data were obtained and are shown below.
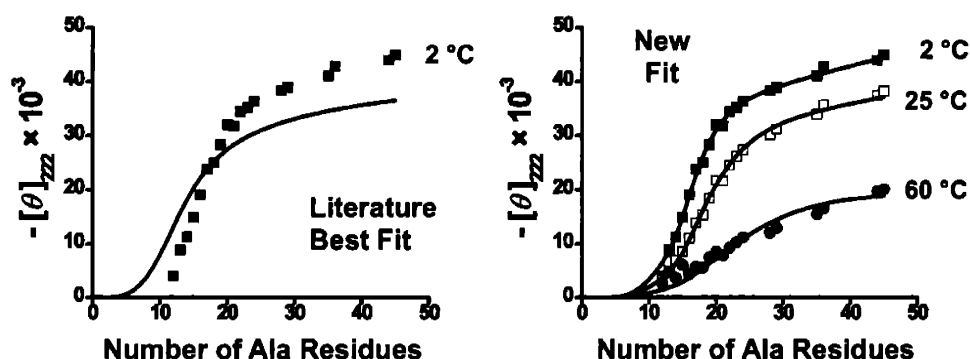

Figure 21: -$[\theta]_{n,222,obs}$ vs. *n* for W-K_m-Inp_2'L-A_n-$^tLInp_2$-K_m (*n* = 4-45) in water A) A **LR prediction using the literature value for the limiting ellipticity,** $[\theta]_{\infty,222}$ **underestimates the experimental data. B) Increasing the limiting ellipticity by 150%, and introducing a cooperative** w_{Ala} **provides a good fit to the experimental data at three temperatures.**

The experimental CD characterization of FH for this series is modeled by LR with the intent of assigning w_{Ala}. The literature LR derived CD prediction seen in Figure 21A has similar shape but the ellipticity range is underestimated. In order for LR to accurately assign *w* values, the values of $[\theta]_{\infty,222}$ and *x* within the length dependent ellipticity function must be known: $[\theta]_{n,222}$ = $[\theta]_{\infty,222}$ (1-*x*/*n*) This error of predicted CD can be attributed to an underestimation of the limiting value $[\theta]_{\infty,222}$ that is well precedent for helical polyalanines by both Kemp[79] and Baldwin.[61] Figure 21B illustrates a fit when

---

[78] Kemp, Miller unpublished data on the use of arginine as a solubilizer.

[79] Wallimann, P.; Kennedy R. J.; Kemp, D. S. "Large Circular Dichroism Ellipticities for N-Templated Helical Polypeptides Are Inconsistent with Currently Accepted Helicity Algorithms" *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1290-1292.

both w and $[\theta]_{\infty,222}$ are treated as free variables.[80] Without a rigorous calibration of

limiting $[\theta]_{\infty,222}$ of helical peptides as measured for CD, the technique can only be used

to solve for relative and not absolute w values. The newly proposed length dependent w

values and $[\theta]_{\infty,222} = -62,000$ deg cm^2 dmol^{-1} are similar with those of previous Kemp

group studies.[79] Despite this agreement, in itself this consistency cannot be used to

validate the model. Whenever possible, it is necessary to exclude other models that may

provide equivalent fits. This thesis explores the possible ways of measuring absolute

helical propensities by defining the relationship between the CD data and the fractional

helicity of alanine based peptides using multiple analytical methods.

Thesis Goals

This thesis focuses on the energetic characterization of polyalanines as a first step

in efforts toward the long-range Kemp group aim: a chemical understanding of the

factors that govern helicity, embodied in an algorithm that accurately predicts the

stability of peptide helices from their amino acid compositions and sequences. The work

presented follows the blueprint for the new spaced and solubilized helical host by

expanding the number of analytical techniques used. It also explores the use of new

highly stabilizing N- and C-terminal caps of the polyalanine core. These new approaches

provide a rigorous cross check of the analytical techniques in an effort to establish a

consistent experimental measure of FH of alanine peptides. Each of the following

chapters will explore the analytical techniques in Figure 22.

---

[80] Details of the error analysis and fitting parameters used can be found in the Ref 77b.
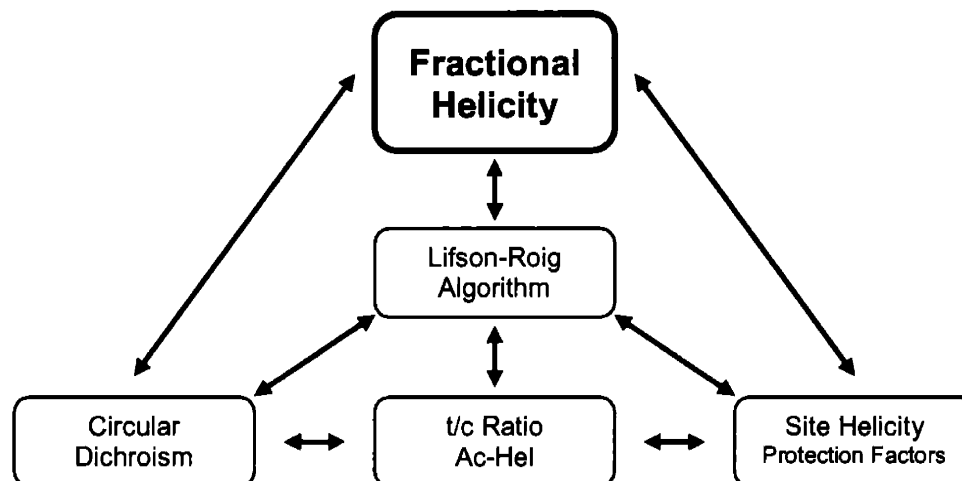
Figure 22: Corning fractional helicity: Cross checking three independent analytical methods in addition to the LR algorithm will be used to develop a consistent picture for the characterization of polyalanine helices in long and short context.

Utilizing the new spaced solubilizing methodology, an expanded series of Ac-Hel capped peptides with 4-14 alanine residues is introduced in Chapter 2 and improves on the previous, criticized limit of six residues. This Chapter presents the first and only energetic analysis of short alanine helices and investigates the response of the t/c ratio to length and temperature changes in the absence of a Lys residue. The results refute Baldwin's contention that the Lys side chain of earlier Kemp group studies interacts with Ac-Hel and largely governs its t/c ratio. A cooperative model for the alanine w value and its temperature dependence is presented and compared with the previous Kemp and Baldwin results. Chapter 2 closes by constructing a cross check that compares t/c and CD as helicity quantifiers. The consistency observed for this cross check validates Ac-Hel as an analytical tool for the measurement of helicity and provides closure to the use of Ac-Hel with short alanine peptides. The correlation of t/c and CD also confirms the need for calibration of the disputed limiting ellipticity $[\theta]_{\infty,222}$.

Chapter 3 focuses upon using hydrogen exchange to further characterize the FH

of polyalanine hosts with protection factors. In the first application, hydrogen exchange

is used to support the model for using length dependent w values used in the Miller study

and is also used to model the t/c and CD data in Chapter 2. The central residue protection

factors of W-K$_m$-Inp$_2$'L -A$_n$ - 'LInp$_2$ -K$_m$ ($n = 5$ - 25) and position dependent PF of $n =15$

are compared and analyzed with a cooperative w_{Ala} LR analysis.

In a second application of hydrogen exchange, protection factors are measured to

quantitate the factional helicity of a new 8-residue polyalanine that is capped by both

highly stabilizing N- and C- terminal stop signals and that exhibits a classical CD

signature. Hydrogen exchange can be used as an independent measurement of fractional

helicity. This fractional helicity value combined with circular dichroism data, provides a

proof of principle for construction of a future series of calibration standards that will

establish $[\theta]_{n,222}$; from CD properties of the entire length series one can solve for $[\theta]_{\infty,222}$

and x of in Equation 5.

SIGNIFICANCE OF POLYALANINES

Why is such an intense effort focused on the properties of a single residue within helices? As discussed above, the polyalanine helix, with its repeating small methyl side chain, creates a minimal context and uniform helical environment ideal for isolating and studying helical stabilizing factors that are independent of their sequence site and interactions with neighboring amino acids side chains. Without an accurate assignment of the w values that describe the helical alanine host, it is not possible to assign absolute w values to guest residues or quantitate other helical stabilizing effects, such as salt bridges, Phe-His interactions and hydrophobic clusters. The accurate measurement of these effects is essential to algorithms that predict helicity and protein folding.

The controversy regarding intrinsic helical propensities is rooted in the over generalized assumptions regarding the active stabilizing mechanisms within a helical host and investigators dismissing conflicting results with minimal discussion. Resolving this dispute requires incisive new experiments that use appropriate measurement tools in conjunction to solve for single variables within a minimal context. The new solubilized-isolated polyalanine may provide this minimal helical context, once fully characterized, that will allow the study of other helical stabilizing mechanisms.

Resolution of controversies surrounding the helical polyalanines may also contribute significantly to the currently active field of computational modeling. Rigorous molecular or *ab initio* modeling of proteins or peptides currently strain the speed and memory limits of the computers accessible to most researchers. In the interests of efficiency, most investigators prune their structures of extra atoms, and polyalanines are often their modeling choices. Unfortunately, reliable experimental data for tests of

modeling predictions are currently problematic, due to the solubility of polyalanines.

Extensions and validations of a database of structural and thermodynamic information for

polyalanines utilizing the *IS-SO* methodology could help fulfill this need.[81]

Polyalanines also have high current scientific visibility for materials scientists and

biomedical researchers. The older literature demonstrates that, in addition to the

currently fashionable helices, polyalanines also readily adopt sheet conformations and

recent *ab initio* modeling of polyalanines in the gas phase has shown similar stabilities

for these conformations.[82] This finding is now paralleled by experimental results. Silk

fibroin, the classic model for β-structure, is a striking case. Alanine and glycine, usually

present in roughly equal amounts, dominate fibroin amino acid compositions, and long

poly-L-alanine regions (PLAs), spaced by glycine-rich regions characterize their

sequences. A recent MAS NMR study of ^{13}C-labeled, silk fibers by Nakazawa and

Asakura[83] shows that PLA is largely α-helical when freshly isolated from larva, but

exposure to protic solvents transforms it into a classic β-fibroin. In silk, PLA is

ambidextrous. Similar findings are reported for resin-bound polyalanines[84] and PLA

---

[81] For an example of the synergy that can develop between computational chemists and their experimental counterparts compare Kennedy, R. J.; Kwok, K. Y.; Kemp, D. S. "Consistent Helicities from CD and Template t/c Data for N-Templated Polyalanines: Progress Toward Resolution of the Alanine Helicity Problem" *J. Am. Chem. Soc.* **2002**, *124*, 934-944 and Dannenberg et al. Ref 82.

[82] Wieczorek, R.; Dannenberg, J. J. "H-Bonding Cooperativity and Energetics of α-Helical Formation of Five 17-Amino Acid Peptides" J. Am. Chem. Soc. 2003, 125, 8124, 8129.

[83] Nakazawa, Y.; Asakura, Y. "High Resolution ^{13}C CP/MAS NMR Study on Structures and Structural Transitions of Antheraea pernyi Silk Fibroin Containing Poly(L-alanine) and Glycine-Rich Regions" *Macromolecules* **2002**, *35*, 2393-2400.

[84] Warrass, R.; Wieruszeski, J. M.; Boutillon, C.; Lippens, G. " High- Resolution Magic Angle Spinning NMR of Resin-Bound Polyalanine Peptides" *J. Am. Chem. Soc.* **2000**, *122*, 1789-1795.

regions of natural and molecularly engineered spider silks,[85] currently of interest for their tensile strength.

Small peptide models for the helix-sheet conformational transition corroborate these findings. In a CD study of the sequence Ac-K-A_n-K-NH$_2$, Blondelle et al.[86] examined concentration and length dependence of β-structure formation, which is dominant for $n > 14$. Significantly, Ac-K-A_n-K-NH$_2$ of this study strongly resembles sequences of the solubilized helical polyalanines reported earlier in this Chapter.

Polyalanines have current medical relevance. Genetic defects resulting in changes in the lengths of short polyalanine regions of proteins that govern development, tumor suppression, or degeneration[87] have been linked to human cancer,[88] mental retardation,[89] congenital limb formation,[90] and muscular dystrophy.[91] Perutz et al.[92] have

[85] Rathore, O.; Sogah, D. Y. "Self-Assembly of β-Sheets into Nanostructures by Poly(alanine) Segments Incorporated in Multiblock Copolymers Inspired by Spider Silk" *J. Am. Chem. Soc.* **2001**, *123*, 5231-5239.

[86] Blondelle, S. E.; Forood, B.; Houghten, R. A.; Perez-Paya, E. "Polyalanine-Based Peptides as Models for Self-Associated β-Pleated –Sheet Complexes" *Biochemistry* **1997**, *36*, 8393-8400.

[87] For a brief review of work prior to 1998 see page 19 of Shriver, S. P.; Shriver, M. D.; Tirpak, D. L.; Bloch, L. M.; Hunt, J. D.; Ferrell, R. E.; Siegfried, J. M. "Trinucleotide repeat length variation in the human ribosomal protein L14 gene (RPL14): localization to 3p21.3 and loss of heterozygosity in lung and oral cancers" *Mutation Research Genomics* **1998**, *406*, 9-23.

[88] Baxter, S. W.; Choong, D. Y. H.; Eccles, D. M.; Campbell, I. G. "Transforming growth factor beta receptor 1 polyalanine polymorphism and exon 5 mutation analysis in breast and ovarian cancer" *Cancer Epidemiology. Biomarkers & Progression* **2002**, *11*, 211-214.

[89] Stromme, P.; Mangelsdorf, M. E.; Shaw, M.A.; Lower, K. M.; Lewis, S. M. E.; Bruyere, H.; Luetcherath, V.; Gedeon, A. K.; Wallace, R. H.; Scheffer, I. E.; Turner, G.; Partington, M.; Frints, S. G. M.; Fryns, J.-P.; Sutherland, G. R.; Mulley, J. C.; Gecz, J. "Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy" *Nature Genetics* **2002**, *30*, 441-445.

[90] Bruneau, S.; Johnson, K. R.; Yamanoto, M.; Kuroiwa, A.; Dubhoule, D. "The Mouse Hoxd13^{spdh} Mutation, a Polyalanine Expansion Similar to Human type II Synpolydactyly (SPD), Disrupts the Function but Not the Expression of Other Hoxd Genes" *Developmental Biology* **2001**, *237*, 345-353.

[91] a) Brais, B.; et al. "Short GCG expansions in the PABP2 gene cause oculophyryngeal muscular dystrophy," *Nature Genetics* **1998**, 164-167. b) Bao, Y. P.; Cook, L. J.; O'Donnovan, D.; Uyama, E.; Rubensztein, D. C. "Mammalian, Yeast, Bacterial, and Chemical Chaperones Reduce Aggregate Formation and Death in a Cell Model of Oculopharyngeal Muscular Dystrophy" *J. Biol. Chem.* **2002**, *277*, 12263-12269.

[92] Perutz, M. F.; Pope, B. J.; Owen, D; Wanker, E. E.; Schertzinger, E. " Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques" *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5596-5600.

recently noted a likely kinship of these disorders with Huntington's disease, which is attributable to genetic defects that grossly extend the length of polyGln regions of the exon-1 peptide of the Huntington protein, with consequent formation of neurotoxic aggregates.

These results highlight a need for a better understanding of the conformational biases of short N- and C-capped polyalanines. Small variations in polyalanine length or the structures of its N- and C-caps result in formation of either unaggregated α-helices or in aggregated β-structures. This ambidextrous potential of polyalanine peptides has hitherto been largely unsuspected by workers in the helicity field, perhaps owing a need to avoid at all costs the time-intensive characterization of peptides that are aggregation-prone. Given the biomedical significance of mis-folded proteins forming aggregates plaques in human diseases,[93] a high priority must be given the development of empirically grounded rules that reliably predict helical or sheet structures for polyalanines.

---

[93] Dobson, C. M. "The Structural Basis of Protein Folding and its Links with Human Disease" *Phil. Trans. R. Soc. Lond. B.* **2001**, *356*, 133-145.

CHAPTER 2: SHORT POLYALANINE HELICES

INTRODUCTION

Chemists often hypothesize about and study complex molecular properties using experimentally defined characteristics of simple molecules. In 1990, the Kemp group introduced[1] Ac-Hel as a stabilizing N-cap and a reporting conformational template.[2] Whereas for simple peptides in solution the shortest detectable helix has 10- 12 residues, Ac-Hel enables the experimental characterization of 3 - 7 residue helices.[3,2c] Just as Pauling's X-Ray analysis of di- and tri-peptides allowed him to hypothesize about the structure of the α-helix in large proteins, Ac-Hel provides a tool[2] for studying the stability of short to medium helical conformers that are a part of the helical manifold of larger peptides. It provides an alternative tool that complements traditional circular dichroism (CD) and protection factor (PF) analyses of peptide helicity. As described in Chapter 1, previously published t/c data and their analyses from the Kemp group have been challenged by other groups.[4] This chapter introduces new data and analyses that clarify the issues raised by these challenges.

---

[1] Kemp, D. S. "Peptidomimetics and the Template Approach to Nucleation of β-Sheets and α-Helices in Peptides" *Tibtech,* **1990**, *8*, 249-255.

[2] a) Kemp, D. S., Allen, T. J.; Oslick, S, "The Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. I . Characterization of the Reporting Helical Template Ac-Hel ," *J. Am. Chem. Soc.,* **1995**, *117*, 6641-6657; b) Kemp, D .S., Allen, T. J.; Oslick, S. Boyd, J. G., "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. II. Characterization of the Helical Structure of Ac-Hel-L-Ala$_6$-OH," *J. Am. Chem. Soc.,* **1996**, *118*, 4240-4248; c) Kemp, D. S.; Oslick, S. L.; Allen, T. J. "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. 3. Calculation of the Helical Propagation Constant *s* from the Template Stability Constants t/c for Ac-Hel$_1$-OH, *n* = 1-6" *J. Am. Chem. Soc.* **1996**, *118*, 4249-4255.

[3] a) Kemp, D. S.; Boyd, J. G.; Muendel, C. C. " The Helical s Constant for Alanine in Water Derived from Template-Nucleated Helices" *Nature* **1991**, *352*, 451-454.

[4] Rohl, C. A.; Fiori, W.; Baldwin, R. L. "Alanine is Helix-Stabilizing in Both Template-Nucleated and Standard Peptide Helices" *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682-3687; Spek, E. J.; Olson, C. A.; Shi, Z.; Kallenbach, N. R. "Alanine is an Intrinsic α-Helix Stabilizing Amino Acid" *J. Am. Chem. Soc.* **1999**, *121*, 5571-5572.

The aim of the research presented in this chapter is to use the _IS-SO_ methodology[5]

discussed in Chapter 1 to extend previous studies of t/c ratios with longer polyalanines

that lack internal solubilizing lysine residues. Analysis of this new series of polyalanines

will allow the assignment of the helical propensity (w_{Ala} value) for alanine based on the

known capping parameters for the Ac-Hel template. Monitoring the response of t/c in a

longer polyalanine series will address two criticisms: 1) Ac-Hel-A_6-NH_2, previously the

longest polyalanine peptide, is too short to form a real helix. 2) The t/c ratios previously

measured for the alanine-rich sequence Ac-Hel-A_5KA_n-NH_2 series are strongly biased by

the presence of the internal solubilizing Lys which, interacts directly with the Ac-Hel

influencing the t/c ratio and are therefore uninterpretable. In addition, this work will

provide a continuous length series ($n = 4$ - 14) to evaluate if the temperature independent

w values for short helices apply to longer ones.

This chapter begins with an explanation of the properties of the Ac-Hel template

and then defines quantitative mass action relationships between t/c ratios and helicity.

The next section continues with how the Lifson Roig[6] (LR) manifold of helical

conformations is influenced by the template and how the model can be tailored to include

Ac-Hel capped peptides and to model t/c ratios.[7] The previous w values for alanine and

[5] a) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Short, Solubilized Polyalanines Are Conformational Chameleons: Exceptionally Helical If N- and C-Capped with Helix Stabilizers, Weakly to Moderately Helical If Capped with Rigid Spacers" _Biochemistry_ **2001**, _40_, 305-309; b) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities" _J. Am. Chem. Soc._ **2002**, _124_, 945-962.
[6] Lifson, S.; Roig, A. "On the Theory of Helix-Coil Transition in Polypeptides" _J. Chem. Phys._ **1961**, _34_, 1963-1974; Qian, H.; Schellman, J. A. "Helix-Coil Theories: A Comparative Study for Finite Length Polypeptides" _J. Phys. Chem._**1992**, _96_, 3987-3998.
[7] Kemp, D. S. "Construction and Analysis of Lifson-Roig Models for the Helical Conformations of α-Peptides" _Helv. Chim. Acta._ **2002**, _85_, 4392-4423.

lysine assigned[8] from the t/c dependence on amino acid sequence and length are revisited as a prelude to the analysis and discussion of new results. The chapter concludes with a qualitative reanalysis of the previously reported t/c data for alanine-lysine peptides.

Chapter 1 introduced a scheme for using multiple analytical techniques to establish a consistent measurement of fractional helicity. Chapter 2 concludes with a constructed cross check of t/c and CD ellipticities as measurements of FH in an effort to revalidate Ac-Hel. This chapter will take steps toward the resolution of controversial aspects of the w_{Ala} assignment and the use of the Ac-Hel template that have been raised in the literature using multiple analytical methods.

BACKGROUND

AC-HEL PROPERTIES

Originally designed and synthesized[9] to overcome the energetically unfavorable initiation step that dominates helix formation for short peptides, the Ac-Hel template is able to stabilize short helical conformers so that they can be detected and analyzed by NMR and CD spectroscopy.[2] This template provides a pre-organized helical structure in the form of a acetyl-prolyl-proline sequence that is constrained to assume helical ϕ, ψ orientations at three of its four backbone $N-C^{\alpha}$ and $C^{\alpha}-CO$ bonds. As seen in Figure 9b of Chapter 1, which is reproduced in Figure 1A of this Chapter, in effect this pre-

---

[8] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. "For Short Alanine-Lysine Peptides the Helical Propensities of Lysine Residues (s Values) Are Strongly Temperature Dependent" _J. Am. Chem. Soc._ **1996**, _118_, 12234-12235.

[9] a) Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd. J. G.; Muendel, C. "Studies of the N-Terminal Templates for α-Helix Formation. Synthesis and Conformational Analysis of (2S, 5S, 8S, 11S,)-1-Acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricylco[2.8.2.0^{4,8}]-tridecane (Ac-Hel$_1$-OH)" _J. Org. Chem._ **1991**, _56_, 6672-6682. b) Kemp, D. S.; Curran, T. P.; Boyd. J.G.; Allen, T.J.. "Studies of the N-Terminal Templates for α-Helix Formation. Synthesis and Conformational Analysis of (2S, 5S, 8S, 11S,)-1-Acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricylco[2.8.2.0^{4,8}]-tridecane (Ac-Hel$_1$-OH)" _J. Org. Chem._ **1991**, _56_, 6683-6697.

organization pays the entropic v^2 conformational price owed by a normal peptide sequence that assumes a helical conformation.[10] When N-linked by an amide bond to a peptide sequence, Ac-Hel provides helical hydrogen bonding sites for the first three amide NH functions of the peptide, as seen in Figure 1B. The first turn of the helix is complete when the first template-linked amino acid adopts helical ϕ, ψ values, forming the first α-helical hydrogen bond. An effective helical template of the Ac-Hel family of structures is expected to substitute a more favorable helical initiation for the unfavorable v^2, making short Ac-Hel initiated helices experimentally detectable.

A   B
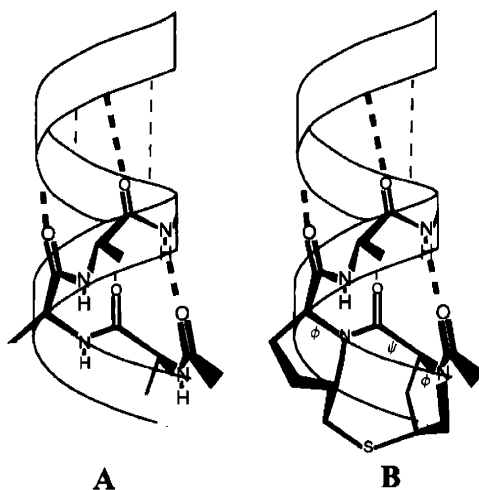
Figure 1: A) Three residues must adopt helical ϕ, ψ angles in order to form the first turn of an α-helix, B) The first two residues, which represent the initiation site, are replaced by the pre-organized Ac-Hel, increasing the stability of short helical peptide conformers.

---

[10] Kennedy, R. J.; Kwok, K. Y.; Kemp, D. S. "Consistent Helicities from CD and Template t/c Data for N-Templated Polyalanines: Progress Toward Resolution of the Alanine Helicity Problem" *J. Am. Chem. Soc.* **2002**, *124*, 934-944.

A synthesis of Ac-Hel-OH completed by Tim Curran[9] and later improved upon by Tim McClure[11] provides a 15-step convergent synthesis. Over a two month period, precursors and 5 grams of Ac-Hel-OH were synthesized for the studies in this thesis. In prior work, the molecular characteristics of the Ac-Hel template and Ac-Hel-peptide conjugates were extensively studied using molecular modeling and spectroscopic techniques.[2] The results indicate that the conjugates have three properties: 1) Short helices are initiated and have helical NMR signatures in both organic[9] and aqueous[2c] solvents. 2) Conformers of the Ac-Hel template detected by NMR spectroscopy reflect the stability of the helix.[2] 3) These conformers reflected conformational communication throughout the length of the helical sequence.[12] This chapter focuses on the first two template properties and provides a working example of the third in Figure 7.

What are the Ac-Hel conformations detected by NMR? During the initial analysis of the Ac-Hel template and its peptide conjugates, two sets of NMR resonances were detected; these correspond to two conformational states characterized by the slow equilibration of the *cis* (c) and *trans* (t) conformations on the N-terminal acetyl group. The structures of these states were unambiguously assigned using a NMR NOE analysis. Staggered (s) and eclipsed (e) conformations of the C-8, C-9 C--C bond were also detected. In total, three conformers[2,13] are detectable by NMR: cs, ts, te as pictured in Figure 2.

[11] McClure, K.; Renold, P., Kemp, D. S. "An Efficient Synthesis for Ac-Hel1-OH," *J. Org. Chem.,* **1995**, *60*, 454-457.

[12] A detailed discussion on conformational communication can be found in the following references: a) Kemp, D. S et al. Ref 9c; b) Deechongkit, S.; Kennedy, R. J.; Tsang, K. Y.; Renold, P.; Kemp, D. S. "An unnatural amino acid that controls polypeptide helicity β-Amino alanine, the first strong C-terminal helix stop signal," *Tetrahedron Lett.,* **2000**, 41, 9679-9683.

[13] Kemp, D. S.; Allen. T. J.; Oslick, S. "Development of a 3-State Equilibrium Model for the Helix Nucleation Template Ac-Hel1-OH" , "Peptides, Chemistry, Structure, Biology," J. Smith and J. Rivier, eds., ESCOM, Leiden, **1992**, pp 352-355

The ts and te conformations are in rapid equilibrium on the NMR time scale, resulting in a single averaged t-state resonance, but owing to amide resonance the $c \rightleftharpoons t$ rotation at the acetamide N-CO bond occurs slowly. As a result, at normal temperatures two sets of NMR are observed: a c-state resonance in which the acetamide CO cannot form the first helical hydrogen bond seen in Figure 2, and a t-state resonance set with chemical shift values that correspond to a weight averaged over the ts and te states.

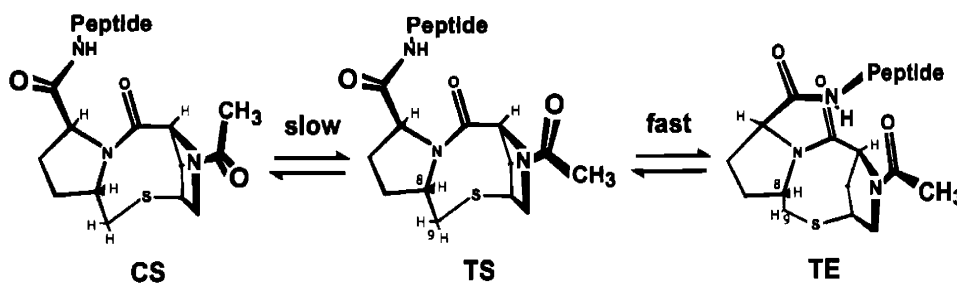**Figure 2: The *cis* (c) and *trans* (t) rotamers of the acetyl bond designate the first letter of each conformation label, and the staggered (s) and the eclipsed (e) conformations of the C-8, C-9 C--C bond designate the second letter. The ce state is not detected. Previous studies have shown that only the te state is capable of initiating helices.[14] Although there is some 3_{10} character of the first turn of the helix, α-helical residues have been shown to dominate longer Ac-Hel conjugates.**

NOE analysis and a quantitative correlation of the t/c values with the te-dependent chemical shifts determined that only the te state is capable of forming helices.[14] The cs and ts states do not have the right geometry to initiate the first turn of the helix. The stability of the helical amino acid sequence is thus reflected in a unique stabilization of

[14] a) The CD spectra of short peptides is responsive to the helix stabilizing additive TFE, and has shown that the cs and ts states have CD spectra typical of unstructured peptides. Only the te state has a helical CD signature. Oslick, S. L. "Characterization of Short Template-Nucleated Helical Peptides" Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 1996. b) NOE evidence for the lysine side chain stabilization of the helix is only seen in the te state: Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. "Template-Nucleated Alanine-Lysine Helices are Stabilized by Position-Dependent Interactions Between the Lysine Side Chain and the Helix Barrel" *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4025-4029.

the te state. The quantitative populations of t and c resonances, expressed as a t/c ratio measured by NMR integration, can thus be used as a measure of the stability of the helical peptide. The larger the t/c ratio, the more stable the helix formed by the attached peptide. Moreover, if the cs ⇌ ts ⇌ te equilibrium of the Ac-Hel template can be quantified as independent of the sequence of the remainder of the linked peptide, the mole fraction of te state, χ_{te}, can be determined from the t/c ratio. Analysis of the t/c response to changes in helical length and sequence allow the assignment of w, the helical propensities of the amino acids in the attached peptide. In initial studies, to minimize the number of independent w values, homopeptides were used.

Chapter 1 introduced the necessary steps required for conformational analysis of peptides and the assignment of w values:

1) A molecule can be assigned to a series of defined conformational states; h or c identifiers for each residue are used creating 2^n peptide conformations.

2) Each conformation is structurally characterized.

3) A temperature range is defined where the conformational manifold is at equilibrium.

4) A spectroscopic method or model can be applied to the manifold to quantify the amount of each conformation, expressed as condition dependent mole fraction χ_i.

Previous NMR studies[2a] establishes equilibrium, and the structural characterization of helical NOE's for Ac-Hel capped peptides satisfies points 2 and 3. The next step is to list the conformational states for the attached peptide for each of the cs, ts, and te Ac-Hel

conformations and see how their populations relate to the spectroscopically determined t/c ratio.[3,2c] This next section will look at both a simplified and Lifson-Roig lists of conformations that make up the state sum for templated peptides based on mass action.

CONFORMATIONAL MODELS: MAJOR HELIX APPROXIMATION

Solving for the helical propensity of alanine (w_{Ala}) requires the use of a peptide conformational model that describes the helical character of each possible conformation in solution. A conceptually simple model retains only the major, highly stabilized conformations. The perresidue helical/nonhelical two state h/c assumption adapted by LR is also used in the major helix approximation.

When Ac-Hel in the te conformational state is attached at the peptide N-terminus, all helical conformers extending from the template are selectively stabilized by the pre-initiated structure, and the magnitude of the template initiation constant $B >> v^2$ implies the dominance of the N- initiated major helix conformers. Helical conformers that are not initiated at the N-terminus are not Ac-Hel stabilized. The initiation constant B of Ac-Hel, shown experimentally to be approximately 70 times greater than the value of v^2, simplifies the list of 2^n possible peptide conformations to a total of (n + 1) major helical conformers.[2c] Incorporating the known template conformations in the peptide manifold, one can write the equilibrium expression for the conformers shown in Figure 3. The simplified state sum (SS) for a pentapeptide manifold of conformers is shown in Equation 1 and follows from the equilibrium expression. This simplified model was applied to t/c analysis of very short peptides for which its underlying approximations can be shown to be valid. The peptide state sums for the cs and ts states, in which the presence of Ac-Hel

does not result in helix stabilization, are approximated as the dominant all non-helical

(ccccc), weight = 1.0.[15]

$$cs\text{-}ccccc \overset{K_1}{\rightleftharpoons} ts\text{-}ccccc \overset{K_2}{\rightleftharpoons} te_0\text{-}ccccc \overset{K_3}{\rightleftharpoons} te_1\text{-}ccccc \overset{w}{\rightleftharpoons} te\text{-}hcccc \overset{w}{\rightleftharpoons}$$

1   K_1 K_1K_2 $K_1K_2K_3=B$ Bw^1

$$te\text{-}hhccc \overset{w}{\rightleftharpoons} te\text{-}hhhcc \overset{w}{\rightleftharpoons} te\text{-}hhhhc \overset{w}{\rightleftharpoons} te\text{-}hhhhh$$

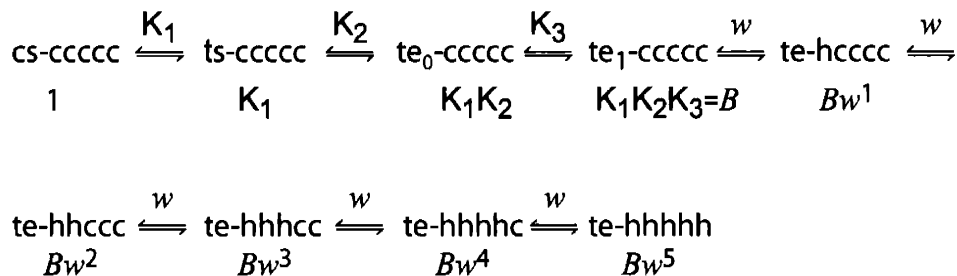Bw^2 Bw^3 Bw^4 Bw^5

Figure 3: Conformational equilibrium mass action model of Ac-Hel-pentapeptide. Major conformers are listed as the template conformer with the residue h/c identifier. *A* and *B* (*A*= K_1 + K_1K_2, *B*= $K_1K_2K_3$) are known template equilibrium constants, *w* represents the helical propensities of the attached polyalanine pentapeptide. The weight of each conformer is listed below.

SS= $1+ K_1 + K_1K_2 + K_1K_2K_3 + K_1K_2K_3w + K_1K_2K_3ww + K_1K_2K_3www + ... + K_1K_2K_3wwwww$

SS= $1 + K_1 + K_1K_2 + K_1K_2K_3\,(w + ww + www + wwww + wwwww)$

SS= $1 + A + B(w + ww + www + wwww + wwwww)$ $A= K_1 + K_1K_2$, $B= K_1K_2K_3$

Equation 1: A set of equalities that describe the state sum (SS) of the conformational manifold of the Ac-Hel capped pentapeptide in Figure 3. Constants *A* and *B* reflect the conformational bias of the template.

The experimental t/c ratio can be modeled using the mass action equilibrium expression and the associated state sum as shown in Equation 2. The template initiation constants *A* and *B* for Ac-Hel have been assigned previously from experimental data.[2c] The ts/cs ratio (*A*) was found to be independent of temperature, solvent additives and peptide structure; it reflects the intrinsic ts/cs bias of Ac-Hel linked to a peptide without helical structure. The value of *B* also has been assigned previously and reflects the

---

[15] This model roughly corresponds to a Zimm-Bragg conformational analysis: Zimm, B. H.; Bragg, J. K. "Theory of the phase transition between helix and random coil in polypeptide chains:" *J. Chem. Phys.* **1959**, *31*, 526-535.

efficiency of the template as a helix initiation site. The remaining equilibrium constants

in the expression are a series of w values for the attached homopeptide. Given the values

of A and B, the mole fraction of the te conformation (χ_{te}) can be assigned from the value

of the t/c ratio; the direct response of χ_{te} to the stability of the attached peptide allows t/c

to be used as a quantitative tool to assign w values. This mass action analysis was used to

assign w values of short Ac-Hel conjugated with polyalanine peptides and peptides

containing alanine and lysine. Ac-Hel peptide conjugates longer than 8 residues require

the use of the LR algorithm to model the minor partial helices not stabilized by Ac-Hel.

$$\frac{[t]}{[c]} = \frac{[ts] + \sum[te]_i}{[cs]} = \frac{[ts]}{[cs]} + \frac{\sum[te]_i}{[cs]} = A + B(w + ww + www + wwww + wwwww)$$

Equation 2: Simplified t/c ratio illustrated for Ac-Hel-A$_5$-NH$_2$. The NMR-derived t/c ratio reflects the populations of the ts and te states divided by the cs state value of 1.0. The concentration of the te state is dependent on the number (i) of helical peptide conformers. The ts/cs is a known constant A, measured independently; values for B have also been measured.
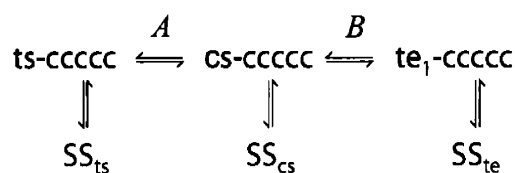
CONFORMATIONAL MODELS: LIFSON-ROIG & AC-HEL-PEPTIDES

The analysis of t/c can also be expanded more rigorously to the complete LR state

sum, including the treatment of isolated h residues in the unstructured state and v^2

initiated helices in all states.[10] In Chapter 1, the helical conformational states generated

by the Lifson-Roig (LR) residue two-state assumption is illustrated for a pentapeptide.

Table 1 indicates all of the helical and nonhelical states for a pentatpeptide.

Energetically, the bold conformations benefit, because B effectively replaces v^2, and these

conformers, having a single helical region, dominate the manifold. Although the bold

conformations identify these major conformers present when Ac-Hel is linked at the

peptide N-terminus, the less stable v^2 weighted helices are also included in the t/c LR

model. Their contributions become significant if $w > 1.3$ and if the length of the peptide

extends beyond 8 residues.

Number of helical residues	Helical Conformations	Non-Helical Conformations
5	**hhhhh**	
4	**hhhhc** + hhhch + hchhh + chhhh	hhchh
3	**hhhcc** + cchhh + chhhc	hhcch + hhchc + hcchh + hchch + hchhc + chhch + chchh
2		**hhccc** + chhcc + cchhc + ccchh + cchch + chcch + chchc + hccch + hcchc + hchcc
1		**hcccc** + chccc + cchcc + ccchc + cccch
0		**ccccc**

Table 1: The total 2^n states of a LR state sum for an Ac-Hel-pentapeptide. The major ($n + 1$) conformers include n single helical regions that abut the template and include the random all c conformation.

$$\text{ts-ccccc} \overset{A}{\rightleftharpoons} \text{cs-ccccc} \overset{B}{\rightleftharpoons} \text{te}_1\text{-ccccc}$$

$$\text{SS}_{ts} \qquad\qquad \text{SS}_{cs} \qquad\qquad \text{SS}_{te}$$

$$\text{SS}_{cs} = 1 + 5v + 10\,v^2 + 7\,v^3 + v^4 + 3\,v^2w + 2\,v^3w + 2\,v^2w^2 + v^2w^3$$

$$\text{SS}_{ts} = 1 + 5v + 10\,v^2 + 7\,v^3 + v^4 + 3\,v^2w + 2\,v^3w + 2\,v^2w^2 + v^2w^3$$

$$\text{SS}_{te} = 1 + 4v + 9\,v^2 + 7\,v^3 + v^4 + 2\,v^2w + 2\,v^3w + 1\,v^2w^2 +$$
$$w + w^2 + \qquad\qquad w^3 + \qquad\qquad w^4 + w^5$$

Figure 4: LR equilibrium expression for modeling t/c ratio of Ac-Hel peptide conjugates, first as a function of the template equilibrium (horizontal equilibria) and then as a function of the subordinate peptide equilibrium (vertical equilibria). The peptide state sums SS$_{cs}$ and SS$_{ts}$ are equivalent and are calculated using the standard LR weightings. SS$_{te}$ is similar to SS$_{cs}$, except the v^2 term for helical conformers that abut the template are removed, v^2w^5 becomes w^5; v and v^2 become w (te$_1$-hcccc) and w^2 (te$_1$-hhccc). The Ac-Hel stabilized weights are listed in vertical columns for comparison with the cs and ts states.

The t/c ratio can be modeled as a ratio of the state sums based on the equilibrium scheme pictured in Figure 4. The te state sum, SS$_{te}$, is dominated with the helical conformers that abut the N-terminus, and B effectively replaces v^2. The equivalent cs and ts state sums are dominated by the random coil states; for short peptides and $w < 1.5$ these generalizations apply. Equation 3 gives the resulting mathematical expression for the t/c ratio. Since the template parameters A and B have been experimentally determined, only the value of the helical propensity remains to be assigned. For a polyalanine, the states sums contain only a single variable w_{Ala}, this simplicity makes Ac-Hel-A$_n$- an ideal conformational system for assigning this key parameter.

$$\frac{[t]}{[c]} \;=\; \frac{[ts]}{[cs]} + \frac{[te]}{[cs]} \;=\; A\frac{SS_{ts}}{SS_{cs}} + B\frac{SS_{te}}{SS_{cs}} \;=\; A + B\frac{SS_{te}}{SS_{cs}}$$

Equation 3: LR definition of the t/c ratio of Ac-Hel peptide conjugates: the concentration of the te and cs states can be modeled fully as the ratio of the corresponding state sums. The main difference between the te and cs state sums is that the te state sum is generated such that v^2 is removed for conformers that abut the template, effectively replacing v^2 with B.

In Chapter 1, it was noted that previous literature studies by several groups assigned w_{Ala} a value between 1.0 and 1.6. Based on the assumption that $w_{Ala} > 1.0$, how will the stability of Ac-Hel capped peptides be reflected in the t/c ratio? Figure 6 lists the dominant helical conformers for an Ac-Hel-A_n length series. Inspection of the list shows that for the major helix approximation the state sum is a geometric series $(1 + w + w^2 + \ldots + w^n)$ If n is increased and w lies close to 1.0, the state sum is nearly proportional to the length n, but a plot of t/c vs. n for $w > 1.1$ is expected to show strong w-dependent upward curvature, and downward curvature if $w < 1.0$. The zero intercept of this plot is the parameter A, the parameter B is defined by the slope of the plot for the first few residues. Given the validity of the fundamental LR assumption that w_{Ala} is independent of the length of the helical sequence, the value of w_{Ala} is defined by the plot curvature observed for subsequent t/c data points.
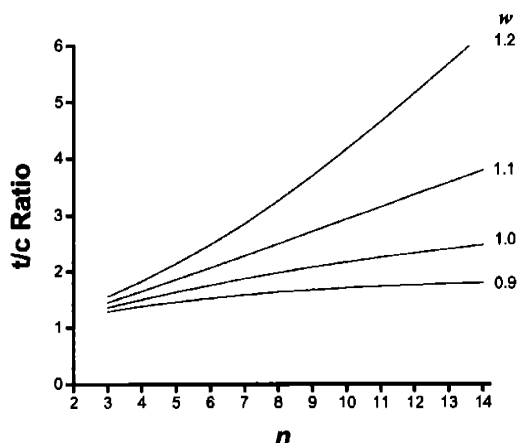
Figure 5: Expected dependence of t/c vs. *n* and *w* calculated for an Ac-Hel homopeptide length series using Equation 3. Upward curvature is observed for *w* > 1.1. Downward curvature is observed for *w* < 1.1. The zero intercept of this plot is the parameter *A*, the parameter *B* is defined by the slope observed for the first few residues. The *w* values are determined by the plot curvature observed for subsequent t/c data points.

Figure 6 compares the change in the dominant te state conformation as peptide length is increased from 5 to 6 residues. It illustrates a unique feature of the t/c ratio as a helicity quantifier. For the following reasons, t/c is more sensitive than CD or hydrogen exchange to any length-dependent changes in the value of *w*. In Chapter 1 it was noted that for any helical peptide described by a LR model with *w* values < 2, a measured helicity property like CD ellipticity or hydrogen-exchange-derived PF reflects the presence of substantial conformational averaging that can hinder detection of the helix stabilizing or destabilizing effects of site mutations or short length extensions. In the non Ac-Hel capped conformations of Figure 6, extending the overall length from a pentamer (*n* = 5) to a hexamer creates four (*n* - 1) new helical states, one each of the possible lengths. If *w* is close to 1.0, these are expected to contribute equally to the state sum. If a length dependence of *w* exists, it will be difficult to detect.

The presence of the Ac-Hel cap influences the relative stabilities of these conformations as apparent from inspection of the dominant helical conformations, in bold, in Figure 6. Of the n - 1 = 4 new helical conformers created in the LR manifold as the length is extended from five to six residues, 3 = (n - 2) are not stabilized by the Ac-Hel cap and make a relatively minor contribution to the state sum. Only one new, dominant conformer is stabilized by Ac-Hel. This longest helical conformer becomes the most stable within the manifold if w is greater than 1.0. The dominance of the longest helical conformer will reveal any length-dependent change in helical stability in plots of t/c vs. n for a series of Ac-Hel capped polyalanines.
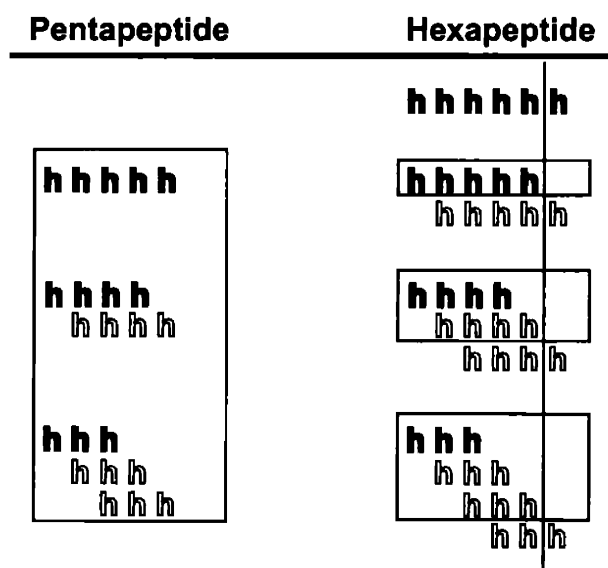
Figure 6: Comparison of the dominant (bold) Ac-Hel initiated and minor (hollow) helical peptide conformers. The boxes indicate conformers that are shared by both manifolds. When comparing the t/c ratios for penta- and hexapeptides, the only new dominant conformer is completely helical and is responsible for a majority of the change in t/c if w > 1.0. The c residue identifiers are omitted for clarity.

PREVIOUS RESULTS

In previous Kemp group studies two peptide motifs have been attached to Ac-Hel. Short polyalanine peptides with one to six residues defined the first series studied in both organic[9] and aqueous[2c] solvents. Unfortunately for longer polyalanines t/c values could not be measured, owing to their insolubility and tendency toward aggregation in aqueous solvents; t/c data for the first three member of the series were used to define the Ac-Hel initiation constant B, leaving only three t/c data points for the assignment of w_{Ala} and its temperature dependence. To extend the sequence length and overcome the water insolubility, second motifs containing single charged residues such as Arg, Lys, His and β-amino alanine were introduced within the polyalanine peptide sequence, extending the length to 10 and some cases 12 residues; t/c values were measured at 2, 25, and 60 °C in water.[8]

The first example of Figure 7 typifies these early studies. It demonstrates a linear length dependence of t/c for Ac-Hel-A_n-NH$_2$ which indicates that $w = 1.0 \pm 0.05$, implying that all peptide conformers are of nearly equal stability. A very small temperature dependence for w_{Ala} over the range of 2 to 60 °C was observed for this series. This result is surprising, since as noted in Chapter 1, in a LR model w carries the temperature dependence of peptide helicity. However, Scheraga had previously noted an unusually small temperature dependence for w_{Ala}, implying that the correspondingly small melting effect expected for short peptides sequences might easily escape detection. The longer heteropeptide sequences in the t/c database would however be expected to show melting properties. The second example of Figure 7 demonstrates this property. This example shows the t/c data for Ac-Hel-A$_4$-R-A$_{(n-5)}$-NH$_2$ series, for which several

noteworthy features are evident. The jump observed for a six-residue peptide when a charged residue replaces the C-terminal alanine indicates that the new helical te state is more stable, implying either that the w value for Arg is larger than w_{Ala} or that the charged residue is stabilizing the C-terminus.[16] The linear increase of t/c beyond the Arg residue further indicates that extension of the helical region beyond an -A_4R sequence results in a length dependent increase in stability. Data obtained when replacing Arg with Orn, which has a shorter positively charged side chain, demonstrates the same qualitative features, but the Orn helices have quantitatively smaller t/c ratios implying that they are less stable than the Arg helices, consistent with the w values for Arg and Orn guests.[17]

The third example in Figure 7 replaces the Arg guest with the unnatural residue, β-amino alanine (β). The t/c data identify this amino acid as assuming a unique role as a helix stabilizing stop signal. An increase in t/c is seen when β is placed at the end of the helix, similar to Arg. The remarkable and exceptional absence of an increase in t/c for the β series as peptide length is extended beyond the site of β indicates the absence of Ac-Hel intiated helices extending into the C-terminal alanine sequence. The lack of increase in t/c for the β series indicates the there is no further helical propagation, with conformational communication between Ac-Hel and the peptide only up to the β residue. For this reason, β has been defined as a helix stabilizing stop signal.[12b]

---

[16] Very specific experiments would be necessary to deconvolute the intrinsic capping parameters, charge interactions and the intrinsic w values for charged amino acids. w_{Lys} is typically reported as position dependent, incorporating the average effects.

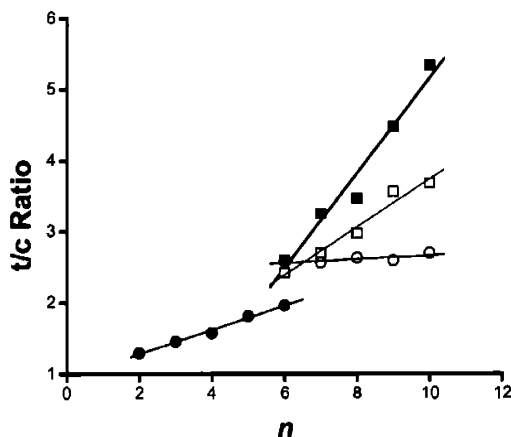[17] Tsang, K. Y.; Kemp, D. S. unpublished results.

Figure 7: Plots of t/c vs. n for peptide series in water at 25 °C. (●) Ac-Hel-A_n-NH_2, (_n_ = 2-6); (■) Ac-Hel-A_4-R-A_(n-5)-NH_2; (□) Ac-Hel-A_4-O-A_(n-5)-NH_2 (○)Ac-Hel-A_4-β-A_(n-5)-NH_2 β= β-amino alanine. Within this size range, the t/c ratio is proportional to the weights of the helical conformers initiated at the template (the single helix approximation). The solid trend lines are calculated with by linear least squares. Over short peptide lengths, linear approximations of the LR function are valid.

The longer series Ac-Hel-A_4-R-A_(n-5)-NH_2 did show a large temperature dependence, the t/c ratio decreased with increased temperature indicating a destabilization of the helical peptide. In a similar series, Ac-Hel-A_4-K-A_(n-5)-NH_2 the temperature dependence was attributed to the loss of the stabilizing lysine side chain interaction with the helix barrel.[14b] Within a larger series[17] of different length polyalanines with different Lys mutants, there is good agreement between the predicted and experimental t/c ratios, as shown in Figure 8. The agreement was based on the assumption that the temperature independent w_{Ala} = 1.0 continues beyond six residues and w_{Lys} is temperature and position dependent. A complete discussion of the reasoning behind these assumptions goes beyond the scope of this thesis and is detailed in the literature.[8]

An intense effort was made to deconvolute the alanine and lysine w values through study of a large peptide database. At 25 °C, t/c values were measured for a mono-Lys series of total lengths 10 and 12 residues with the Lys site varied from position 3 to 6, and an additional di-Lys series was prepared in which the spacing between the pair varied from 0 to 5 and the separation between Ac-Hel and the first Lys was varied from 2 to 5.[17] Although the introduction of charged residues into the peptide sequence provided an experimentally expedient access to a larger database of Ac-Hel stabilized peptides, analysis of its t/c data required assignments for both w_{Ala} and w_{Lys}. As noted in Chapter 1, amino acids bearing a positively charged side chain are expected to stabilize any helical conformation in which it appears at the C-terminus, but only moderately stabilize or even destabilize conformations in which these amino acids appear at other sites. This implies that w_{Lys} must be strongly site dependent. Although some investigators have argued that an average value of $<w_{Lys}>$ can be used successfully for quantitative modeling, simple inspection of the t/c data for the mono-Lys showed that at least 5 values of w_{Lys} would be required for accurate modeling.

Unambiguous assignment of all of these as well as w_{Ala} from the t/c data appeared to be daunting. Provided the data were found to support it, a plausible assumption applied w_{Ala} values assigned from the Ac-Hel-A_n-NH$_2$ ($n = 1$ - 6) series to the much longer data set and varied the w_{Lys} values based on the distance from the template. The average best-fit w_{Lys} value at 25 °C resulting from this modeling was 3.0. If correct, this modeling result would imply that these peptides are primarily stabilized by the presence of Lys residues, consistent with the calculations of the Scheraga group, but not with the assignments of the Baldwin group and others.
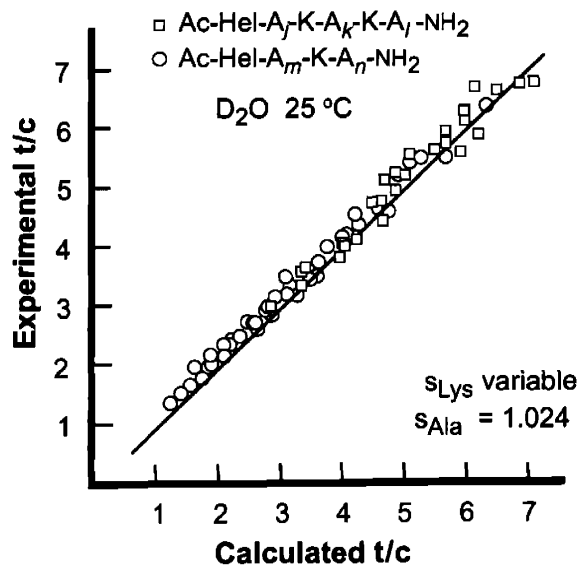
Figure 8: Correlation between expected and experimental t/c ratios for a series of mono and di-lysine Ac-Hel polyalanine sequences. Analysis by D. S. Kemp[17] using the major helix approximation $s = w / (1 + v)$.

The excellent linear fit for the large database of peptides seen in Figure 8 can be taken as validation of the assumptions underlying the calculation of the t/c model. However this fit does not invalidate any of a potential range of alternative multi-parameter models or assumptions. The key assumption underlying Figure 8 is a lack of length and temperature dependence for $w_{Ala} = 1.0$ when the length n is extended beyond the experimental limit of six residues. The *IS-SO* experiments of Chapter 1 open up the possibility of testing this assumption experimentally; C-terminal caps based on the *IS-SO* principle should allow measurement of t/c data series for Ac-Hel-A_n-*IS-SO* peptides in which n extends beyond 12 residues. A definitive finding would be a similarity of the t/c data for this new series to those of the Arg guest series of Figure 7. In this case, the underlying assumption of Figure 8 is rendered invalid. This eventuality would underline the need, as discussed in Chapter 1, for verification of all key helicity conclusions by

independent experimental methods. This chapter closes with a verification of t/c data using CD ellipticity values.

NEW RESULTS

The new *IS-SO* methodology was applied to this problem by terminating the Ac-Hel-A_n- sequence with -^{t}LInp$_2$-K$_4$-W-NH$_2$, in which ^{t}LInp$_2$ isolates the -A_n- core from the charged water solubilizing K$_4$ region, and the C-terminal W provides a UV reporter, required for accurate CD quantitation. This construct provides the first opportunity to extend the helical polyalanine region beyond six residues, allowing one to test the assumptions that the temperature independent $w_{Ala} = 1.0$ applies to polyalanine lengths extended beyond six residues. If these assumptions are correct the t/c ratio will increase linearly with length throughout the range of $n = 3$ - 14.

A **B**
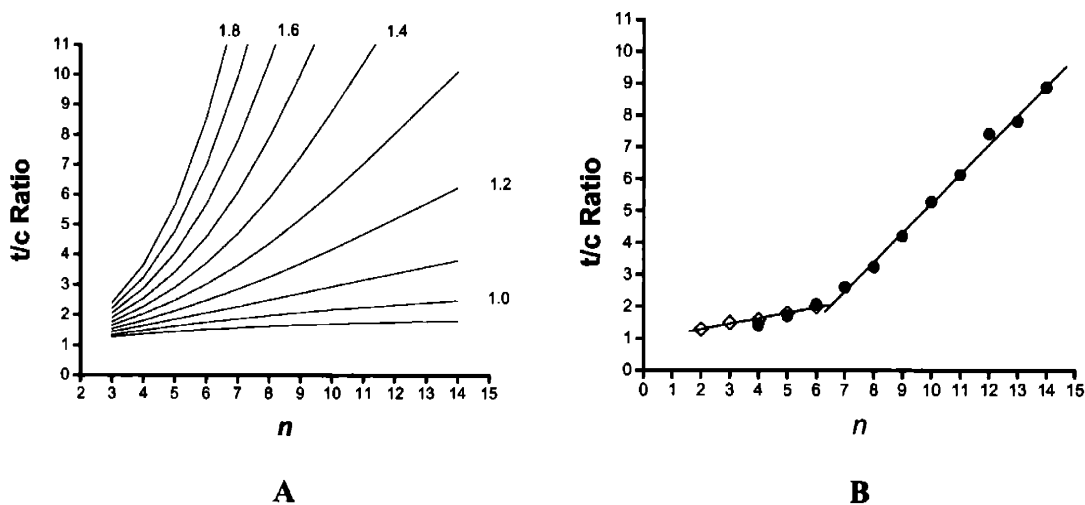
Figure 9: A) Model data for t/c vs. *n* generated using the LR analysis B) Experimental data for (●)Ac-Hel-A_n-^{t}LInp$_2$-K$_4$-W-NH$_2$ at 2 °C in water. These data show a good agreement with corresponding length data for the series (◊) Ac-Hel-A_n-NH$_2$, (*n* = 2-6) indicating that the *IS* region has a minimal effect on the alanine region. Two different linear regions are apparent in the plot of t/c vs. *n*. A 5% error is expected for t/c measurements.

Figure 9B plots the experimental t/c ratios for Ac-Hel-A_n-$^lLInp_2$-K_4-W-NH_2. There are two distinct linear regions. Peptides less than six residues appear to have w_{Ala} = 1.0 in agreement with the shorter Ac-Hel-A_n-NH_2, (n = 2-6) series. The break at $n > 6$ indicates an increase in stabilization for longer polyalanine helices and demonstrates the sensitivity of t/c to length extension that was explained in Figure 6. The t/c response of this polyalanine series resembles the Arg mutant series in Figure 7. Whereas the LR algorithm assumes that each amino acid in a homopolymers has the same contribution to helical stability, the t/c ratio indicates that the w value for alanine is length dependent, falsifying the LR and Kemp group assumptions for short and medium sized helices and invalidating the analysis of Figure 8. In order to rule out aggregation as the cause for this t/c change, analytical ultracentrifugation (AUC) was used immediately following t/c measurement to verify that the members of the peptide series were monomeric under NMR measurement conditions.[18]

Is there literature precedent for the change of helical stability with length? Karplus and Gao using *Ab Initio* calculations[19] have predicted that strength of hydrogen bonds cooperatively increases with length. An α-helix has three continuous hydrogen bond strands that run up the sides of the helix, each linking every third amide. One could imagine iterating the incremental increase of each strand, hydrogen bond cooperativity, as additional residues are added thus generating a $w(k)$ function. Such an analysis was carried out, but a simple pair of length dependent w values that change value at $n = 6$ was explored and found to give an equally good fit to the data.[10]

---

[18] Ac-Hel-A_{14}-$^lLInp_2$-K_4-W precipitated out of solution at 60 °C, but prior to heating the peptide was monomeric by AUC measurement at low temperatures; these results are reported in Chapter 4.

[19] Gou, H.; Karplus, M. "Solvent Influence on the Stability of Peptide Hydrogen Bond: A Supramolecular Cooperative Effect" *J. Phys. Chem.* **1994**, *98*, 7104-5105.

Traditionally the LR conformational model uses a single w value for alanine. Can the new data set be modeled using two alanine w values, one for short helices and one for long helices representing the two regions of the t/c vs. n graph? Such a condition can be modeled because the LR generated SS can be sorted by the length of the helical conformation (k) and separate w values applied to yield a new SS that is used when calculating a t/c ratio.[10] Figure 10 shows successful modeling of the t/c dataset at three temperatures.
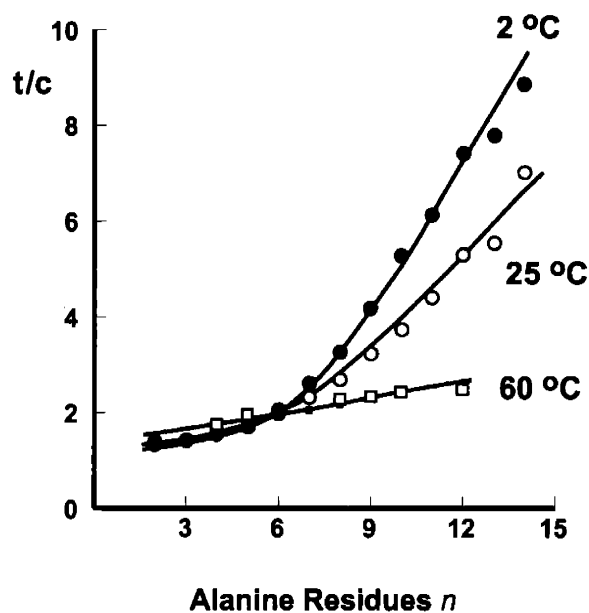
Figure 10: Experimental data points for Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W-NH$_2$ at 2, 25, 60 °C in water and the LR data. The solid trend lines are generated by the LR algorithm. A 5% error is expected for t/c measurements.

The LR algorithm was used to model the t/c response to the length of the Ac-Hel-alanine peptide. The literature value of $v = 0.048$ is implemented for peptide initiated helices, and template initiation parameters A and B were used as previously defined, $A =$

0.80-0.85, B = 0.15-0.20. The remaining variables are the two helical propensities of the polyalanine peptide for long and short conformers.

The best linear least squares fit was obtained if the SS is split into two regions, $n \leq 4$ and $n > 4$. For conformations with $n \leq 4$ the w was set at 1.0 consistent with previous studies. Larger conformations with $n > 4$ were assigned a new w'. In order to model the smooth transitions the w' values of n = 5, 6 were assigned values of $w'^{1/3}$ $w'^{2/3}$. The values for w' were determined to be 1.29 at 2 °C and 1.22 at 25 °C.[20] The 60 °C data can be fit with a range of w' values ranging from 1.01 < w' < 1.06. Details of how these changes are included on a classical LR algorithm can be found in the literature.[10]

How do these new results affect the previous assignments for the internally solubilized polyalanine peptides? Figure 11 illustrates the previous work with internally solubilized polyalanines, changes in t/c ratio are seen when an alanine residue is replaced with a lysine residue. The quantitative characterization of the w_{Lys} lies beyond the scope of this work, but the trend of increasing t/c when a Lys mutant is made at the C-terminal end of the helical sequence makes it clear that w_{Lys} is very site dependent. In this context, except when mutation occurs at site 3, an Ala $\longrightarrow$Lys mutation is helix stabilizing. However, the length dependent curvature of the new t/c values for the Ac-Hel-A$_n$-*IS-SO* series closely parallels that seen for the monolys mutant. Moreover, similar behavior is seen for t/c data between 2 and 60 °C.[21]

---

[20] Both B and w' were fit as free variables. The B values obtained from the analysis were consistent with the independent determination in earlier studies.

[21] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. "For Short Alanine-Lysine Peptides the Helical Propensities of Lysine Residues (*s* Values) Are Strongly Temperature Dependent" *J. Am. Chem. Soc.* **1996**, *118*, 12234-12235.
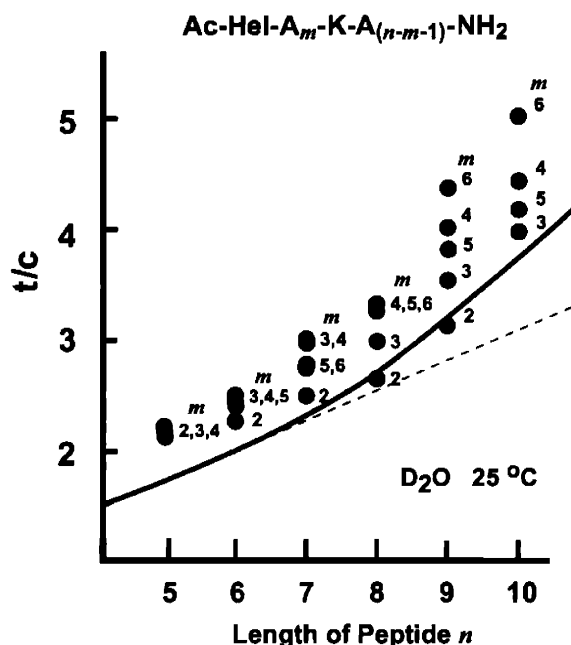
Figure 11: Comparison of the t/c ratios of Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W-NH$_2$ and Lys mutants Ac-Hel-A$_m$-K- A$_{(m-n-1)}$-NH$_2$ at 25 °C. The increase in t/c ratio indicates that lysine can be an intrinsic helical stabilizing residue when placed right of center in a polyalanine peptide sequence.

The values of A, B, w and w' obtained from t/c analysis can be used with the LR algorithm to determine the fractional helicity of the peptide as one method of determining FH. These parameters quantitatively define mole fractions for members of the manifold of helical lengths. Chapter 1 outlines a scheme in which several different analytical methods can be used to determine the fractional helicity of alanine peptide in order to settle literature controversies. This next section compares the t/c ratios of Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W-NH$_2$ with the CD ellipticities obtained for the same series.

VALIDATION OF AC-HEL AND T/C USING CIRCULAR DICHROISM

As mentioned in Chapter 1, the CD derived FH is dependent on the value for the $[\theta]_{\infty,222}$ and x, as shown in Equation 4. The current literature debate regarding the values

of these parameters makes it difficult to determine FH accurately using CD ellipticities measured for polyalanine peptides. The controversy regarding $[\theta]_{\infty,222}$ exists because there is no 100% helical peptide standard, due to the fraying. Although there are some protein and coiled coils standards that set $[\theta]_{\infty,222} = -37,000$ deg cm^2 dmol^{-1} res^{-1}, these standards all rely on helices constrained by hydrophobic packing of secondary structures within globular proteins or helical coiled coils; unaggregated peptides on the other hand are completely solvated by water. A brief review of literature attempts to resolve this issue is described below.

The helicity of short and medium sized peptides have been shown to be dependent on the concentration of helix stabilizing additives such as trifluoroethanol (TFE) and hexafluoro-2-propanol (HFP), and titrations of this kind have been used to estimate[22] $[\theta]_{n,222}$ and by extrapolation $[\theta]_{\infty,222}$. The effect of these additives often levels out around 16 mol %; if this is taken as the titration end point, calculations yield $[\theta]_{\infty,222} = -44,000$ deg cm^2 dmol^{-1} res^{-1}.

As noted in Chapter 1, a report by the Kemp group shows that Ac-Hel-capped Baldwin-Marqusee peptides can yield experimental $[\theta]_{n,222,obs}$ values significantly more intense than these extrapolated limits.[23] In this study we included a function that used two different $[\theta]_{\infty,222}$ values, one for short and one for long helices to calculate $[\theta]_{n,222}$. The dramatic helix stabilizing affect of Ac-Hel produces a manifold in which longer helices are more abundant than for those of non-stabilized peptides, explaining the

[22] Luo, P.; Baldwin, R. L. "Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water" *Biochemistry* **1997**, *36*, 8413-8421.

[23] Wallimann, P.; Kennedy R. J.; Kemp, D. S. "Large Circular Dichroism Ellipticities for N-Templated Helical Polypeptides Are Inconsistent with Currently Accepted Helicity Algorithms" *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1290-1292

increase in CD. Using two $[\theta]_{\infty,222}$ values for specific conformers lengths (k) in a LR

analysis approximated experimental $[\theta]_{n,222,obs}$ values satisfactorily. In this study for

helices of lengths $k < 10$, x in Equation 4 was set to 2.5 and $[\theta]_{\infty,222}$ to -44,000 deg cm^2

dmol^{-1} res^{-1} in accord with common usage, but for longer helices ($k > 10$) the

$[\theta]_{\infty,222}$ value of -44,000 was replaced with -61,000 deg cm^2 dmol^{-1} res^{-1}. Additional

support for this later assignment later came from a polyalanine length series. In

conjunction with the *IS-SO* development, Justin Miller measured the CD properties of

W-K$_m$-Inp$_2$^{t}L-A$_n$-^{t}LInp$_2$-K$_m$ ($n = 4$ - 45) as pictured in Chapter 1, Figure 21. Fitting this

data set with $[\theta]_{\infty,222}$ and w as free variables provided additional evidence that -61,000

deg cm^2 dmol^{-1} res^{-1} may be the limiting ellipticity for polyalanine helices; the

corresponding w_{Ala} assignment was 1.3 at 2 °C.[5b] Other investigators have observed

experimental ellipticities that exceed the current literature value for FH = 1.0.[24]

$$FH_{helix} = \frac{[\theta]_{n,222,obs}}{[\theta]_{n,222}} \qquad [\theta]_{n,222} = [\theta]_{\infty,222}\left(1-\frac{x}{n}\right) \qquad 2 < x < 5$$

Equation 4: CD derived experimental determination of fractional helicity as the observed perresidue molar ellipticity $[\theta]_{n,222,obs}$ divided by the limiting perresidue molar ellipticity $[\theta]_{n,222}$ for length n at 222 nm. The limiting perresidue molar ellipticity is calculated from a value of the perresidue molar ellipticity for a helix of infinite length $[\theta]_{\infty,222}$ with moderate length correction (1 - x/n). Experimental fits for x have ranged from $2 < x < 5$.[25] If the values for $[\theta]_{\infty,222}$ and x are defined, $[\theta]_{n,222}$ can be calculated. The current controversy over the value of $[\theta]_{\infty,222}$ for polyalanine peptides casts uncertainty over the calculated values of FH.

[24] Lazo, N. D.; Downing, D. T. "Circular Dichroism of Model Peptides Emulating the Amphipathic α-Helical Regions of Intermediate Filaments" *Biochemistry* **1997**, *36*, 2559-2565.

[25] Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W., Kallenbach, N. R. "The Helix-Coil Transition in Heterogeneous Peptides with Specific Side-Chain Interactions: Theory and Comparison with CD Spectral Data" *Biopolymers* **1991**, *31*, 1605-1614.

Can one construct a rigorous quantitative test of the quality of correlation between experimental t/c values and CD ellipticities, without accurate values for the limiting ellipticity $[\theta]_{\infty,222}$? The short answer is yes. To achieve this important test, we have taken a two step approach. First, using the LR algorithm as applied to the Ac-Hel peptides and plausible ranges for w and $[\theta]_{\infty,222}$, model data have been constructed for both t/c and $[\theta]_{n,222,calc}$; a functional dependence between this pair of parameters has been sought that remains robustly linear for values lying within the range of accurate t/c measurements and typical values of w. Second, the linearity of this functional dependence is tested for the experimental CD ellipticities and Ac-Hel t/c values. If linearity is observed, these parameters can be taken as providing mutually consistent quantitative measures of helicity.

FH and t/c are not expected to correlate linearly. As shown earlier, t/c is expected to be proportional to the state sum for the helical te state, and like the state sum, t/c has a potential range from 1 to infinity. By contrast, $[\theta]_{n,222,obs}$ is proportional to FH which has a range of 0 to 1. One can transform the experimental ellipticity data from $[\theta]_{n,222,obs}$ to $[\theta]_{n,222,obs} / ([\theta]_{\infty,222} - [\theta]_{n,222,obs})$; like t/c the latter ellipticity transformation has a range from 0 to infinity.

Accurate NMR integration limits define the useful range of t/c from 1 to 10. Can one expect a linear correlation between t/c and transformed ellipticity over this t/c range? Figure 12 addresses this question. The data set on which this Figure is based was constructed by setting $[\theta]_{\infty,222}$ to -61,000 deg cm^2 dmol^{-1} res^{-1} and varying w and the helical length over large ranges. It was analyzed by assigning three different values of $[\theta]_{Test,222}$ used in the transformation function $[\theta]_{n,222,obs} \rightarrow [\theta]_{n,222,obs} / ([\theta]_{Test,222} - $

$[\theta]_{n,222,obs}$), exploring experimental conditions where $[\theta]_{\infty,222}$ might assume a value over the range of -44,000 to -77,000 deg cm^2 dmol^{-1} res^{-1}. These analyses generate the three different correlations seen in Figure 12. [26] Two features are evident. All three correlations are exceptionally linear; hence one does not need to know the value for $[\theta]_{\infty,222}$ in order to test the quality of the experimental t/c vs. $[\theta]_{n,222,obs}$ correlation. Moreover for a given choice of $[\theta]_{test,222}$, the t/c and $[\theta]_{n,222,calc}$ data calculated for varying w and n fall on the same line. This was not unexpected; in a LR model the temperature dependence of helicity is embodied in the varying values of w, and for each of the three data groups, a good linear correlation is seen for data calculated over a large range of w. The quality of the correlation between temperature and length dependent t/c and CD data can clearly be tested by this versatile analysis.

---

[26] If the CD data is generated and transformed with a value of $[\theta]_{\infty} = [\theta]_{Test}$ the model datasets have the same linear correction and the slope is the same for $[\theta]_{\infty} = [\theta]_{Test} = -37,000; -44,000; -61,000; -71,000$.
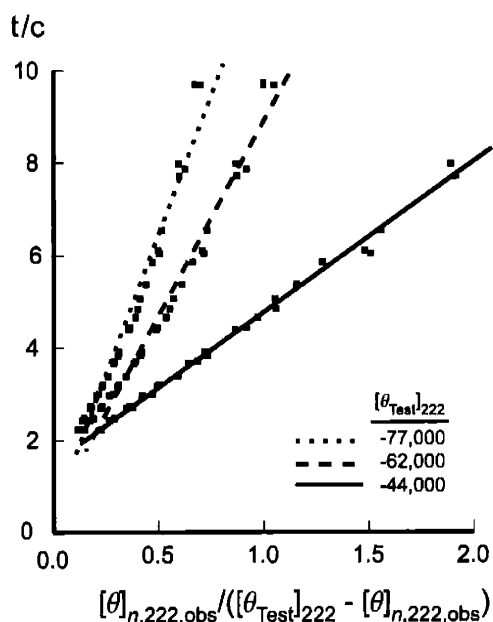
Figure 12: Plot of t/c vs. $[\theta]_{n,222,obs} \rightarrow [\theta]_{n,222,obs}$ **/ (**$[\theta]_{Test,222}$ **-** $[\theta]_{n,222,obs}$**) testing three different values for the controversial limit** $[\theta]_\infty$**. Linearity is observed for all of these values as long as the dataset is limited to the accurate range of t/c, 1 -10. Each of the three data sets incorporates a range of** w **values from 1.05 to 1.35 and helical lengths 6 - 14. Helical lengths and large** w **values that produced t/c ratios larget than 10 were not included. Large changes in slope are apparent if** $[\theta]_{test}$ **used to transform the ellipticity data is significantly larger or smaller than the actual** $[\theta]_\infty$**.**

Figure 13 plots experimental t/c values for the length series Ac-Hel-A$_n$-^{l}LInp$_2$-K$_4$-W-NH$_2$ at three temperatures as a function of the experimental CD ellipticity data that has been transformed into $[\theta]_{n,222,obs} \rightarrow [\theta]_{n,222,obs}$ / ($[\theta]_{Test,222}$ - $[\theta]_{n,222,obs}$). The value of $[\theta]_{Test}$ was set as -61,000 deg cm^2 dmol^{-1} res^{-1} consistent with other Kemp group studies of highly helical polyalanine peptides. An excellent correlation coefficient is observed CC = 0.98. The demonstrated linearity for this series of polyalanine peptides, which lack the possibility of interference from helical lysine residues, validate the t/c method and establish t/c data as equivalent to CD data as quantitative measures of helicity.
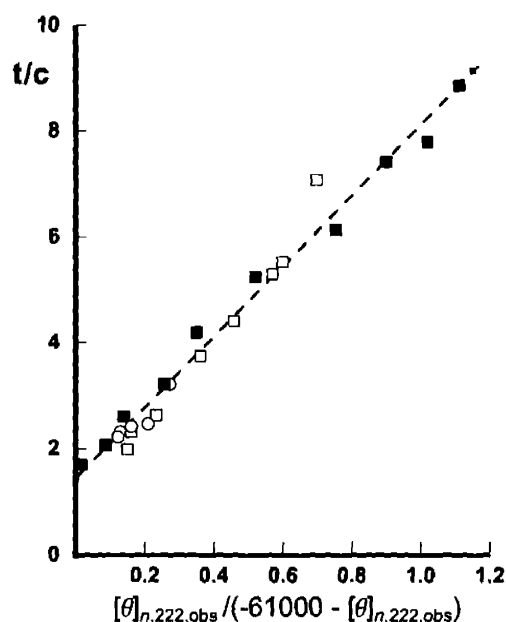
Figure 13: Comparison of experimental measurements of fractional helicity. Ratios of t/c are plotted against the experimental ellipticity at 222 nm $[\theta]_{n,222}$, transformed by $[\theta]_{n,222,obs}$ / ($[\theta]_{\infty,222}$ - $[\theta]_{n,222,obs}$). The correction to experimental ellipticity modifies the bounds to mirror the t/c bounds of 1 to infinity. The correlation at 2 (■), 25 (□), 60 (○) °C is linear indicating that both t/c and CD monitor helicity quantitatively. The experimental CD results at 2 and 25 °C can be seen in Figure 14.

This analysis can also provide independent evidence for the need of a revised value of $[\theta]_{\infty,222}$ for polyalanine peptide helices between 4 and 14 residues. The slope of the experimental data set is ca 7.0. Based on the LR modeling in Figure 12, the theoretical correlation with $[\theta]_{Test,222}$ = -62,000 deg cm^2 dmol^{-1} res^{-1} has a slope of 8.0 ± 0.5. If one varies the value of $[\theta]_{Test,222}$ such that the experimental data set is within this range, taking account of error limits, the analysis suggests that $[\theta]_{\infty}$ lies between -61,000 and -66,000 deg cm^2 dmol^{-1} res^{-1}. Although this analysis does not define the value of $[\theta]_{\infty}$ exactly, it does provide further support that the value of $[\theta]_{\infty}$ is significantly greater than -44,000 deg cm^2 dmol^{-1} res^{-1}.

As discussed in Chapter 1 both CD ellipticities and t/c ratios reflect the average properties of helical peptides. The correlation between experimental t/c and CD ellipticities as modeled by LR further demonstrates for the need to rigorously determine $[\theta]_{\infty,222}$. Independently these methods are unable to effectively recalibrate CD, due to conformational averaging. Chapter 3 presents a detailed discussion of hydrogen exchange as a method that when combined with t/c and CD has the potential to provide the needed calibration. A detailed discussion of the methodology will be presented there. Despite the uncertainty of the $[\theta]_{\infty,222}$ = -61,000 and -66,000 deg cm^2 dmol^{-1} res^{-1} suggested in the $[\theta]_{Test}$ series, does this averaged value reflect the trends seen in the CD data of Ac-Hel-A$_n$-^{l}LInp$_2$-K$_4$-W-NH$_2$?

As mentioned earlier in this chapter, the values of A, B, w and w' obtained from t/c analysis can be used with the LR algorithm to determine the fractional helicity of the peptide as one method of determining FH. These FH values can be combined with the limiting ellipticities $[\theta]_{n,222}$ in Equation 4 to predict CD ellipticities as a function of w and w'.[10,5b] Based on the value of $[\theta]_{Test,222}$ and a CD calibration standard that will be presented in Chapter 3, two values for $[\theta]_{\infty,222}$ are proposed. For shorter peptides with $n \leq 9$, the function is -53,000(1 - 2.5/n) and for longer peptide with $n > 9$ the function is $-61,000(1 - 2.5/n)$.
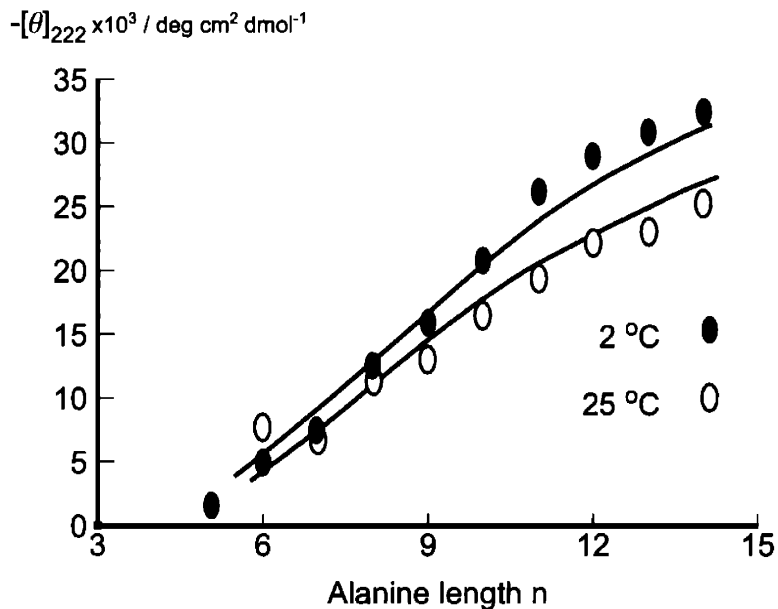
Figure 14: Experimental Circular Dichroism results for Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W. The values of -[θ]$_{n,222}$ are plotted vs. alanine length. The solid lines indicate the calculated ellipticities using the w values obtained from the a LR analysis using w values from the t/c study and the two-term ellipticity function.

The proposed increases in [θ]$_{\infty,222}$ model the experimental CD data. Figure 14 plots length dependence of experimental -[θ]$_{n,222,obs}$ data for the Ac-Hel-Ala$_n$- at 2 and 25 °C. Applying a LR analysis using w' and B calculated from t/c data with the [θ]$_{n,222}$ function described above generates the 2 and 25 °C curves shown. T he calculated curves reproduce the sigmoid features of the length dependence of the experimental [θ]$_{n,222}$ data, and differences between calculated and experimental ellipticities lie within the error of measurement and may reflect the averaged properties of t/c and CD used to generate [θ]$_{Test}$. This correlation provides final closure to the demonstration that experimental t/c and [θ]$_{n,222,obs}$ data are fully equivalent measures of helicity.

SUMMARY

This chapter presents two analytical methods for determining the helicity of peptides. Using the *IS-SO* methodology, the helical alanine region was extended to 14 residues, at which a t/c ratio of 10 was observed at 2 °C in water, the practical limit for accurate t/c measurement. This length extension falsified the previous assumption of Kemp et al. that w_{Ala} is a constant 1.0 and temperature independent between 2 and 60 °C, as observed for the Ac-Hel polyalanine series with n less than 6 residues. The value of 1.3 for w' at 2 °C in this new series is roughly halfway between the original Kemp prediction of 1.0 and Baldwin's value of 1.6. The results also show that the previous assigned apportionment of helix stabilization energy between alanine and lysine must be reapportioned to reflect the new temperature dependent value of w'. A preliminary reanalysis of w_{Lys} indicates that when Lys is located at the C-terminal region of the helix, $w_{Lys} > w_{Ala}$ consistent with the earlier Kemp work.

The t/c results were validated by the linear correlation of t/c and experimental ellipticities, and the t/c derived values of w' produced a good fit for the experimental $[\theta]_{n,222,obs}$ when the length dependent ellipticity function is supported with experimentally determined values of $[\theta]_{n,222}$. The ellipticity calibration function used in this chapter must be regarded as preliminary and provisional because both CD and t/c ratios reflect the average properties of the helical peptides. The LR modeled correlation between t/c and CD ellipticity leaves a rigorous assignment of $[\theta]_{\infty,222}$ as a remaining unresolved issue.

Chapter 1 outlined the scheme in Figure 15 where three experimental methods

and the LR algorithm and analysis are used to generate a consistent measure of peptide

helicity. This chapter has validated the direct comparison of CD and t/c. The LR

algorithm was also used to define the w values obtained from these experimental

measurements. Chapter 3 will look at how hydrogen exchange can be added to the

scheme to determine accurate fractional helicities and probe the experimental evidence to

justify the use of a length dependent w values in the LR analysis. Chapter 3 also presents

hydrogen exchange as a method that when combined with t/c and CD have the potential
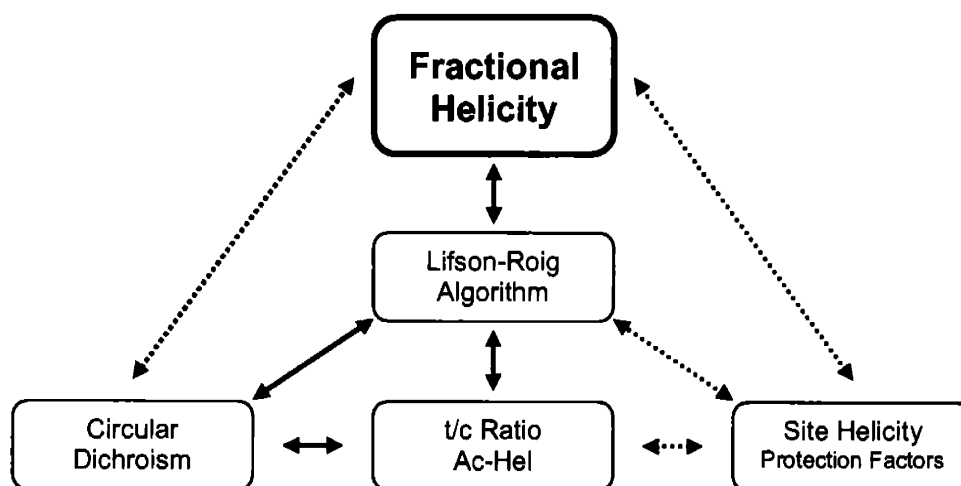
to calibrate $[\theta]_{\infty,222}$.

.

Figure 15: Solid lines indicate methods compared in this chapter. Dotted lines indicate methods that will be used in the next chapter.

CHAPTER 3: LONG POLYALANINE HELICES

INTRODUCTION

Helicity measurements at individual sites along the peptide backbone provide incisive information. Measured by hydrogen exchange utilizing isotopic labeling, the fractional site helicities provide a direct indication of the helical stability at each residue of a polypeptide in solution. In this chapter, hydrogen exchange will be used to demonstrate and quantify the exceptional helicity of the highly stabilizing N- and C-capped polyalanines and to define the length dependence of w_{Ala} values along with the distribution of helical conformers in the polyalanine regions of the spaced-solubilized (SO-IS-Ala$_n$-IS-SO) motif introduced in Chapter 1, in which the helical region is limited to the polyalanine core.

For the first time Chapter 2 analyzes the t/c values measured for polyalanines longer than six residues in the series Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W-NH$_2$ ($n = 4$ - 14) and assigns a pair of length dependent values to w_{Ala}: $w_{Ala} = 1.0$ for $n < 6$, $w_{Ala} = 1.29$ for $n > 6$ at 2 °C.[1] This length dependent change in w_{Ala}, previously undetected in early Kemp group studies[2] of -Ala$_4$Lys- peptides, is inconsistent with the Lifson-Roig assumption[3] that peptide stability is a linear function of length; no earlier studies by other groups have suggested this anomaly. Owing to the errors in determining large t/c values

---

[1] Kennedy, R. J.; Kwok, K. Y.; Kemp, D. S. "Consistent Helicities from CD and Template t/c Data for N-Templated Polyalanines: Progress Toward Resolution of the Alanine Helicity Problem" *J. Am. Chem. Soc.* **2002**, *124*, 934-944.

[2] a) Kemp, D. S.; Oslick, S. L.; Allen, T. J. "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. 3. Calculation of the Helical Propagation Constant *s* from the Template Stability Constants t/c for Ac-Hel$_1$-OH, *n* = 1-6" *J. Am. Chem. Soc.* **1996**, *118*, 4249-4255 b) Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. "For Short Alanine-Lysine Peptides the Helical Propensities of Lysine Residues (*s* Values) Are Strongly Temperature Dependent" *J. Am. Chem. Soc.* **1996**, *118*, 12234-12235

[3] a) Lifson, S.; Roig, A. "On the Theory of Helix-Coil Transition in Polypeptides" *J. Chem. Phys.* **1961**, *34*, 1963-1974; b) Qian, H.; Schellman, J. A. "Helix-Coil Theories: A Comparative Study for Finite Length Polypeptides" *J. Phys. Chem.* **1992**, *96*, 3987-3998.

with accuracy and aggregation problems with longer polyalanines, the polyalanine series

in Chapter 2 reached the experimental measurement limit of t/c at $n = 14$ and 2 °C. No

further helicity information can be obtained from the t/c method for longer peptides. The

aim of this Chapter is to refute or validate the length dependence of w_{Ala} by exploring

longer polyalanines that are unaggregated and unlike the Baldwin-Marqusee[4] -$(A_4K)_m$-

peptides, lack internal lysine solubilizing residues. Although Chapter 2 also presents

evidence of the correlation between t/c and circular dichroism, CD cannot be used to

provide an independent calibration of helicity until $[\theta]_{\infty,222}$ is rigorously defined.

The spaced-solubilized (SO-IS-Ala$_n$-IS-SO) motif[5] developed in Chapter 1 is the

system of choice and hydrogen exchange is the least assumption dependent analytical

method for studying helices formed by larger peptides. By applying this technique to the

spaced-solubilized (SO-IS-Ala$_n$-IS-SO) motif in the form of a site scan of -Ala$_{15}$-, we will

first determine if the general form of the site-dependent helicities predicted for A_{15} from

LR models match the experimental data. If the site scan is successful and models the

fraying discussed in Chapter 1, measurement of the central residue hydrogen exchange-

based protection factors for a series of polyalanines between 5 - 25 alanine residues will

investigate if the length dependent w values seen for short polyalanines in Chapter 2 are

applicable in this longer polyalanine context. These exploratory experiments have the

potential to falsify or validate the generalization of length dependent w values for alanine.

---

[4] Marqusee, S.; Robbins, V. H.; Baldwin, R. L. "Unusually Stable Helix Formation in Short Alanine-Based Peptides" *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286-5290.

[5] a) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Short, Solubilized Polyalanines Are Conformational Chameleons: Exceptionally Helical If N- and C-Capped with Helix Stabilizers, Weakly to Moderately Helical If Capped with Rigid Spacers" *Biochemistry* **2001**, *40*, 305-309; b) Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities" *J. Am. Chem. Soc.* **2002**, *124*, 945-962.

In addition, hydrogen exchange will be used to measure the site helicity for each

residue of the highly stabilized N- and C- capped polyalanines. The test peptides[6]

malonyl-Hel-A_8-β-NH_2 and Ac-$^\beta$aspartyl-Hel-A_8-β-NH_2 were originally studied and

found to be highly helical by circular dichroism. This Chapter will define the overall

fractional helicity for the octaalanine core as measured by hydrogen exchange; current

work in the Kemp group has studied a polyalanine length series of this type. This

fractional helicity value combined with circular dichroism data, has the potential to

define a series of calibration standards in the form of $[\theta]_{n,222}$ for the entire length series;

this can be used to solve for $[\theta]_{\infty,222}$ and x.

The next section will address the relevant background on hydrogen exchange that

is necessary in order to analyze and understand the experimental results.

BACKGROUND

AMIDE HYDROGEN EXCHANGE MECHANISM

The proposed exchange experiments in this chapter monitor the H to D exchange

rate of an amide hydrogen atom that belongs to the peptide backbone. The process can

be monitored for the entire peptide without assignment of exchange properties to

particular NH groups by the change in mass as detected by mass spectrometry.[7] This

chapter relies instead on the NMR methods[8] utilizing ^{15}N labeling to isolate site-specific

---

[6] β = beta-amino alanine, $^\beta$Asp ≡ beta-linked aspartate
[7] Engen, J. R.; Smith, D. L. "Investigating protein structure and dynamics by hydrogen exchange MS" *Anal. Chem.* **2001** *73*, 256A-265A.
[8] Rohl, C. A.; Baldwin, R. L. "Exchange Kinetics of Individual Amide Protons in ^{15}N-Labeled Helical Peptides Measured by Isotope-Edited NMR" *Biochemistry* **1994**, *33*, 7760-7767

hydrogen atoms from the other amide and non-amide hydrogens of the peptide. These

exchange rates are expected to correlate with site helicities.
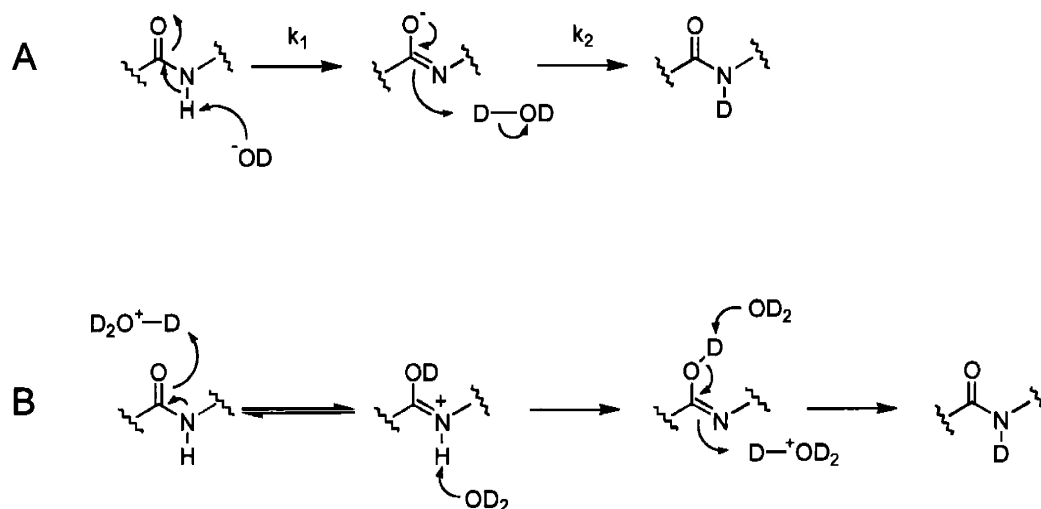
Figure 1: Hydrogen exchange mechanisms A) base catalysis B) acid catalysis. Both replace a backbone amide hydrogen with a deuterium atom. For the base catalyzed mechanism, the first step is rate determining and the overall rate equation is -d[NH]/dt = k₁ [NH][OH⁻]

There are two mechanisms by which hydrogen exchange can take place. In both

cases the amide hydrogen atom is replaced with a deuterium atom from the solvent, as

illustrated in Figure 1. The base catalyzed mechanism requires that of the amide

functionality only its NH be solvent accessible for replacement. The acid catalyzed

mechanism involves protonation of the oxygen carbonyl prior to replacement, thus

requiring that both ends of the amide be solvent exposed. Englander and coworkers

studied the pH and temperature dependence of these processes on short alanine di- and

tri- peptides and longer poly-D-L-alanines.[9] The pH profiles measured as a part if our

[9] Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. "Primary Structure Effects on Peptide Group Hydrogen Exchange" _Proteins: Struct., Funct., Genet._ **1993**, _17_, 75-86.

studies are shown in Figure 2; these are similar to those reported by the Englander group.[10] A large difference in the rate constants of the two mechanisms, with base catalysis being more efficient, results in a pH minimum between 3.0 and 4.0; in general at pH's below 3.5 the acid catalysis is dominant and above pH 3.5 the base catalysis is dominant. Due to experimental error, it is not possible to exclude D_2O catalysis close the pH minimum. Englander also failed to detect general catalysis from the buffer. To avoid structural ambiguities that attend the acid catalyzed mechanism, the studies in this thesis were set up to monitor exchange under basic conditions carried out in the pH region > 3.5.

---

[10] Englander[9] calculated $logk_1$ under basic conditions for : N-Ac-Ala-N'MA $logk_1 = 9.71$ (L), 9.01 (R) and poly-D-L-Alanine $logk_1 = 9.90$ at 5 °C in 0.5M KCl Units: M^{-1} min^{-1}. The solvent exposed rate constant, $logk_1 = 9.12$ M^{-1} min^{-1} calculated from the data at pH 3.85 in Figure 2 is within the range of Englanders results.
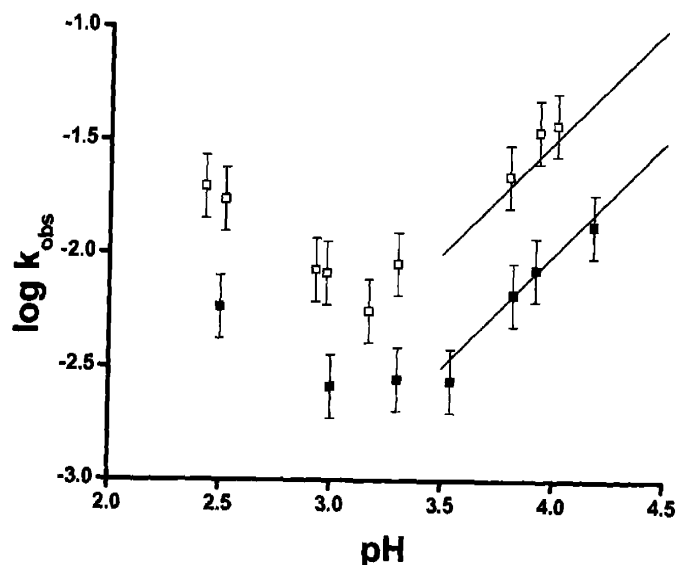
Figure 2: Site-specific hydrogen exchange pH profiles for the central residues of $WK_5\text{-}Inp_2'L\text{-}A_n\text{-}'LInp_2\text{-}K_5\text{-}NH_2$ measured by HSQC NMR to isolate the ^{15}N labeling site. ($\square$) $n = 5$: ^{15}N labeled Ala at site 3 ($\blacksquare$) $n = 15$: ^{15}N labeled Ala at sites 10, 11, 12. Trend lines show the expected first order slope of 1.0 on the side of basic catalysis. The exchange was measured at 2 °C in 20 mM $NaH_2PO_4/$ D_2O. pH is measured using a glass electrode at room temperature; no correction is made for deuterium or temperature. A 5- 10% error in k_{obs} was observed between trials. k_{obs} is the pseudo-first order rate constant.
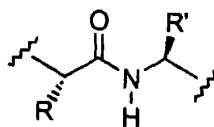
Figure 3: Side chains R and R′ adjacent to the amide NH influence the rate of exchange based.

In addition to pH, as cited above, temperature and the hydrophobic bulk and the electronic properties of the adjacent amino acid side chains also influence the rate of hydrogen exchange. Englander[9] has studied these effects and created a database of reference rates for the amino acids side chains in short di- and tri-peptides that do not

adopt secondary structure; these are required as reference standards for analysis of exchange data for heteropeptides that do adopt secondary structures that contain solvent shielded amide NH groups. Within WK_5-Inp_2-^lL-A_n-$^lLInp_2$-K_5-NH_2, the N-terminal alanine that is adjacent to a ^lL residue may be influenced by its large hydrophobic bulk. Englander's corrections show that under basic conditions the solvent exposed rate constants in comparison to alanine were reduced by 14% for valine, and similarly a reduction of 21% and 23% were observed for leucine and isoleucine. The decrease in rate would produce an anomalously high protection factor if the proper reference rate was not used. With the exception of the alanine that is adjacent to the ^lL residue, the analysis of experimental data in this chapter does not require the use of these correction factors, as the entire helical peptide sequence is a polyalanine. If Englander's assumptions are correct the only reference rate constants required for the data analysis of this thesis are those of short polyalanines that are unable to form helices. In order to maximize the helicity of the polyalanines, where based on the w values the protection factors are expected to be small, and to maintain experimentally measurable rates all the experiments were conducted at 2 °C.

SITE HELICITY

How does hydrogen exchange provide a quantitative measurement of helical secondary structure? The formation of hydrogen bonds in α-helices is assumed, based on protein[11] and other models, to completely prevent the exchange of amide NHs. As illustrated in Figure 4, the exchange is predicted to be exceptionally slow for an amide

---

[11] Goodman, E. M.; Kim, P. S. "Periodicity of Amide Proton Exchange Rates in a Coiled-Coil Leucine Zipper Peptide" *Biochemistry* **1991**, *30*, 11615-11620.

NH that is fully hydrogen-bonded within a helical peptide, but occurs at fast rates for solvent exposed N-terminal amide NHs.
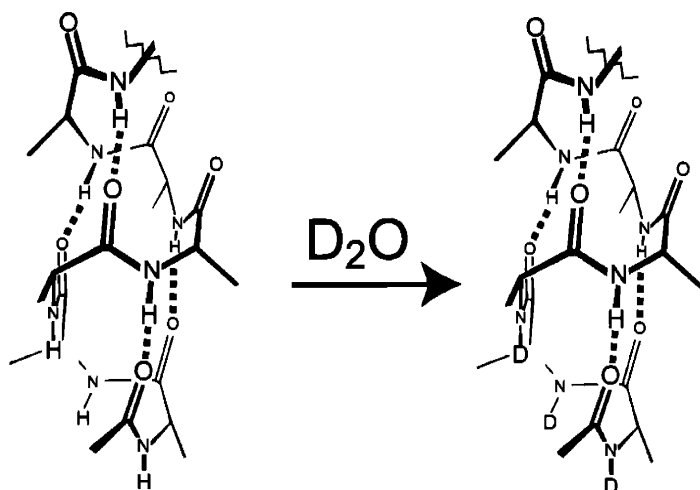
Figure 4: Structural representation of hydrogen exchange, showing the rapid exchange of non-hydrogen bonded amides and the slow exchange of amide NH's hydrogen bonded within the helix. A fully helical conformation that is assumed by a peptide of *n* residues that is N- and C-capped by amides is expected to have 3 solvent exposed NHs at its N-terminus and *n* - 2 hydrogen bonded amide NHs; one of these is formed by the C-terminal residue or NH₂ cap.

Experimental evidence shows that in the native structures of proteins, hydrogen exchange is very slow, with half lives of days to weeks, due to the strong cooperative folding that maintains a single compact tertiary structure. By contrast, the exchange of denatured proteins is fast.[12] Unaggregated peptides in solution may adopt many helical conformers that have residue sites in equilibrium between the slowly exchanging helical (h) conformations and nonhelical (c) conformations where exchange is much faster. The average solvent accessibility of a backbone amide NH will change, based on the magnitude of the equilibrium constant w defined between the hydrogen bonded and solvent exposed states. Qualitatively, slower exchange indicates that the residue is, on

---

[12] Englander, S. W.; Kallenbach, N. R. "Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids" *Q. Rev. Biophys.* **1984**, *16*, 521-655.

average, more helical. Comparing the rate of exchange between sites along the peptide

backbone allows one to identify local helical sequences that are protected from solvent

and that correspond to highly helical regions, as well as regions that are poorly structured

and largely solvent exposed.
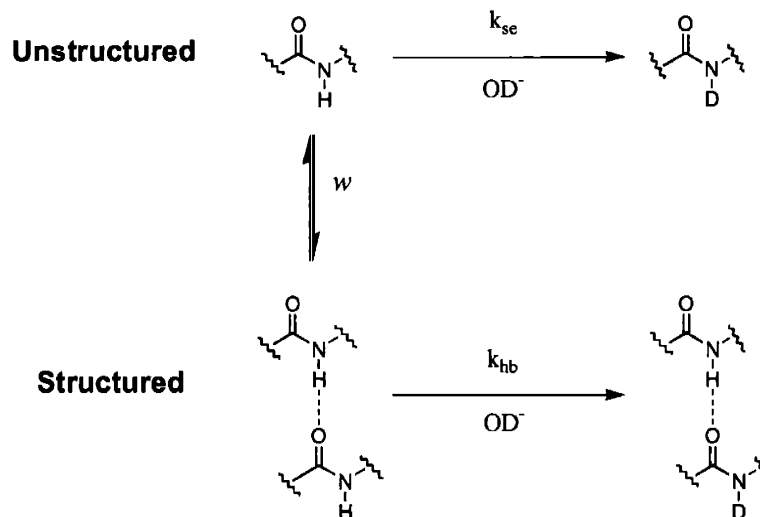
Figure 5: Effect of hydrogen bonding on exchange rate for an amide NH in a single helical conformation. The equilibrium constant w governs the formation of helices and hydrogen bonds. The rate constant k_{se} corresponds to the solvent exposed amide and k_{hb} corresponds to hydrogen bonded amide NH.

How does the exchange rate correspond quantitatively to the measurement of

helicity at a specific site i? Under the mechanism illustrated in Figure 5, the exchange of

the amide NH is dependent on the rate constants (k) for the solvent exposed (k_{se}) and

hydrogen bonded (k_{hb}) states along with the corresponding amide concentrations ($NH_{i,se}$,

$NH_{i,hb}$), Equation 1a. As noted above, the rate constant for exchange of the hydrogen

bonded state approaches zero, Equation 1b. In the base catalyzed pH region, the

experimentally observed pseudo-first order rate constant (k_{obs}) is known to be first order

with respect to the hydroxide concentration and the total amide concentration ($NH_{i,T}$).

Based on the equality of Equation 1c, the experimentally observed rate constant at site i is equal to the product of the solvent exposed rate constant and the mole fraction of solvent exposed amide, Equation 1e.

a) $\quad \dfrac{d[NH_{i,T}]}{dt} = -k_{se}\,[OH^-][NH_{i,se}] - k_{hb}\,[OH^-][NH_{i,hb}]$

b) $\quad k_{hb} \cong 0$

c) $\quad \dfrac{d[NH_{i,T}]}{dt} = -k_{se}\,[OH^-][NH_{i,se}] = -k_{obs}\,[OH^-][NH_{i,T}]$

d) $\quad k_{obs} = k_{se}\,\dfrac{[NH_{i,se}]}{[NH_{i,T}]}$

e) $\quad \chi_{NHi,se} = \dfrac{[NH_{i,se}]}{[NH_{i,T}]}$

Equation 1: Rate equations for amide hydrogen exchange in the pH region where base catalysis dominates. a) The overall exchange rate of an amide NH at site i is a sum of the rate terms and the concentrations of the solvent exposed amide ($NH_{i,se}$) and the hydrogen bonded amide ($NH_{i,hb}$) conformers. b) Under the assumption that $k_{hb} \cong 0$, the second term of a) can be neglected. c, d, e) The experimentally observed rate constant k_{obs} is proportional to k_{se} times the mole fraction of solvent exposed amide NH protons at site i, $\chi_{NHi,se}$.

The ratio of the experimentally observed exchange rate constant for a particular helical amide NH (k_{obs}) and the exchange rate constant for the same residue that is known to be fully solvent exposed (k_{se}) is defined as the protection factor (PF_i) at amide NH_i. Experimentally, the only way to measure k_{se} is to use a short peptide sequence that is unable to form helical hydrogen bonds, such that the total amide concentration ($NH_{i,T}$) is equal to the solvent exposed concentration ($NH_{i,se}$); k_{se} is assumed to be length

independent. The ratio of the intrinsic k_{se} and k_{obs} is then by definition inversely proportional to the mole fraction of conformations that do not have hydrogen bonding at site i ($\chi_{NHi,se}$). As shown in Equation 2, one can also solve for the mole fractions of conformations that have helical hydrogen bonding at site i ($\chi_{NHi,hb}$). Protection factors range from 1.0, if no helix is present, to infinity if the site is 100% helical.

$$PF_i = \frac{k_{se}}{k_{obs}} = \frac{[NH_{i,T}]}{[NH_{i,se}]} = \frac{1}{\chi_{NHi,se}} = \frac{1}{1-(\chi_{NHi,hb})}$$

Equation 2: A protection factor (PF) is defined as the ratio of the experimental rate constants for the partially helical amide NH at site i amide (k_{obs}) and a model non-helical amide NH that is free to exchange with solvent (k_{se}). Based on the model illustrated in Equation 1 PF$_i$ is equal to the inverse of the mole fraction of amide NH that are solvent exposed ($\chi_{NHi, se}$) or the inverse of the quantity unity minus the mole fraction that does have α-helical hydrogen bonding ($\chi_{NHi, hb}$).
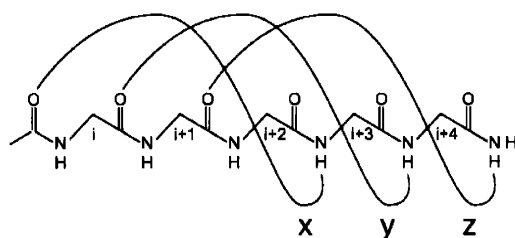
Figure 6: Hydrogen bonding scheme for a pentapeptide. The α-helical hydrogen bonding pattern requires that at least three preceeding residues adopt helical ϕ, ψ angles to achieve the definitive i to $i + 4$ hydrogen bonding pattern.

It is important to notice that site helicities defined at each α-carbon ($\chi_{\alpha C_i, H}$) differ from the mole fraction of amides with hydrogen bonds ($\chi_{NH_i, hb}$). In the Lifson Roig (LR) algorithm, where h and c identifiers are assigned per residue based on ϕ, ψ angles, all conformations where at site i an h appears and is part of helical sequence of three or more residues are included in a calculation of $\chi_{\alpha C_i, H}$. How does calculation of the mole fraction of sites that have hydrogen bonding ($\chi_{NH_i, hb}$) compare with site fractional helicity ($\chi_{H, \alpha C_i}$) as defined by the helical ϕ, ψ of a single residue? As illustrated in Figure 4, although the first three residues have helical ϕ, ψ angles, their amide NHs exchange rapidly due to the lack of hydrogen bonds hence, $\chi_{NH_i, hb} = 0$. The α-helical hydrogen bonding pattern requires that at least three preceeding residues adopt helical ϕ, ψ angles for a hydrogen bond to form. As illustrated in Figure 6, for the fifth residue ($i + 4$) to be protected, all of the ($i + 1$), ($i + 2$), ($i + 3$) sites must be helical. Note that site ($i + 4$) does not have to be helical for its amide NH to form a hydrogen bond. An appropriate LR

algorithm[13] can be constructed to accommodate this more complex identifier prerequisite for a calculation of $\chi_{NHi,hb}$.

Part A

	helical conformers					α-carbon site weights				
	1	2	3	4	5	1	2	3	4	5
A	h	h	h	h	h	+	+	+	+	+
B	h	h	h	h	c	+	+	+	+	-
C	c	h	h	h	h	-	+	+	+	+
D	h	h	h	c	c	+	+	+	-	-
D'	h	h	h	c	h	+	+	+	-	-
E	c	h	h	h	c	-	+	+	+	-
F	c	c	h	h	h	-	-	+	+	+
F'	h	c	h	h	h	-	-	+	+	+
Helical Sites Counted	4	6	8	6	4					

Part B

	helical conformers						PF site weights		
	1	2	3	4	5	(6)	4	5	(6)
A	h	h	h	h_x	h_y	h_z	+	+	+
B	h	h	h	h_x	c_y		+	+	-
C	c	h	h	h	h_y	h_z	-	+	+
D	h	h	h	c_x	c		+	-	-
D'	h	h	h	c_x	h		+	-	-
E	c	h	h	h	c_y		-	+	-
F	c	c	h	h	h_z		-	-	+
F'	h	c	h	h	h_z		-	-	+
							4	4	4

Figure 7: Comparison for a pentapeptide of the helical conformations included in the calculations of fractional site helicity ($\chi_{\alpha Ci,H}$ part A) and the mole fraction of sites with hydrogen bonds ($\chi_{NHi,hb}$ part B). Part A: a plus sign indicates each helical site as designated by the h identifier. Part B: a plus sign indicates each residue that is preceded by three helical residues establishing the hydrogen bonds labeled x, y and z in Figure 6. Site 6 presents the terminal amide NH_2, residues must be capped with amides to be able to adopt helical ϕ, ψ angles; Note from Figure 6 that both N-acetyl and terminal NH_2 are involved in hydrogen bonding.

Figure 7 illustrates and contrasts states that are included in the calculation of fractional site helicity ($\chi_{\alpha Ci,H}$) and states that are included in the mole fraction of sites protected by hydrogen bonding ($\chi_{NHi, hb}$). In this pentapeptide example, the fractional site helicity of the fifth residue would include the weight for all conformations that have an h in site 5: A, C, D', F and F'. To calculate the mole fraction of the amide NH at site 5 that

---

[13] Kemp, D. S. "Construction and Analysis of Lifson-Roig Models for the Helical Conformations of α-Peptides" *Helv. Chim. Acta.* **2002**, *85*, 4392-4423.

is protected from exchange conformations: B and E are added to the count along with A and C while F and F' are not counted because these helical conformations do not extend three residues toward the N-terminus. Similar comparisons can be made for the sites 4 and 6. As required, the first three hydrogen atoms are not protected (PF = 1.0).

Figure 8A illustrates the difference between the mole fractions calculated by the LR algorithms for fractional site helicity ($\chi_{\alpha C i, H}$) and sites with hydrogen bonding ($\chi_{NH i, hb}$). Both site measurements exhibit curved site dependence reflect the fraying discussed in Chapter 1 where the ends of the potentially helical peptide sequence are expected to be less helical. Figure 8B illustrates what happens when $\chi_{\alpha C i, H}$ is used to approximate a protection factor. Due to the difference in sets of helical conformers used to generate $\chi_{\alpha C i, H}$ and $\chi_{NH i, hb}$ the data sets are qualitatively similar but not quantitatively identical. The LR algorithm should be used to calculate fractional helicity based on $\chi_{NH i, hb}$ obtained from experimental protection factor data.

If the LR algorithm successfully models the experimental protection factor data one would expect to see the following trends. 1) The maximum protection factor is displaced by two residues toward the C-terminus from the central residue. 2) A symmetrical decrease in protection factor is observed between the site of maximum protection and the N- and C-termini.

A   B
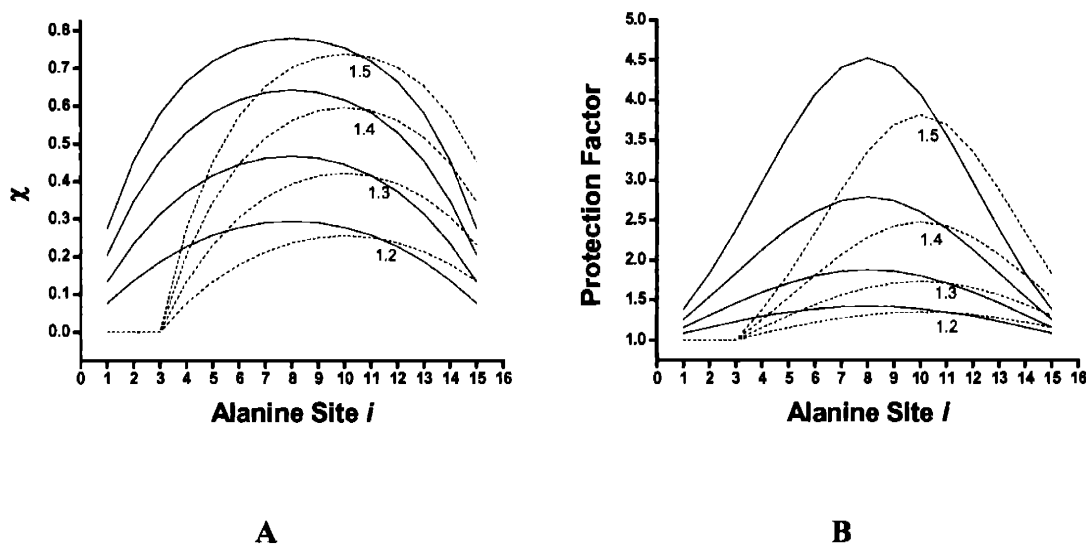
Figure 8: Dependence of site measurements on w based on the LR algorithm (v = 0.048). A) Comparison of fractional site helicity $\chi_{\alpha Ci, H}$ (solid) and mole fraction of sites with hydrogen bonding $\chi_{NHi, hb}$ (dotted). B) PF data as calculated from $\chi_{\alpha Ci, H}$ and $\chi_{NHi, hb}$. The w values are placed at the intersection of the corresponding curves.

PREVIOUS STUDIES

Rohl, Baldwin and coworkers have reported results for a series of experiments that utilized hydrogen exchange to quantitate the helicity of alanine rich peptide containing lysine residues for solubilization. The first study in 1992 looked at a length series Ac-(AAKAA)$_m$Y-NH$_2$ (m = 1-10).[14] Poor dispersion of the amide ^{1}H signals forced the analysis to consider the fractional occupancy of the peptide backbone as whole as measured by acid catalyzed hydrogen exchange at pH 2.5. The study concluded that the average w, for the peptide treated as a homopolymer, is 1.58. A second study utilized ^{15}N site labeling to isolate the individual ^{1}H NMR signal of a single site, measured both acid and base catalyzed exchange rates for Ac-(AAKAA)$_4$-Y-NH$_2$ and concluded[8] that

---

[14] Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. "Kinetics of Amide Proton Exchange in Helical Peptides of Varying Chain Lengths. Interpretation by the Lifson-Roig Equation" *Biochemistry* **1992**, *31*, 1263-1299.

base catalyzed exchange rates measure the extent of hydrogen bonding. Although this thorough study indicates that the LR algorithm is able to model the protection factors, systematic deviations from the predicted protection factors were observed for alanine residues centered near the lysine residues sites.

The experiments with homopeptides in this chapter are designed to gather similar information, but are not designed to support or refute the Baldwin results. A direct comparison will have to wait until the influences of both alanine and lysine are better understood. The two purposes of these experiments are to refute or verify the length dependent w values seen in Chapter 2 for Ac-Hel-A$_n$-^{t}LInp$_2$-K$_4$-W-NH$_2$ (n = 4 - 14) and explore a larger range of lengths. One of the benefits of using the sequence WK$_5$-Inp$_2$^{t}L-A$_n$-^{t}LInp$_2$-K$_5$-NH$_2$ is that there is no need for a complex analysis of the differing influence of the alanine and lysine side chains on the PF of neighboring residues. Relying on similar methodologies to the Baldwin studies, ^{15}N amide labeling is used here to measure single residue sites, and to simplify the structural interpretations hydrogen exchange will be conducted under basic conditions.

RESULTS: A$_{15}$ SITE SCAN

By applying hydrogen exchange to the spaced-solubilized (*SO-IS*-Ala$_n$-*IS-SO*) motif in the form of a site scan of -A$_{15}$-, we will first determine if the general form of the site-dependent helicities predicted for -A$_{15}$- from LR models matches the experimental data for polyalanines. As shown in the LR model in Figure 8B the maximum protection factor is not predicated to appear at the center of the helical sequence, but rather 2 residues off-center toward the C-terminus. In order to verify this prediction, ten peptides were synthesized with single or multiple ^{15}N isotopically labeled alanines in the sequence WK$_5$-Inp$_2$'L-A$_{15}$-'LInp$_2$-K$_5$-NH$_2$ and hydrogen exchange rate constants were measured for individual sites along an -A$_{15}$- sequence at pH 3.5 - 4.0. The rate constants were corrected to pH 4.0 and protection factors were calculated using Equation 2. The three sites that were predicted by LR to have the maximum protection factors were labeled using three isotopes in the same peptide; these three residues were found to have the same chemical shift.[15] The results are shown in Figure 9. The reference rate for a solvent exposed peptide was measured using a single isotope at the central alanine of WK$_5$-Inp$_2$'L-A$_5$-'LInp$_2$-K$_5$-NH$_2$ which has a CD signal consistent with an unstructured peptide.

---

[15] Unless the rate constants of the three residues differ my more than 25%, it is difficult to solve for three independent rate constants from a single NMR signal; otherwise the average rate constant results. The LR algorithm does not predict such a large change in PF between the three central sites, and no fit deviations were noted when the data were analyzed.
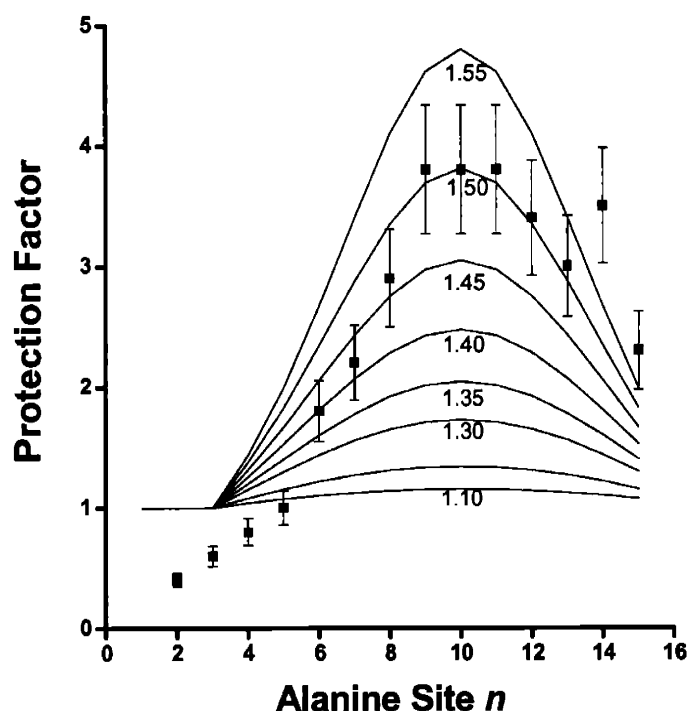
Figure 9: Protection factor site scan of WK$_5$-Inp$_2$'L-A$_{15}$-'LInp$_2$-K$_5$-NH$_2$ at pH 4.0. **The experimental points are plotted with the corresponding error bars representing a 14% error. Shown as solid lines, the dependence of χ_{NH} and PF on w as modeled by an standard LR analysis ($v = 0.048$). The exchange was measured at 2 °C in 20 mM NaH$_2$PO$_4$/ D$_2$O. pH is measured using a glass electrode at room temperature; no correction is made for deuterium or temperature. A 5- 10% error in k_{obs} and k_{se} was observed between trials. The error bars of 14% represent a maximum bound on the error in PF.**

The protection factor results for A$_{15}$ follow the general shape of the LR predictions. Although no single w value models the site dependence, there is an incremental decrease in protection factor from a central maximum at both the N- and C-terminus. Site 14 is a single anomaly that does not follow the trend, with a significantly higher protection than its neighboring residues; further investigation into this site is needed to determine if a rate modifying structural feature of this highly protected site can be assigned. Unique to the N-terminus, sites 2 - 4 exchange more quickly than the

nonhelical solvent exposed penta-alanin, hence PF < 1.0. Baldwin and coworkers observed a similar trend at the N-terminus.[8] Our analysis of the rapid exchange of these sites will be detailed in the discussion section. The LR algorithm does accurately predict the site of maximal protection and within error the protection factor values at the C-terminal region follows the predicted trend with $w = 1.55$.

RESULTS: CENTRAL RESIDUE A_n LENGTH SERIES

The $-A_{15}-$ site scan is successful reflecting the fraying discussed in Chapter 1 and corroborated the LR prediction of the site of maximum protection factor. Yet a length series is more sensitive to a length dependent w values than trying to fit cooperative w values to a site scan due to the inherent error in PF measurement. The measurement of the central residue hydrogen exchange-based protection factors for a series of polyalanines between 5-25 alanine residues will investigate if the length dependent w values seen for short polyalanines in Chapter 2 are applicable in this longer polyalanine context.

Based on the LR prediction, confirmed by the agreement seen in the A_{15} site scan, a series of polyalanines were synthesized with ^{15}N at the three contiguous sites of predicted maximum protection for the length series $WK_5\text{-}Inp_2'L\text{-}A_n\text{-}'LInp_2\text{-}K_5\text{-}NH_2$ ($n = 5\text{-}25$) for example sites 9, 10 and 11 were labeled in $-A_{15}-$. The LR analysis does not predict a significant change at these positions. A comparison between the standard LR prediction factors calculated based on $\chi_{NHi, hb}$ and the experimental protection factor data are shown in Figure 10ABC.

A

B

C
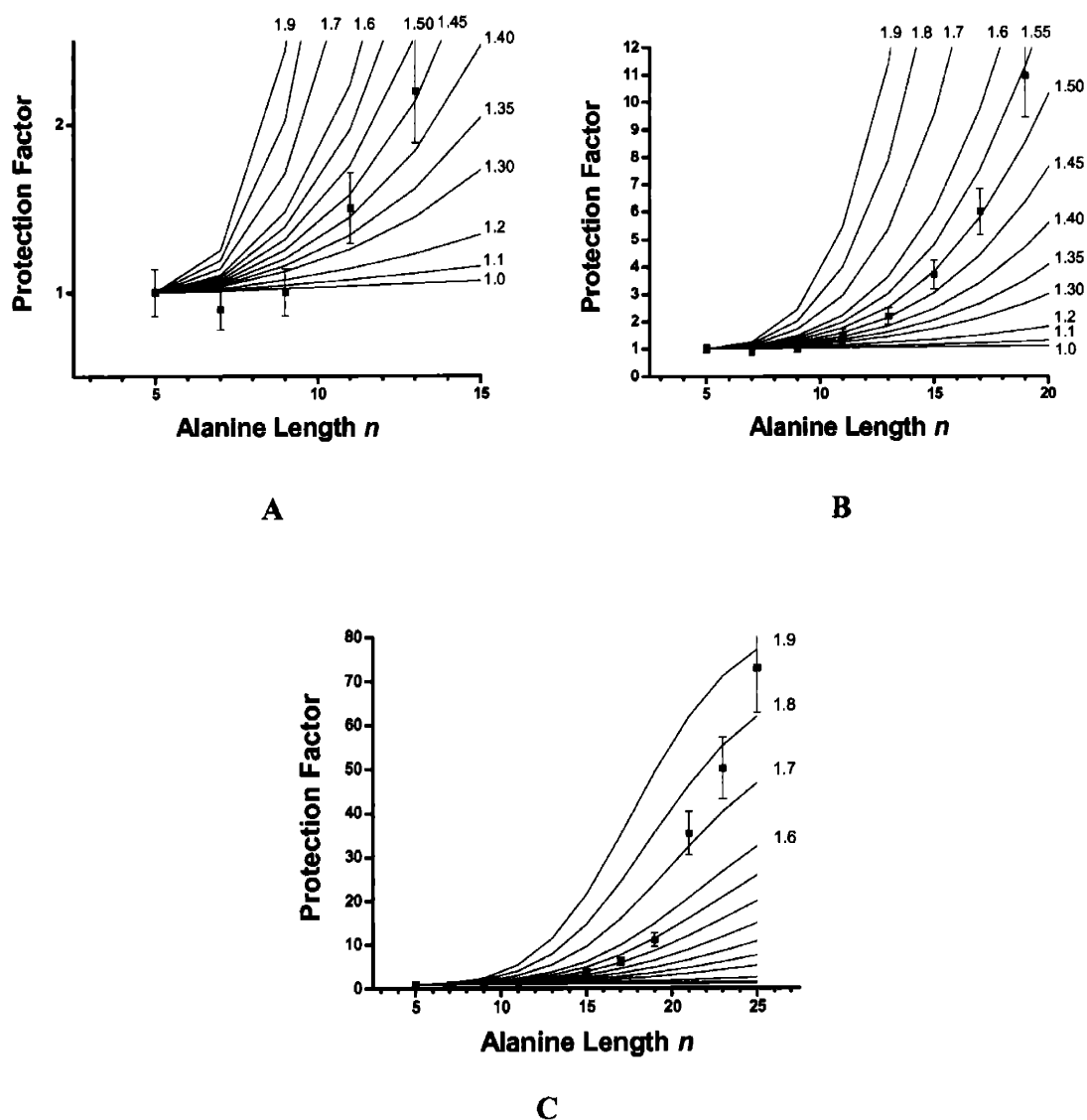
Figure 10: Protection Factor Data for the series $WK_5\text{-}Inp_2{}^tL\text{-}A_n\text{-}{}^tLInp_2\text{-}K_5\text{-}NH_2$ measured at the three residues predicted by LR to exhibit the maximum protection factor. The exchange was measured at 2 °C in 20 mM $NaH_2PO_4/$ D_2O and extrapolated to pH 4.0. pH is measured using a glass electrode at room temperature; no correction is made for deuterium or temperature. A 5- 10% error in k_{obs} was observed between trials, the 14% error bars represent a maximum bound on the error.

Figure 10A plots the first five points with odd lengths n = 5 - 13.   Overall no single w value predicts the behavior of short polyalanine peptides. The central residues

of polyalanines with $n < 10$ show little to no protection from exchange. Within error, these short polyalanines have w values of 1.0 to 1.1 consistent with short Ac-Hel polyalanines observed in Chapter 2 and previous Kemp group studies.[2] Increased protection is observed for lengths 11 and 13 that correspond to w values between 1.4 and 1.5.

The protection factors and w_{Ala} values needed to model the behavior continue to increase with length. As illustrated in Figure 10B, polyalanines with 15, 17 and 19 residues show moderate protection and w values between 1.40 and 1.55. An increase is observed for polyalanine greater than 19 residues. Polyalanines with lengths of 21, 23 and 25 residues are highly protected. The corresponding w values are between 1.7 and 1.9. These values are dramatically larger than those assigned previously. In order to rule out aggregation as a cause of the slow exchange, analytical ultra-centrifugation was used to verify that NMR sample used was monomeric, these results are reported in the experimental section, Chapter 4.

DISCUSSION

It is apparent that that no single value of w can account for the increasing helicity observed as PF increases with polyalanines length. The length dependent w values observed for short polyalanine in the series Ac-Hel-A_n-$^{'}LInp_2$-K_4-W-NH_2 ($n = 4$ - 14) continue to increase with longer polyalanine sequences WK_5-$Inp_2$$^{'}L$-$A_n$-$^{'}LInp_2$-$K_5$-$NH_2$ ($n = 5$ - 25). Length dependent w values were also assigned by Justin Miller to explain the circular dichroism results for the same series, although both $w(k)$ and the limiting ellipticity were used as free variables.[5b]

These findings require construction of a new type of LR algorithm in which each conformation of length k within the manifold of helical conformers with lengths $k < n$ is modeled using a different value of the length-dependent function $w(k)$, which must be assigned recursively.[1] A LR analysis can be used to sort the stated sum for each peptide length based on the length of the conformers and assign individual $w(k)$ values.

Prior to the recursive fitting, in order to obtain a w value that incrementally increases with length, that PF data set was smoothed using the least squares polynomial, $\ln(\text{PF}) = 0.00627n^2 + 0.00625n - 1.5$, for ($n = 9 - 25$). Figure 11 illustrates the good fit between the experimental data and this fitting polynomial. The error in the PF measurements for lengths 21, 23 and 25 may be larger than 14% as instrument time limited the range of decay to 1 - 2 half-lives instead of the standard 5-6.

A recursion was initiated using the length-dependent w values for short polyalanines as assigned by the t/c analysis of Chapter 2. These adequately modeled the smoothed central PF values for lengths $n = 9, 11$. In the first recursion step of the PF analysis, the smoothed central PF of length $n = 13$, was fitted by varying $w(12) = w(13)$ used to calculate weighting terms of the state sum for conformations with helical regions of lengths 12 and 13. This fitting process is continued throughout the length series. In each remaining recursion step that weights a smoothed central PF of length n a new state sum is constructed that employs $w(k)$ values $k < (n - 1)$ that have been assigned in previous recursion steps, but assigns a single new $w(n)$ for both helical conformers n and $n - 1$. The complete recursion generates a state sum that can be used to model PF values for any peptide in the length series. The results of the recursive analysis are listed in Table 1, and the corresponding fit to the experimental data is illustrated in Figure 12.

Figure 11: ln (Protection Factor) vs. n: The experimental data points ln(PF) are plotted. The line shows the reasonable fit is observed to the polynomial ln(PF) = $0.00627n^2$ + $0.00625n$ -1.5, for ($n = 9$ - 25). The error in the PF measurements for lengths 21, 23 and 25 may be larger than 0.14 as instrument time limited the range of decay to 1 - 2 half-lives instead of the standard 5-6.

Polyalanine Length n	w value
1-4	1.0
5	1.12
6	1.26
7	1.43
8, 9	1.46
10, 11	1.48
12, 13	1.49
14, 15	1.50
16, 17	1.51
18, 19	1.55
20, 21	1.58
22, 23	1.63
24, 25	1.69

Table 1: Table of length dependent w values from the recursive analysis of the smoothed data set.

Figure 12: LR modeling of protection factors using the length dependent w values from Table 1.

Is there any literature justification for a length dependent w value for polyalanine helical conformers of increasing length? Gou and Karplus[16] have modeled strings of amides using *ab initio* calculations and noticed that a hydrogen bond between amides at the string center is 2.9 time as strong as the hydrogen bond in a amide dimer, indicating an increase in hydrogen bond strength with an increase in numbers of consecutive hydrogen bonds. This hydrogen bond strengthening is often described as hydrogen bonding cooperativity. Dannenberg has modeled peptides in helices using molecular orbital[17] calculations and found that hydrogen bonding cooperativity is an essential energetic feature. As discussed in Chapter 1, a helix has three continuous strings of

---

[16] Gou, H.; Karplus, M. " Solvent Influence on the Stability of Peptide Hydrogen Bond: A Supramolecular Cooperative Effect" *J. Phys. Chem.* **1994**, *98*, 7104-5105.

[17] Kobko, N.; Dannenberg, J. J. "Cooperativity in Amide Hydrogen Bonding Chains. Relation between Energy, Position, and H-Bond Chain Length in Peptide and Protein Folding Models" *J. Phys. Chem. A* **2003**, *107*, 10389-10395; Wieczorek, R.; Dannenberg, J. J. "Hydrogen-Bond Cooperativity, Vibrational Coupling, and Dependence of Helix Stability on Changes in Amino Acid Sequence in Small 3_{10}-Helical Peptides. A Density Functional Theory Study" *J. Am. Chem. Soc.* **2003**, *125*, 14065-14071.

hydrogen bonds that connect i and $i + 4$ residues on adjacent loops. Figure 13A illustrates the beginning of the first string of hydrogen bonds in the helical hexamer. In a nonamer conformation the string is extended by another hydrogen bond; Gou and Karplus along with others observed a strengthening of the central hydrogen bond in the chain.[16] As the polyalanine chain lengthens the longer helical conformers would be expected to become more stable under the cooperativity assumption. The stability of longer helices is modeled in the LR algorithm by an increase in w_{Ala}.

Figure 13: Hydrogen bonding cooperativity for a hex-alanine and nona-alanine. As a single continuous chain, marked as bold, grows from a six to a nine residue conformer, the strength on the interior bond has been shown to increase by *ab initio*[16] and molecular orbital[18] calculations.

Does the implementation of the length dependent LR model account for the discrepancies between the experimental and predicated PF values? As a check to validate or discount the cooperative w values, a comparison of the experimental protection factors observed at all of the sites in $-A_{15}-$ and the values predicted from an LR analysis implementing the length dependent w values was made. Figure 14A illustrates the differences between what the LR model predicts using the length dependent w values and the experimental data. The central residues and the C-terminal region are modeled well. Yet PF values in the N-terminal region are consistently overestimated by both standard

and length-dependent LR algorithms. As discussed earlier, three amide NHs at the N-terminus of the completely helical conformer are completely solvent exposed and available for exchange, despite the fact that the backbone adopts helical ϕ, ψ values for conformations in this region. The experimental protection factors of less than 1.0 at sites 2 and 3 indicate that these sites are exchanging faster than the solvent exposed nonhelical pentaalanine reference. This result suggests that there may be a structural cause of this increase, since a nonhelical region is expected to match the solvent exposed rate constant observed for the short peptide sequence.

A   B

Figure 14: Cooperative fit for A_{15} site scan. A) The LR algorithm used to model the A_{15} site scan incorporated the length dependent w values from Table 1 and the N- and C-terminal capping values for $^tL.^{5b}$ B) Two models offer a possible explanation of the fast exchange seen at the N-terminus. The top of the bars represent the PF where k_{se} has been corrected for the mole fraction of the first three N-terminal sites at each site. The opaque bars include helical conformers with length $k > 9$, the dashed bars $k > 6$, in the calculation of the mole fraction of fast exchanging protons.

Are there plausible reasons why the rate constants might increase for amide NHs at the N-terminus of a helical region? The carbonyls of these first three amides participate in helical hydrogen bonding, and the localized partial positive charge of the helix dipole or the cooperative stabilization of the helix may increase this rate constant for base-catalyzed hydrogen exchange. This property of amide NH sites at the peptide N-terminus is shared with sites of all conformers with lengths $k < n$ for which solvent-exposed amide NHs belong to the first three residues of a helical region. Referring to Equations 1 and 2, one sees that an accurate calculation of PF values for helices appears to require a replacement of the single rate constant k_{se} for solvent exposed amide NH functions by two rate constants: k_{nh}, which equals the rate constant observed for the pentapeptide standard, and a larger rate constant k_f.

Within any conformer and a particular NH site, there are three possible types of amide NHs: a) solvent exposed NHs belonging to nonhelical residues. b) solvent exposed NHs that belong to one of the first three residues of a helical conformer. c) NHs that are not solvent exposed because they are helically hydrogen bonded (sites $i + 3$ through k of the helical region). Given $w(k)$ values, a LR analysis allows calculation if the mole fractions χ_a, χ_b, χ_c. Within the set of solvent exposed NHs, the normally exchanging mole fraction $\chi_{nh} = \chi_a / (\chi_a + \chi_b)$ and the rapidly exchanging mole fraction $\chi_f = \chi_b / (\chi_a + \chi_b)$.

Using the adjustment model described in Equation 3, the modified rate constant k_{se} can be expressed as the sum of two products: the rate constant k_{nh} for the solvent-exposed NHs that are isolated from helical conformations times their mole fraction χ_{nh} plus the rate constant k_f for the faster exchanging NHs that belong to the first three

residues of a helical region times their corresponding mole fraction χ_f. As noted above, LR modeling using w values assigned from the central residue exchange experiments yields the required mole fractions.

$$k_{se} = \chi_{nh}\ k_{nh} + \chi_f\ k_f$$

Equation 3: Modification to the reference rate k_{se} used to calculate protection factors.

The mole fraction of solvent exposed NHs that belong to one of the first three sites (χ_b) of a helical conformer is largest at the N-terminus residue and approaches zero at the residue of maximum protection. Residue sites 2 and 3 of $-A_{15}-$ have the largest mole fraction[18] of rapidly exchanging conformations, $\chi_{nh}= 0.56$ and 0.44 respectively.[19] Residue sites 2 and 3 are ideal to approximate k_f because these site are always solvent exposed $\chi_a + \chi_b = 1.0$. Using the rate constant for the reference peptide $-A_5-$ as k_{nh} it is possible to solve for k_f; the k_f rate constants for sites 2 and 3 equal 0.154 min^{-1} and 0.074 min^{-1} respectively. Although the fast rate constants k_f differ by 2-fold, this preliminary adjustment model allows one to evaluate how both values adjust k_{se} and the resulting PF. This simple model assumes that the first three helical residues in a conformer are equivalently effected by the helix dipole, a more complex model may be able to solve for a universal k_f when the structural cause is known from experiment. As a first approximation only, the best that can be expected from this test is a demonstration that a correction, however crude, can explain the magnitude of the deviations seen in Figure 14A.

---

[18] Site 1 is not suitable for this analysis due to the protection caused by the bulk of the ^tL residue.
[19] The mole fractions of χ_{nh} and χ_f were calculated utilizing a major helix approximation (Chapter 1) and applying the length dependent LR helical weights.

Each of these rate constants were then used to adjust the reference rate k_{se} used to calculate protection factors for sites 2-10 of -A_{15}-. Using the value of 0.154 min^{-1} results in an overestimation of the length dependent LR prediction for all residue sites 2 - 10. The results of incorporating the fast rate calculated from residue site 3 (0.074 min^{-1}) by adjusting k_{se} and the resulting PF$_i$ are illustrated Figure 14B.

Two adjustments to the mole fractions of sites with solvent exposed NHs that belong to one of the first three residues of a helical conformer were made in this adjustment model. The first assumes that the N-terminal rate increase will begin when the helical conformer reaches six residues, the second version starts at nine residues, based on the hydrogen bonding patterns in Figure 13. The six residue assumption within the adjustment model does increase the experimental protection factor to correspond with the LR prediction for the first few resides, but dramatically overestimates the protection factor at sites 8 and 9. This overestimation is diminished if the length at which this effect takes place is increased to 9 residues.

This family of adjustment models for the solvent exposed rate constant used to calculate protection factors results in a reasonable fit to the predicted LR calculation. Further experimentation designed to isolate these independent variables and to explain the plausible model assumptions used to calculate χ_f are necessary before concluding the cause of the fast exchange of the N-terminal region of A_{15}. The important conclusion from this preliminary study is that the magnitude of the change necessary appears to be related to the mole fraction of N-terminal sites as a function of w in a LR analysis.

Three important conclusions from this study of A_{15} PF values. First, a simple Englander calculation of N-terminal peptide PF values from site dependent exchange rate

constants will seriously underestimate these values, and accurate PF modeling must correct the value of k_{se} to accommodate unusually rapid exchange of NH protons that is expected at sites of conformations that lie within the first three N-terminal residues of a helical region. Second, a crude first model for this correction can explain the magnitudes of the deviations observed for Englander PF values. Third, central residue PF values calculated by the Englander value of k_{se} are almost certainly accurate within experimental error.

In Chapter 2, length dependent w values were proposed to model t/c values measured for Ac-Hel-A_n- peptides of lengths between 4 - 14 residues. In this Chapter for studies at 2 °C the use of protection factors calculated from hydrogen exchange rate constants was successful in expanding the length range of an -A_n- series. The experimental -A_{15}- protection factor site scan did match the LR prediction for residue site most predicted from exchange.

The most important result of this study is the validation of length dependent w values for the length series from 5 to 25 residues. If hydrogen bonding cooperativity correctly explains the length dependent w values of Table 1 then this effect may be a universal property of helices. If verified for heteropeptides, helices formed by biopeptides that are significantly longer than the average 11 to 13 residues found in globular proteins may exhibit unexpected high helicity. An interesting candidate has been recently reported by Kuhlman.[20] A 34 residue sequence from ribosomal protein L9 that lacks core packing stabilization is completely helical within its native context and the fragment in water retains high helicity. Using the Table 1 $w(25)$ value of 1.7 as a lower

[20] Kuhlman B.; Yang, H. Y.; Boice J. A.; Fairman R.; Raleigh, D. P. "An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability" *J. Mol. Biol.* **1997**, *270*, 640-647.

bound, one calculates that relative to $w(13) = 1.48$ the length dependent w value increases

helicity for a 34 residue peptide by at least 100 fold. Much larger stabilization

increments may well be characteristic of the helical structures in proteins.

In the efforts toward establishing reliable analytical techniques to provide incisive

experiments on simple helical peptides, hydrogen exchange has provided a potential

explanation to resolve part of the controversy regarding the w value of alanine. In

Chapter 1 and Chapter 2 using utilizing short polyalanines the Kemp group assigned w_{Ala}

$= 1.0$ and the Baldwin group using longer alanine rich peptides assigned a value of 1.6.

Based on the protection factor measurements in this Chapter, it appears that both

assignments are reasonable when applied appropriately to short and long alanine rich

peptides. Although all of the subtleties of the previous experiments need to be explored,

the length dependent values of polyalanines appear to be an important step in

characterizing this series. This series can now be used as a host to assign w values to

guest amino acids that are introduced at specific sites.

STABILIZING N- AND C- CAPS FOR POLYALANINES

The final section in this thesis demonstrates a novel use of protection factor measurements to demonstrate and quantify the high helicity of capped polyalanines. As an introduction this section focuses on the capping issue by reviewing previous work in the Kemp group. Many research groups have developed a variety of methods to stabilize short peptide sequences that would not normally form secondary structure. This type of structural engineering has taken many approaches, including: incorporation of side chain salt-bridges and covalent bonds as well as stabilizing amino acids N- and C- caps. Small molecule templates like Ac-Hel have been developed within this capping approach. All share the same goal of stabilizing the single completely helical conformation, such that it dominates the manifold of partially helical conformers.

To achieve this goal the Kemp group has focused on using a variety of N- and C-terminal caps to stabilize polyalanines. Recently Wolfgang Maison and Eva Arce sought to increase the capping acetyl capacity of Ac-Hel by replacing the N-acetyl with a variety of hydrogen bonding donors, some of which included negative charges, as seen in Figure 15 Both malonyl-Hel and a acetyl-$^\beta$aspartyl-Hel (β-linked) dramatically increase the fractional helicity as measured by circular dichroism by 66% and 45% respectively in comparison to Ac-Hel.[21] in the series X-Hel-A_4KA_3-^{l}LInp$_2$-K_4-W-NH$_2$. Although ac-$^\beta$aspartyl-Hel is a slightly weaker helix stabilizing N-cap than malonyl-Hel, the Ac-$^\beta$aspartatyl-Hel cap is of paticular interest because of its ability to extend the peptide sequence beyond the N-terminus, while maintaining helix stabilization and the structural

---

[21] Maison, W.; Arce, E.; Renold, P.; Kennedy, R. J.; Kemp, D. S. "Optimal N-Caps for N-Terminal Helical Templates, Effects of Changes in H-Bonding Efficiency and Charge" *J. Am. Chem. Soc.* **2001**, *123*, 10245-10254.

integrity of the N-terminal stop signal.[22] Additional residues such as the isolator (*IS*) and solubilizer (*SO*) residues can be added if necessary. Figure 16 illustrates the proposed conformation that allows the carboxylate anion to stabilize the helix with an additional hydrogen bond acceptor and a negative charge.[23]

Figure 15: Highly Stabilizing N-terminal Caps: A) Ac-Hel B) Malonyl-Hel C) Acetyl-$^\beta$Aspartyl-Hel (β-linked)

Figure 16: Replacing the N-acetyl with malonate (R=H) provides an additional hydrogen bond acceptor for the first alanine amide NH, in addition to a stabilizing negative charge at the N-terminus. Ac-$^\beta$aspartyl (R= NHCOCH$_3$) can be used to extend the peptide chain in the N-terminal direction.

---

[22] The structure of X-Hel- prevents helical structure toward the N-terminus.

[23] Recent work by Gabe Job and Bjoern Heitmann in the Kemp group has shown a decrease in CD ellipticity for short peptides when the carboxylate anion is protonated.

The success with X-Hel modifications reflects the advantages of constructing conformationally preorganized capping structures that complement charge and hydrogen bonding features of the α-helix. Due to the difficulties of designing and synthesizing a constrained tricyclic molecule that can provide three hydrogen bond donors for the three amide carbonyls at the C-terminal end region, the Kemp group originally looked at constructing a single positively charged amidinium ring derived from functionalized side chains of unnatural amino acids linked at the peptide C-terminus. The first proof of principle was demonstrated by Eloise Young with the synthesis of (S)–α-(2-aminoethyl)-methionine, Figure 17A, which when incorporated in a peptide sequence, can be converted to a cyclic amidinium ion by BrCN treatment.

The Kemp group has since focused on using other positively charged amino acids to stabilize the C-terminus. Early work of Sam Tang and Peter Renold in the group looked at the position dependence when lysine and arginine residues are introduced into Ac-Hel initiated alanine rich peptides, Figure 17A and B respectively. Increased helicity was observed when lysine or arginine was placed near the C-terminus.[24] Similar results for lysine and arginine have been obtained by Bob Moreau within the SO-IS-Ala$_n$-IS-SO motif. As noted in Chapter 1, lysine and arginine can add ambiguity to the length of the helical region since they can either act as caps terminating the helix, or they can adopt helical ϕ, ψ angles and propagate the helical backbone.

The use of positively charged unnatural amino acids was further explored in the Kemp group. Following Katrin Grobke's[25] and Peter Walliman's studies of the helix

---

[24] Kemp group unpublished results,
[25] Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. "Template-Nucleated Alanine-Lysine Helices are Stabilized by Position-Dependent Interactions Between the Lysine Side Chain and the Helix Barrel" _Proc. Natl. Acad. Sci. U.S.A._ **1996**, _93_, 4025-4029.

stabilization of lysine analogs,[26] Songpong Deechongkit discovered that β-aminoalanine,

Figure 17E, is a strongly helix-stabilizing C-terminal cap. It is also a helical stop signal

that prevents propagation of the helical backbone.[27] The t/c data relevant to β-amino

alanine is shown in Figure 7 of Chapter 2. Peter Rudolf's synthesis of a β-aminoalanine

analog, Figure 17D, added a second hydrogen bond donor but was shown to have an

equivalent capping capacity comparable to β-aminoalanine as measured by CD for Ac-

Hel-A_{12}-X-InpK$_2$-W-NH$_2$. The capping capacity of β-aminoalanine and its commercial

availablilty makes it an ideal choice for futher studies and practical applications.

Figure 17: Structures of amino acid and amino acid derivatives explored as stabilizing C-terminal caps.

---

[26] Kemp Group unpublished results

[27] Deechongkit, S.; Kennedy, R. J.; Tsang, K. Y.; Renold, P.; Kemp, D. S. "An unnatural amino acid that controls polypeptide helicity β-Amino alanine, the first strong C-terminal helix stop signal," _Tetrahedron Lett._, **2000**, 41, 9679-9683.

This chapter will now present the first example of helicity data for polyalanines in which stabilizing N- and C- terminal caps are incorporated in the same sequence. The single completely helical conformation benefits from the stabilizing features of both caps; with the potential to make it the dominant conformation. The circular dichroism studies will compare polyalanine cores capped with both stabilizing, and neutral caps. As discussed in Chapter 1, circular dichroism remains the standard for quantitative measurements of helicity, since the ease of CD acquisition allows for quick comparison of the efficiencies of a series of caps. The perresidue molar ellipticity at 222 nm ($[\theta]_{n,222}$) is the standard quantity used to calculate and compare FH in Equation 5, since $[\theta]_{n,222}$ corrects for peptide concentration, cell length and the number of helical residues as noted in Equation 4.

$$[\theta] = \frac{\theta}{c\ l} \qquad [\theta]_n = \frac{\theta}{c\ l\ n}$$

Equation 4: Molar ellipticity $[\theta]$ is proportional to the difference in molar extinction coefficients between left and right circularly polarized light, reported in degrees θ, divided by the peptide concentration and cell path length (deg cm^2 dmol^{-1}). Perresidue molar ellipticity $[\theta]_n$ also corrects for the number of residues, n, within the helical region (deg cm^2 dmol^{-1} res^{-1}).

$$FH_{helix} = \frac{[\theta]_{n,222,obs}}{[\theta]_{n,222}} \qquad [\theta]_{n,222} = [\theta]_{\infty,222}\left(1 - \frac{x}{n}\right) \qquad 2 < x < 5$$

Equation 5: CD derived experimental determination of fractional helicity as the observed perresidue molar ellipticity $[\theta]_{n,222,obs}$ divided by the limiting perresidue molar ellipticity $[\theta]_{n,222}$ for length n at 222 nm. The limiting perresidue molar ellipticity is calculated from a value of the perresidue molar ellipticity for a helix of infinite length $[\theta]_{\infty,222}$ with moderate length correction $(1 - x/n)$. Experimental fits for x have ranged from $2 < x < 5$.[28]

Figure 18 illustrates the powerful effect that stabilizing N and C-terminal caps can have on short helical polyalanines. All four of the peptides in this comparison have an $-A_{12}-$ core, yet the uncertainty of the helical length contributed to the helix from the capping regions, makes a calculation of $[\theta]_n$ difficult. Even if the length of the helical region is uncertain, molar ellipticity ($[\theta]$) can be used to make qualitative comparisons of capping efficiencies. The CD spectrum for A: $W-K_4-Inp^tL-A_{12}-^tLInp_2-K_4-NH_2$ is standard for an unstructured peptide; at 222 nm there is little or no ellipticity. The presence of ^tL at the N- and C-terminus was initially assumed to be a neutral cap, but is actually slightly destabilizing.[29] The alanine core capped with lysine residues, B: $W-K_3-A_{12}-K_3-NH_2$ has a moderately helical CD signature. The NMR investigation of this sequence reported in Chapter 1 demonstrated the extension of the helix into the three lysine residues at the C-terminus, this ambiguity of helical length in this sequence prevents an accurate calculation of its per-residue ellipticity.

[28] Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W., Kallenbach, N. R. "The Helix-Coil Transition in Heterogeneous Peptides with Specific Side-Chain Interactions: Theory and Comparison with CD Spectral Data" *Biopolymers* **1991**, *31*, 1605-1614.

[29] Maison, W.; Kennedy, R. J.; Miller, J. S.; Kemp, D. S. "C-Terminal Helix Capping Propensities in a Polyalanine Context for Amino Acids Bearing Nonpolar Aliphatic Side Chains" *Tetrahedron Lett.* **2001**, *42*, 4975-4977.

A comparison of the CD spectra of the last two peptides in the series demonstrates the chameleon like properties of short polyalanine peptides. With the neutral or slightly destabilizing ^tL cap, $-A_{12}-$ does not show any helical character. But when Ac-Hel is added as an N-terminal cap to yield C: Ac-Hel-A_{12}- $^tLInp_2$-K_4-W-NH_2, the standard helical minima are observed at 208 and 222nm. Both of these minima become more intense, hence more helical, if Ac-Hel is added at the N-terminus and β-aminoalanine is placed at the C-terminus, to yield D: Ac-Hel-A_{12}-β-Inp-K_2-W-NH_2.[30] The CD data for these four peptides indicated that an alanine core is strongly influenced by helix stabilizing caps, a dramatic increase if helicity is observed when both Ac-Hel and β are placed at the N- and C- terminus respectively.

---

[30] Although rarely seen in proteins and peptide CD spectra, in comparison to 208 nm the larger intensity of the 222 nm band may reflect a completely helical structure. If 12 residues is used to calculate the perresidue ellipticity $[\theta]_{12,222,obs}$ and the limiting ellipticity $[\theta]_{12,222}$ the value of FH = 0.90.
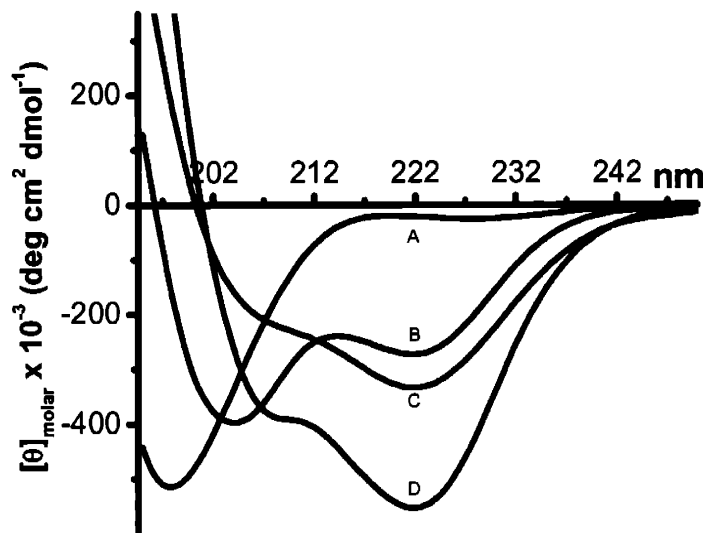
A: W-K_4-Inp$'$L-$\mathbf{A_{12}}$-$'$LInp$_2$-K_4-NH$_2$

B: W-K_3-$\mathbf{A_{12}}$-K_3-NH$_2$

C: Ac-Hel-$\mathbf{A_{12}}$-$'$LInp$_2$-K_4-W-NH$_2$

D: Ac-Hel-$\mathbf{A_{12}}$-β-Inp-K_2-W-NH$_2$

Figure 18: Circular Dichroism data for a series of capped A_{12} peptides at 2 °C in water. Helicity is monitored at 222 nm. A) No helicity is detectable for A_{12} when capped by tert-leucine, a partially destabilizing cap. B) Moderate helicity is observed for the lysine capped peptide, that was studied by NMR in Chapter 1. C) Use of the stabilizing N-cap Ac-Hel, markedly increases the helicity of A_{12}. D) Use of both Ac-Hel and the stabilizing C-cap β (β-amino-alanine) dramatically increases the helicity of A_{12}. Molar ellipticity is plotted due to the uncertain length of the helical region that may extend into the cap region of B.

HYDROGEN EXCHANGE OF HIGHLY HELICAL CIRCULAR DICHROISM STANDARDS

The goal of this section is to provide the first example of a CD helicity calibration using short polyalanines bearing the most effective N- and C- caps. These CD measurements of short, N- and C- capped polyalanines that are not normally helical will demonstrate the how effectively the capping effects, described above, can increase helicity. Utilizing hydrogen exchange to measure the PF for each residue site will be used to assign of FH. This FH value can be combined with the observed ellipticities ($[\theta]_{n,222,obs}$) to solve for the limiting ellipticity $[\theta]_{n,222}$. If this proof of principle experiment is successful, a length series of highly helical N- and C- capped polyalanines can be used to determine the length dependence x and the perresidue ellipticity for a peptide of infinite length $[\theta]_{\infty,222}$ in Equation 5

In an joint project with Peter Wallimann, the series Ac-Hel-$(A_4K)_m$-A_2-NH_2 ($m = 1 - 5$) was previously studied using CD. Although only stabilized at the N-terminus the 100% helical length dependence as described in Equation 5, using -44,000 as $[\theta]_{\infty,222}$, was exceeded for all three 17, 22, and 27 residue peptides. This initial study[31] suggested that the value of $[\theta]_{\infty,222}$ could be as large as -61,000 cm^2 dmol^{-1} res^{-1}. Justin Miller corroborated this result when CD properties of for W-K_m-Inp_2'L-A_n- 'LInp_2-K_m ($n = 4$ - 45), as pictured in Chapter 1, Figure 21, were measured. Fitting this data set with $[\theta]_{\infty,222}$ and w as free variables provided additional evidence that -61,000 deg cm^2 dmol^{-1} res^{-1} may be the limiting ellipticity for polyalanine helices.[5b] Both of these projects

---

[31] Wallimann, P.; Kennedy R. J.; Kemp, D. S. "Large Circular Dichroism Ellipticities for N-Templated Helical Polypeptides Are Inconsistent with Currently Accepted Helicity Algorithms" *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1290-1292.

provide evidence that $[\theta]_{\infty,222}$ is underestimated for partially helical longer peptides. The experiments in this Chapter explore if this is also true for highly helical short peptides.

Short polyalanines are not usually helical. Sheri Oslick in the Kemp group looked at the CD properties of (H)-A_6-OH and found it was unstructured,[32] and Justin Miller did not see helical CD signals for -A_n- shorter than A_{14} within the SO-IS-Ala_n-IS-SO series[5] capped with ^{t}L. Based on these results one would not expect to see any helical CD signals for an -A_8- polyalanine core.
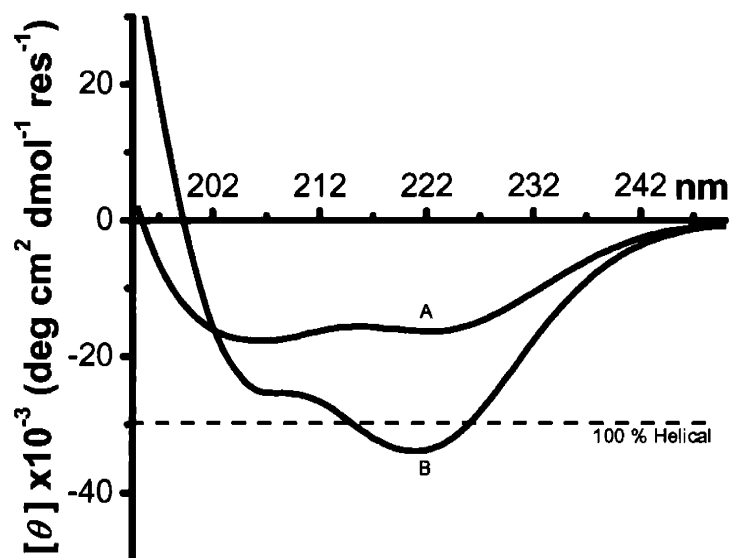
A: Ac-Hel-A_8-^{t}LInp$_2$-K$_4$-W-NH$_2$

B: Ac-$^\beta$Asp-Hel-A_8-β-Inp$_2$-K$_2$W-NH$_2$

Figure 19: CD data for highly helical octa-alanine peptides at 2 °C in water. The addition of βAsp, replacing Ac on the Hel template, and β-amino alanine dramatically increase the helicity of a alanine sequence that would not otherwise be helical.

[32] Oslick, S. L. "Characterization of Short Template-Nucleated Helical Peptides" Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 1996.

Figure 19 compares the perresidue molar ellipticities for a pair of octaalanine cores,[33] the first bearing Ac-Hel as an N-cap and a second peptide C-capped with β-amino alanine and bearing Ac-$^\beta$aspartyl-Hel as an N-cap. When N-capped by Ac-Hel the -A$_8$- does show a standard helical CD spectrum with a double minimum. A dramatic increase in helicity is observed when the N-terminal cap is strengthened by replacing Ac with $^\beta$Asp and β-amino alanine is added to the C-terminus. What is truly remarkable is that this capped octaalanine peptide shows a $[\theta]_{222}$ that also exceeds the 100% limit calculated using -44,000 as $[\theta]_{\infty,222}$. This case provides additional evidence that the CD scale is in need of highly helical standards, for both short sequences and long sequences.

Hydrogen exchange was then used to measure the decay of the amide NHs at each residue site of peptide B of Figure 19, Ac-$^\beta$Aspartyl-Hel-A$_8$-β-NH$_2$. The average protection factor is 17, corresponding to a $\chi_{NHi, hb}$ of 0.94 at pH 3.5. The average protection factor measured for malonyl-Hel-A$_8$-β-NH$_2$ is also 17 corresponding to $\chi_{NHi, hb}$ of 0.95 at pH 3.5.[34] Due to the dispersion of amide NH chemical shifts, ^{15}N isotopic labeling was not necessary to measure site hydrogen exchange. The identity of the ^{1}H amide shifts was assigned by Kemp group member Bjoern Hietmann.

Figure 20 compares the dramatic effect of using the Ac-$^\beta$Asp-Hel N-cap and β-aminoalanine as a C-cap on a short polyalanine (B) with prediction of $\chi_{NHi, h} = 0$ for an octaalanine that is not highly stabilized. Note that unlike the first three residues of a regular helix, the first three alanine amide NHs are protected from exchange by the three

---

[33] The polyalanine core was used to determine the perresidue molar ellipticity $[\theta]_{8,222,obs}$. The two proline residues are not expected to make a large contribution to the helical signal at 222 nm (Woody, R. W. Improved Calculation of the nπ* Rotational Strength in Polypeptides" J. Chem. Phys. 1968, 49, 4797-4805; Sreerama, N.; Woody, R. W. Circular Dichroism of Peptides and Proteins, In *Circular Dichroism: Principles and Applications*, Berova, N.; Nakanishi, K.; Woody, R. W.; Eds.; Wiley and Sons, 2000).

[34] Current work had shown that at pH 3.5 the malonate acid may not be full deprotonated, explaining the efficiency of malonyl vs. Ac-$^\beta$aspartyl seen here and Maison's CD study.[29]

hydrogen bond donors of the Ac-$^\beta$Asp-Hel N-cap. In addition all of the alanine residues

are highly helical, there is no monotonic decease of helicity toward the N- and C-termini,

and the conformational averaging observed for -A_{15}- is not present. This result indicates

that the dominant conformation in solution is the completely helical sequence that

benefits from both the N- and C-terminal caps.

A   B
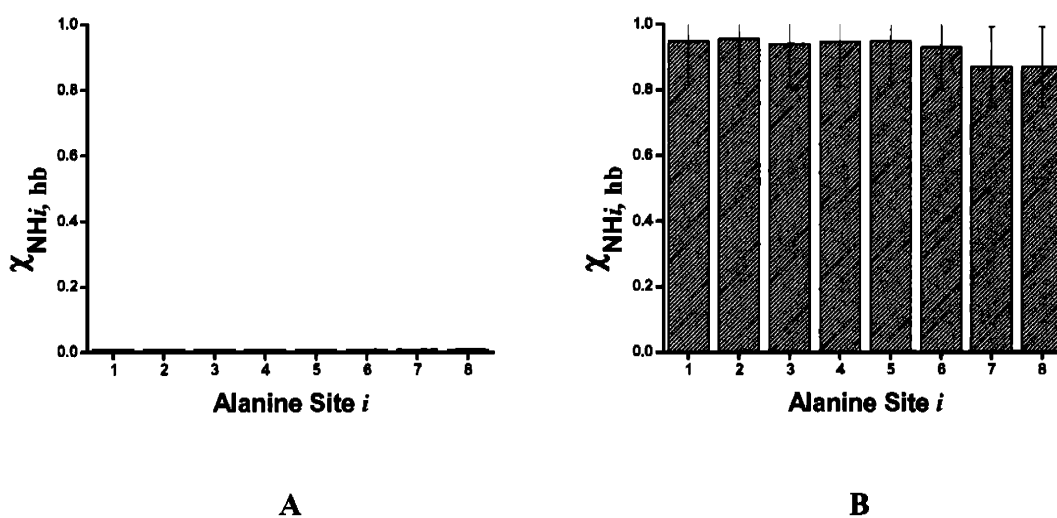
Figure 20: Comparison of the predicted mole fraction of hydrogen bonded sites $\chi_{NHi,\ hb}$ octaalanines: A) No helical character is expected when capped with a neutral cap as noted in Figure 12. B) With highly stabilizing N- and C- caps Ac-$^\beta$Asp-Hel-A_8-β-NH$_2$ is very highly helical with mole fractions of hydrogen bonded sites between 0.86 and 0.96. The Error bars are set at 14%. The solvent exposed reference for this experiment was the average for all three alanines in Ac-A_3-OH. Hydrogen change was measured in 20mM NaH$_2$PO$_4$/ D$_2$O and extrapolated to pH 3.5.

The high helicity observed for this peptide provides proof of principle and makes

this a strong potential candidate for a larger series to calculate the length dependent

circular dichroism in Equation 5. Hydrogen exchange will be used to assign the PF$_i$ at

each residue site and to calculate FH.[35] This FH value, combined with the observed

ellipticities, $([\theta]_{n,222,obs})$ will yield the limiting ellipticity $[\theta]_{n,222}$. This study is currently

under investigation in the Kemp group[36] and should yield the series of calibration

standards that will establish $[\theta]_{n,222}$ from CD ellipticities for each alanine length.[37]

Highly stabilizing N- and C- terminal caps were successful at inducing FH values

comparable Figure 20 to alanine cores between 4 and 25 residues.[38] The entire length

series will be used to solve for $[\theta]_{\infty,222}$ and x of in Equation 5.

SUMMARY AND FUTURE POTENTIAL

In this Chapter hydrogen exchange has played a key role in determining the

helical behavior of two very different types of peptides. In the first part of this chapter,

hydrogen exchange was used to validate the length dependent w values for polyalanines

first detected in Chapter 2. The mole fraction of sites that are hydrogen bonded vary

dramatically from the N- to C-terminus for the ^{L} capped $-A_{15}-$ peptide. Protection

factors give individual information for each site, whereas CD and t/c ratios cannot. In

addition the site scans of similar polyalanines have the potential to explain the increased

rate of exchange observed at the N-terminal region of helical peptides if an appropriate

model can be used to explain the phenomenon. In the second part, hydrogen exchange

was used to quantify the helicity for highly helical peptides that are stabilized with N-

and C- terminal caps. These locally defined helices, with a stable secondary structure

have the potential to be used to engineer a variety of useful conformational motifs such as

---

[35] $\chi_{NHi,h} \cong \chi_{\alpha Ci}$ when the mole fraction of all sites is greater than 0.85.

[36] Bjoern Heitmann and Gabe Job

[37] The entire series will be able to determine values for x and $[\theta]_{\infty}$. Although the single point for $n = 8$ does yield $[\theta]_{8,222.} = -36,000$ ($[\theta]_{8,222,obs}$ is -34,000 at 2 °C in water, divided by the fractional helicity of 0.94)

[38] Isolators (*IS*) and solubilizers (*SO*) were needed at both the N- and C- terminus for this series.

CD calibration standards, robust proteins like coiled-coils or molecular cylinders for defined length that may be useful as building blocks for molecular recognition. In addition, owing to their conformational homogeniety these highly helical molecules are excellent candidates of NMR structural analysis. Currently in progress in the group, these experiments will define the hydrogen bonding pattern and have the potential for distinguishing a temperature dependent helical structural change from loss of helical structure needed to explain the anomalous intensity observed in CD spectra at 222 nm, as seen in Chapter 1.

The global aim of this thesis has been to use three independent methods to quantitate fractional helicity of polyalanines. In combination, all of these methods can provide cross checks of phenomena observed for peptides that adopt helical structures. The interrelationships between the methods in Figure 21 have been strengthened through experiments presented here. These cross checks should help unify the tools used to measure fractional helicity and settle present and future disagreements about the appropriateness of these experimental methods.

Future work with a series of highly helical peptides will provide calibration standards for circular dichroism, allowing accurate calculation of FH from CD measurements. This accuracy will help establish a long-range Kemp Group aim: a chemical understanding of the factors that govern helicity, embodied in an algorithm that accurately predicts the stability of peptide helices from their amino acid compositions and sequences. Amino acid mutants for all of the natural amino acids can be studied by CD and measure the amino acid propensity and sequence dependence of fractional helicity. In the Kemp group, Bob Moreau's preliminary relative helical propensities for mutants in

the $SO\text{-}IS\text{-}Ala_n\text{-}IS\text{-}SO$ motif have shown a rank ordering of helical propensities in excellent agreement with Blaber data[39] from protein and peptide models. The site helicities from protection factors obtained in this Chapter provide an important step in defining the helical properties of polyalanine as helical host, such that absolute w values for other residues and quantitative site and context effects can be reliably assigned.
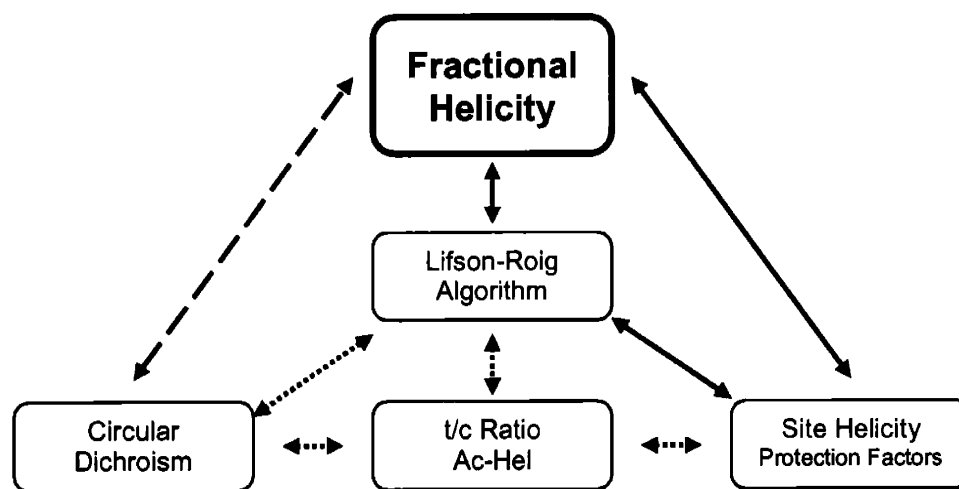
Figure 21: Solid lines indicate methods compared in this chapter, Dashed lines indicate improvement on the direct relationship between FH and CD.

[39] Blaber, M.; Zhang, X.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W. A.; Matthews, B. W. "Determination of α-Helix Propensities within the Context of a Folded Protein" *J. Mol. Biol.* **1994**, *235*, 600-634.

CHAPTER 4: EXPERIMENTAL PROCEDURES

SMALL MOLECULE SYNTHESIS

Ac-Hel-OH was synthesized via the convergent 16-step synthesis described by Kim McClure.[1] Fmoc-Hel-OH was prepared by Wolfgang Maison and used without further purification.[2]

PEPTIDE SYNTHESIS

Peptides were synthesized on a 0.025 or 0.050 mmol scale by a Pioneer Peptide Synthesizer using Fmoc (9-fluorenylmethoxycarbonyl) solid phase peptide synthesis. Three equivalents of Fmoc-amino acids (Novabiochem) were activated by 3 equivalents of both HATU (*O*-(7-azabenzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate from Applied Biosystems) and DIEA (diisopropylethylamine) before the 30 minute, continuous flow coupling with the PAL-PEG-PS resin (polystyrene (PS) functionalized, polyethylene glycol (PEG) Peptide Amide Linker (PAL) from Applied Biosystems). Resin washes used biosynthesis grade DMF (*N,N*-dimethylformamide). Deblocking used a 48:1:1: (v:v:v) mixture of DMF: piperidine : DBU (1,8-diazabicyclo[5.4.0]undec-7-ene). After the final wash, the resin was washed with methanol and dried with nitrogen gas.

Extended couplings of 1 hour were necessary to achieve adequate yields with 1 equivalent of Ac-Hel-OH and Fmoc-Hel-OH and ^{15}N labeled amino acids. To prevent a

---

[1] McClure, K.; Renold, P., Kemp, D. S. "An Efficient Synthesis for Ac-Hel1-OH" *J. Org. Chem.*, **1995**, *60*, 454-457.

[2] Maison, W.; Arce, E.; Renold, P.; Kennedy, R. J.; Kemp, D. S. "Optimal N-Caps for N-Terminal Helical Templates, Effects of Changes in H-Bonding Efficiency and Charge" *J. Am. Chem. Soc.* **2001**, *123*, 10245-10254.

deletion sequence, a second coupling using a non-labeled alanine followed the initial coupling.

Peptides were cleaved manually from the resin using a fresh cocktail made up of 82.5% trifluoroacetic acid, 5% phenol, 5% water, 5% thioanisole, 2.5% 1,2-dithioethane for two hours. Dropwise addition into 30 ml of diethyl ether precipitated the peptide as a white solid which was isolated by centrifugation and decanting the ether. The solid was washed and isolated 3 times with diethyl ether before being dried under vacuum.

PEPTIDE PURIFICATION

The crude peptide was taken up in 15 mL water (18.2 MΩ from a Millipore Gradient A10 water purification system) a 50 μL sample was used for crude analytical HPLC analysis. Analytical HPLC was preformed on a Waters 2690 HPLC with auto injector and a 996 PDA UV-detector. Using a YMC ODS-AQ 200Å, 4.5 × 150 mm stainless steel column (Waters) and Water (0.05% TFA): Acetonitrile gradients of 95:5 to 5:95 over 30 minutes were used to judge synthesis success. Gradient optimization followed to obtain separation of the major product over a 30% change in acetonitrile with a 1%/min gradient.

Semi-preparative HPLC was preformed on a Waters 600 HPLC equipped with a 996 PDA UV-detector and fraction collector. Using a YMC ODS-AQ 200Å, 10 × 150 mm stainless steel column (Waters) gradients of 1%/min acetonitrile over a 30% window were used to fractionate the major products. The final purity of the peptide was judged by analytical HPLC and mass spectrometry. Peptide purity was greater than 95 %. Semi-preparative HPLC samples were freeze dried and stored as a solid at 10 °C.

Both analytical and semi-preparative HPLC used monitored UV detection at 214

and 280 nm.

PEPTIDE IDENTIFICATION: MASS SPECTROMETRY

Mass spectra were obtained on crude and purified peptides samples using a

Waters Micromass ZMD 4000 mass spectrometer with electrospray ionization.   The

instrument was calibrated in the m/z range of 50 to 2000 using NaI/ RbI (2.0/0.05 µg/µL)

in 1:1 water: ispropanol; the multiple Lys residues create a charged m/z manifold within

this calibration range. A tolerance of ±0.75 mass units is observed.

New Peptides	Electrospray Mass Spectrometry ESI+ $[m + z\text{H}]/z$ Found (Expected)
AcHel-A_4-^{t}LInp$_2$K$_4$W-NH$_2$	539.77 (539.32); 405.06 (404.74)
AcHel-A_5-^{t}LInp$_2$K$_4$W-NH$_2$	842.78 (844.00); 562.45 (563.00)
AcHel-A_6-^{t}LInp$_2$K$_4$W-NH$_2$	879.80 (879.52); 586.59 (586.68)
AcHel-A_7-^{t}LInp$_2$K$_4$W-NH$_2$	915.14 (915.03); 610.20 (610.36); 457.91(458.02)
AcHel-A_8-^{t}LInp$_2$K$_4$W-NH$_2$	950.83 (950.55); 634.58 (634.04); 476.01 (475.78)
AcHel-A_9-^{t}LInp$_2$K$_4$W-NH$_2$	986.12 (986.07); 657.94 (657.72); 395.08 (395.03)
AcHel-A_{10}-^{t}LInp$_2$K$_4$W-NH$_2$	1021.55 (1021.59); 681.30 (681.40); 511.15 (511.30)
AcHel-A_{11}-^{t}LInp$_2$K$_4$W-NH$_2$	1057.17 (1057.11); 705.76 (705.07); 528.76 (529.06)
AcHel-A_{12}-^{t}LInp$_2$K$_4$W-NH$_2$	1093.02 (1092.63); 728.97 (728.75); 546.97 (546.82)
AcHel-A_{13}-^{t}LInp$_2$K$_4$W-NH$_2$	1127.84 (1128.14); 752.31 (752.43); 564.41 (564.58)
AcHel-A_{14}-^{t}LInp$_2$K$_4$W-NH$_2$	1163.29 (1163.66); 775.87 (776.11); 582.11 (582.34)
(H)- ^{t}LInp$_2$K$_4$W-NH$_2$	1051.71(1051.71); 526.30 (526.36); 351.18(351.18); 263.58 (263.58)
Ac-$^\beta$D-Hel-A_8-β-NH$_2$	1067.29 (1067.49); 534.10 (534.25)
Ac-$^\beta$D-Hel-A_8-β-InpK$_2$W-NH$_2$	840.70 (840.92); 540.81 (540.95)
malonyl-Hel-A_8-β-NH$_2$	994.38 (994.44) [ESI-]

β = beta-amino alanine, $^\beta$D= beta linked aspartate

600 MHz NMR EXPERIMENTS

A collaboration with Prof. Harald Schwalbe, along with his students Elke Duchardt and Julia Wirmer, provided the technical skills to operate the Bruker 600 MHz at the MIT Department of Chemistry Instrumentation Facility. Chapter 1, Figure 18 shows two 2D NMR experiments carried out on (H)-K_3-A_{12}-K_3-WNH$_2$, all 12 alanine residues are ^{15}N labeled, at 2 °C in water. Figure 18A shows a ^{1}H-^{1}H NOESY, acquired with a mixing time of 400 ms and ^{15}N filtering;[3] indicating the i to $i + 1$ amide NH to NH connectivities of only the helical alanine core. Figure 18B shows a ^{1}H-^{1}H NOESY acquired with the same mixing time and ^{15}N decoupling; indicating the i to $i + 1$ amide NH to NH connectivities of both the helical alanine and lysine amide NHs. A singly ^{15}N labeled alanine peptide was used to identify the N-terminus.

AC-HEL: T/C RATIO MEASUREMENT

^{1}H NMR spectra for t/c measurements were acquired on the 500 MHz spectrometer running Varian software (VNMR) at the MIT Department of Chemistry Instrumentation Facility. The temperature was monitored using a standard methanol reference. For the series Ac-Hel-A_n-^{t}LInp$_2$K$_4$W-NH$_2$ ($n = 4$ - 14), 1 - 2 mM peptide samples in D$_2$O were prepared after repeated lyophilization from D$_2$O. TMSP (2, 2, 3, 3, D$_4$ Sodium 3-trimethylsilylproprionate) was used as an internal reference and 10 μL of TFA was added to the NMR sample. Samples were equilibrated for 10 minutes at the set probe temperature before calibrating the 90° ^{1}H pulse; NMR ^{1}H spectra were obtained using 256 scans, a delay of 3 second and a 4 second acquisition time. Integration of the t

---

[3] The ^{15}N filtering is generated by a magnetization transfer from the amide proton to the ^{15}N labeled amide and back, (HSQC) eliminating ^{1}H NMR signals for protons not bonded to ^{15}N atoms.

and c NMR signals were used to calculate t/c ratios as previously described.[4]   The t/c

values for n = 5, 6, were consistent with AcHel-A$_n$-NH$_2$ as previously reported.[4] The t/c

values for the series Ac-Hel-A$_n$-^{t}LInp$_2$K$_4$W-NH$_2$ (n = 4 - 14) are listed in Table 1.

Ac-Hel t/c Measurements			
	Temperature °C		
Ala$_n$	2	25	60
4	1.42	1.51	1.76
5	1.71	1.72	1.96
6	2.07	1.99	2.04
7	2.61	2.33	2.27
8	3.24	2.68	2.28
9	4.19	3.22	2.34
10	5.27	3.74	2.44
12	7.41	5.3	2.49
14	9.92	7.06	xxx

Table 1: Temperature dependent t/c values for Ac-Hel-A$_n$-^{t}LInp$_2$K$_4$W-NH$_2$ (n = 4 - 14). No data is available for n = 14 at 60 °C as the peptide precipitated out of solution. A 5% error is observed.

CIRCULAR DICHROISM

CD spectra were obtained on an Aviv 62DS spectrometer with a thermoelectric

temperature controller using 1, 5 and 10 mm QS cells (suprasil, Hellma).   CD

measurements were taken every 0.5 nm and averaged over 5 scans.   After blank

correction, the data were smoothed using polynomial fitting (Aviv software). Peptides

were dissolved in water and concentration was determined on a Cary 300 UV-Vis

[4] Renold, P.; Tsang, K-Y.; Shimizu, L. S.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 12234-12235; Allen, T. J. "Studies of Template-Initiated α-Helix Propagation in Short Peptides" Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 1986 (Region 3).

spectrometer utilizing the Trp chromophore (ε_{280} = 5560 cm^{-1} M^{-1});[5] peptide samples were typically 10-100 μM. As indicated the number of alanines was used as n to calculate the perresidue molar ellipticity$[\theta]_{n,222,obs}$.[6] The perresidue molar ellipticities for the series Ac-Hel-A$_n$-^{t}LInp$_2$K$_4$W-NH$_2$ (n = 4 - 14) are listed in Table 2.

$[\theta]_{n,obs,222} \times 10^{-3}$		
Ala$_n$	**Temp °C**	
	2	**25**	**60**
4	5.9	9.3	12.5
5	6.0	9.6	13.7
6	8.9	9.8	11.6
7	10.7	9.8	10.7
8	15.1	12.5	11.1
9	17.9	13.7	10.6
10	22.5	16.9	11.4
12	27.1	20.7	11.7
14	33.4	25.4	14.3

Table 2: CD ellipticity data, $[\theta]_{n,obs,222}$, for Ac-Hel-A$_n$-^{t}LInp$_2$K$_4$-W-NH$_2$ (n = 4 - 14).

[5] Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities" _J. Am. Chem. Soc._ **2002**, _124_, 945-962.

[6] The polyalanine core was used to determine the perresidue molar ellipticity $[\theta]_{8,222,obs}$. The two proline residues of Ac-Hel are not expected to make a large contribution to the helical signal at 222 nm (Woody, R. W. Improved Calculation of the nπ* Rotational Strength in Polypeptides" _J. Chem. Phys._ **1968**, _49_, 4797-4805; Sreerama, N.; Woody, R. W. Circular Dichroism of Peptides and Proteins, In _Circular Dichroism: Principles and Applications_, Berova, N.; Nakanishi, K.; Woody, R. W.; Eds.; Wiley and Sons, **2000**).

To isolate the $[\theta]_{n,222}$ of only the alanine helix, the molar ellipticity of (H)-^{f}LInp$_2$K$_4$-NH$_2$ was subtracted from the molar ellipticity of the Ac-Hel-A$_n$-^{f}LInp$_2$K$_4$-NH$_2$ series members. The resulting molar ellipticity was divided by the number of alanine residues, yielding perresidue molar ellipticity. The result of this correction is illustrated in Figure 1.
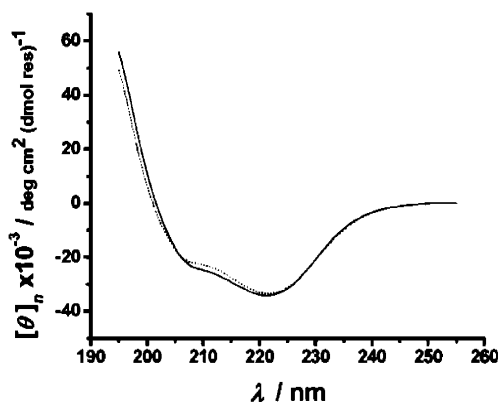
Figure 1: Plot of perresidue molar ellipticity $[\theta]_{n,obs}$ for Ac-Hel-A$_{14}$-^{f}LInp$_2$K$_4$-NH$_2$. The original helical CD curve (dotted), - ^{f}LInp$_2$K$_4$-NH$_2$ corrected CD curve (solid).

Before comparing the t/c and CD ellipticities for Ac-Hel-A$_n$-^{f}LInp$_2$K$_4$-W-NH$_2$ ($n = 4$ - 14), the CD contributions of the cs, ts and te states[7] were also removed by subtracting the limiting ellipticities at 222 nm.[8] The mole fraction of cs, ts and te states can be calculated from the t/c ratio.[9] The corrected values are listed in Table 3.

$$[\theta]_{222} = [\theta]_{222} + \chi_{te} \times [\theta]_{222,te} + \chi_{cs+ts} \times [\theta]_{222,cs+ts}$$

Equation 1: Correction of $[\theta]_{222}$ for the CD spectra of the cs, ts and te states.

[7] Oslick, S. L. "Characterization of Short Template-Nucleated Helical Peptides" Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 1996.

[8] Kemp, D. S.; Allen, T. J.; Oslick, S. L.; Boyd, J. G. *J. Am. Chem. Soc.* **1996**, *118*, 4240-4248.

[9] Kemp, D .S., Allen, T. J.; Oslick, S. Boyd, J. G., "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. II. Characterization of the Helical Structure of Ac-Hel-L-Ala$_6$-OH," *J. Am. Chem. Soc.* **1996**, *118*, 4240-4248.

Experimental t/c and $[\theta]_{n,222,obs}$

	Limiting Ellipticities at 222 nm **deg cm² (dmol)⁻¹ [Molar]**	
	Ac-Hel Correction Used for all temperatures	Cap Correction at 222 nm (ᴸL-Inp₂-K₄-W-NH₂)
	te -750	2 C -9602
		25 C -3452
	cs + ts -24600	60 C 17038

$\chi_{te} = (1-(1.79/(1+t/c)))$

te, cs, ts Correction $[\theta]_{222}$ corr = $\{n[\theta]_{n,222,obs} - \chi_{te}[\theta]_{222,te} +(1 - \chi_{te}) [\theta]_{222(cs+ts)}\}/n$

	A_n			deg cm² (dmol n)⁻¹	Ellipticities -$[\theta]_{222}$ x10⁻³		-$[\theta]_{222}$ x10⁻³
Temp	n	t/c	χ_{te}		Corrected for cs, ts, te	% Correction	Cap Corr & AcHel Corr
2 °C	4	1.42	0.260	5.9	1.3	77.9	3.7
	5	1.71	0.339	6	2.7	55.0	4.6
	6	2.07	0.417	8.9	6.5	27.4	8.1
	7	2.61	0.504	10.7	8.9	16.8	10.3
	8	3.24	0.578	15.1	13.7	9.0	14.9
	9	4.19	0.655	17.9	16.9	5.6	18.0
	10	5.27	0.715	22.5	21.7	3.4	22.7
	12	7.41	0.787	27.1	26.6	1.8	27.4
	14	9.92	0.836	33.4	33.1	1.0	33.8
25 °C	4	1.51	0.287	9.3	4.9	47.7	5.7
	5	1.72	0.342	9.6	6.3	34.3	7.0
	6	1.99	0.401	9.8	7.3	25.6	7.9
	7	2.33	0.462	9.8	7.9	19.8	8.4
	8	2.68	0.514	12.5	11.0	12.4	11.4
	9	3.22	0.576	13.7	12.5	8.8	12.9
	10	3.74	0.622	16.9	15.9	5.8	16.3
	12	5.3	0.716	20.7	20.1	3.0	20.4
	14	7.06	0.778	25.4	25.0	1.7	25.2
60 °C	4	1.76	0.351	12.5	8.4	32.4	4.2
	5	1.96	0.395	13.7	10.7	22.2	7.3
	6	2.04	0.411	11.6	9.1	21.3	6.3
	7	2.27	0.453	10.7	8.7	18.4	6.3
	8	2.28	0.454	11.1	9.4	15.5	7.2
	9	2.34	0.464	10.6	9.1	14.2	7.2
	10	2.44	0.480	11.4	10.1	11.5	8.4
	12	2.49	0.487	11.7	10.6	9.2	9.2
	14	precipitated		14.3			

Table 3: Tabulation of the t/c and ellipticity data for Ac-Hel-A$_n$- ᴸLInp₂K₄-W-NH₂ (n = 4 - 14), including the calculations used to correct $[\theta]$ for *IS-SO* region and the Ac-Hel states: cs, ts and te.

HYDROGEN EXCHANGE

Hydrogen exchange samples were prepared by making 1- 2 mM peptide solutions in H_2O. The pH of this solution was adjusted with 1 μL additions of 1M NaOH or HCl to achieve the desired experimental pH. The adjusted sample was then lyophilized to dryness. After equilibration in an ice bath, the sample was dissolved the 20 mM NaH_2PO_4 /D_2O buffer (pH 4.0) transferred to an NMR tube and quickly inserted into the NMR spectrometer (pre-equilibrated at 2 °C). All pH readings were taken using a 6 mm calomel electrode connected to a Radiometer pH meter calibrated with standards of 4.0 and 7.0 at room temperature. pH is measured using a glass electrode at room temperature; no correction is made for deuterium or temperature. pH measurements before and after exchange did not show a significant change. Sample pH did deviate from run to run, possibly due to residual TFA. The observed rate constants with pHs between 3.5 and 4.1 were adjusted to pH 4.0, based on the first order rate dependence of hydroxide, to calculate protection factors. A 5- 10% error in k_{obs} was observed between trials, hence a 14 % error is applied to PF values.

The first slice of an HSQC experiment (ni = 1, phase =1) was used to obtain a 1D, ^{15}N filtered, ^1H spectrum of only the labeled ^{15}N amide NHs. An array of spectra were obtained every 5 - 10 minutes (nt = 64, at = 3 sec. Total acquisition time = 1 min). Both the peak height and peak area gave satisfactory fits to a first order exponential decay, calculated using the Origin[10] graphing program. Standard 1D ^1H parameters were used to monitor the exchange of Ac-$^\beta$D-Hel-A_8-β-NH_2 and malonyl-Hel-A_8-β-NH_2. Presaturation of the H_2O NMR signal was not necessary.

---

[10] Origin © Microcal Software, Inc.

The rate constants and protection factors discussed in Chapter 3 are listed in the following tables:

Peptide	Avg. Rate at pH 3.5 min^{-1}	Protection Factor	$\chi_{NHi,hb}$
Ac-A$_3$-OH	0.0303	1.0	0.00
Ac-$^\beta$aspartyl-Hel-A$_8$-β-NH$_2$	0.00192	16	0.94
malonyl-Hel-A$_8$-β-NH$_2$	0.00193	16	0.94

Table 4: Average Protection Factors for highly helical octaalanines at 2 °C in 20 mM NaH$_2$PO$_4$/ D$_2$O, pH 3.5.

Alanine Site Ac-$^\beta$aspartyl-Hel-A$_8$-β-NH$_2$	Avg. Rate at pH 3.5 min^{-1}	Protection Factor	$\chi_{NHi,hb}$
1	0.00150	20	0.95
2	0.00132	23	0.96
3	0.00174	17	0.94
4	0.00157	19	0.95
5	0.00149	20	0.95
6	0.00203	15	0.93
7,8	0.00376	8.1	0.88

Table 5: Protection factor measurements for Ac-$^\beta$aspartyl-Hel-A$_8$-β-NH$_2$ at pH 3.5 in 20 mM NaH$_2$PO$_4$/ D$_2$O. The average exchange rate constant for Ac-A$_3$-OH was used as the solvent exposed reference rate constant k$_{se}$.

^{15}N label at site A_i	pH	rate constant at exp. pH min^{-1}	rate constant at pH 4.0 min^{-1}	% error	PF
1	3.38	0.0039	0.0174	0.4	2
2	3.38	0.0191	0.0857	1.5	0.4
3	3.38	0.0121	0.0544	1.3	0.6
4	3.38	0.0092	0.0411	0.7	0.9
5	3.80	0.0220	0.0357	3.6	1
6	3.80	0.0119	0.0193	1.0	1.8
7	3.78	0.0092	0.0157	1.1	2.2
8	3.7	0.0060	0.0123	1.1	2.9
9,10,11	3.92	0.0078	0.0094	1.8	3.7
12	3.79	0.0063	0.0104	4.1	3.4
13	3.88	0.0092	0.0123	2.3	2.9
14	3.89	0.0078	0.0102	1.5	3.4
15	3.8	0.0093	0.0151	10.4	2.3

Table 6: Protection factor measurements for WK$_5$-Inp$_2$^{t}L-A$_{15}$-^{t}LInp$_2$-K$_5$-NH$_2$ in 20 mM NaH$_2$PO$_4$/ D$_2$O. The experimental rate constant is listed, along with the corrected rate constant at pH 4.0. The % error in fit of the rate constant is listed. The average exchange rate constant for Ac-A$_3$-OH was used as the solvent exposed reference rate constant k$_{se}$.

Alanine Length n	^{15}N Labels at site A_i	pH	rate constant at exp. pH min^{-1}	rate constant at pH 4.0 min^{-1}	% Error	PF
5	3	4.01	0.03499	0.03415	1.7	1.0
7	4	3.46	0.01023	0.03774	1	0.9
9	6,7	3.34	0.00715	0.03527	0.9	1.0
11	7,8,9	3.50	0.00695	0.02329	1.7	1.5
13	8,9,10	3.51	0.00498	0.01629	0.9	2.2
15	9,10,11	3.92	0.00776	0.00942	1.8	3.7
17	10,11,12	3.50	0.00175	0.00587	0.6	6.0
19	11,12,13	3.50	0.00095	0.00319	2.2	11.0
21	12,13,14	3.49	0.00029	0.00100	3.1	35.3
23	13,14,15	3.49	0.00020	0.00070	4.6	50.1
15	14,15,16	3.52	0.00015	0.00048	2.4	73.0

Table 7: Protection factor measurements for WK$_5$-Inp$_2$^{t}L-A$_n$-^{t}LInp$_2$-K$_5$-NH$_2$ (n = 5 - 25) in 20 mM NaH$_2$PO$_4$ / D$_2$O. The experimental rate constant is listed, along with the corrected rate constant at pH 4.0. The alanine sites (i) with ^{15}N labels are listed. The % error in the fit of the rate constant is listed. The average exchange rate constant for Ac-A$_3$-OH was used as the solvent exposed reference rate constant k$_{se}$.

ANALYTICAL ULTRACENTRIFUGATION

Analytical ultracentrifugation experiments were performed at the Facility for the Study of Complex Macromolecular Systems at MIT funded by NSF-0070319.[11] Data were acquired using a Beckman Optima Xl-A centrifuge and a six-channel cell in a An-60 Ti analytical rotor.

The peptides in Table 8 were all found to be monomeric species by sedimentation equilibrium analysis[12] at 40,000 and 45,000 rpm. Samples were tested at μM and mM concentration to take into account the concentration of solutions used in the CD and t/c – NMR experiments. Accurate molecular weights ($\pm$ 10 %) were obtained for μM samples in 0.1M NaCl, consistent with the need to screen the charged lysine residues.[12] The consistent linear distribution of $\ln(c_r)$ vs. $r^2/2$ for μM and mM peptide samples, in water and 0.1M NaCl is indicative of monomeric species.

$$MW_{ap} = \frac{RT}{(r^2/2)(1-\rho)\omega^2} \ln(c_r)$$

Equation 2: Calculation of apparent MW from AUC data ($R \equiv 8.314 \times 10^7$ erg mol^{-1} K^{-1}, $T \equiv$ temperature in K, $c_r \equiv$ UV-determined concentration as a function radial position (r), $\rho \equiv$ solvent density at temp T, $\omega \equiv$ angular velocity)[13]

---

[11] This facility is maintained by the Imperiali group in the Chemistry department.

[12] Miller, J. S. "I. Efforts Towards the Synthesis of CP-263,114 II. On Solubilized, Spaced Polyalanines and Amino Acid α-Helix Propensities" Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 2001.

[13] Laue, T. M.; Stafford III, W. F. "Modern Applications of Analytical Centrifugation." *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 75-100; McRorie, D. K.; Voelker, P. J. *Self-Associating Systems in the Analytical Ultracentrifuge.* **1993**, Beckman Instruments, Palo Alto, CA; Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. "Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins." *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E.; Rowe, A. J.; Horton, J. C. Eds.), **1992**, pp 90-125, Royal Society of Chemistry, Cambridge.
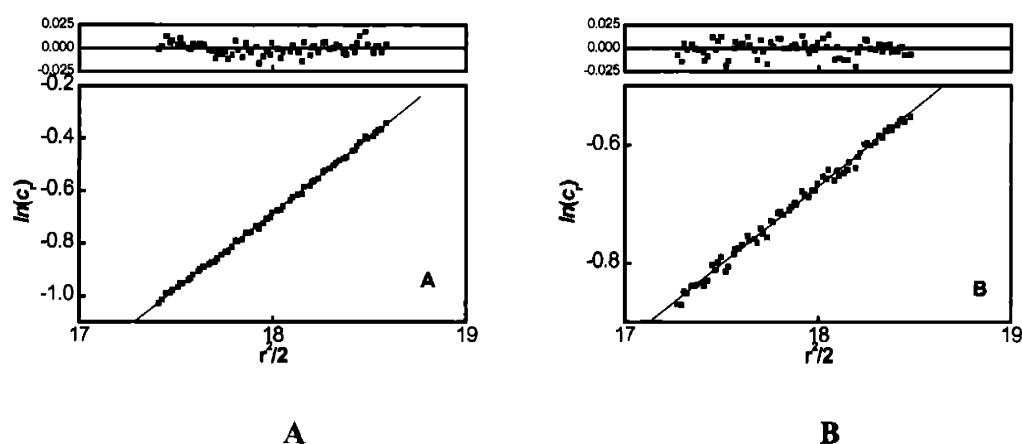
Figure 2: Sedimentation Equilibrium Analysis at 45,000 rpm for AcHel-A_{14}-$^fLInp_2K_4W$-NH_2: A) 20 μM peptide in 0.1M NaCl; B) 5 mM peptide in D_2O.

Peptide and Conditions	MW (D)	MW_{ap} by AUC (D)	% Error	R Value
AcHel-A_{14}-$^fLInp_2K_4W$-NH_2	**2327**			
mM, D_2O, 45k rpm		1256	-46	0.996
μM, H_2O, 45k rpm		1111	-53	0.996
μM, 0.1 M NaCl/ H_2O, 45k rpm		2405	3	0.999
Ac-$^\beta$D-Hel-A_8 -β-InpK_2W-NH_2	**1621**			
μM, H_2O, 45k rpm		1290	-20	0.995
μM, 0.1 M NaCl/ H_2O, 45k rpm		1509	-7	0.998
Ac-$^\beta$D-Hel- A_8 -β-NH_2	**1067**			
μM, 9:1 H_2O D_2O, 45k rpm		592	-45	0.995
μM, 9:1 H_2O D_2O, 45k rpm		601	-44	0.987
WK_5-Inp$_2^f$L-A_{17}-$^fLInp_2$-K_5-NH_2	**3108**			
mM, 0.20 mM NaH_2PO_4 /D_2O, 40k rpm		2928	-6	0.997
WK_5-Inp$_2^f$L-A_{23}-$^fLInp_2$-K_5-NH_2	**3534**			
mM, 0.20 mM NaH_2PO_4 /D_2O, 40k rpm		3352	-5	0.998
WK_5-Inp$_2^f$L-A_{27}-$^fLInp_2$-K_5-NH_2	**4146**			
mM, 0.20 mM NaH_2PO_4 /D_2O, 40k rpm		3667	-11	0.998

Table 8: AUC results for peptide at 2 °C.

LIFSON-ROIG MODELING

Lifson Roig calculations in Chapters 2 and 3 were performed by Prof. Dan Kemp.

Chapter 3 Cooperative LR

The recursion process uses the classical LR state sum as a template. Each of its terms and coefficients are unchanged in form, the sole change is replacement of each v^2w^j term, which weights helices of length $(j+2)$, by $v^2w(j)^j$; the constant w of the classical state sum is replaced by a length-dependent $w(j)$. Computational software Mathematica contains a function Coefficient that simplifies this transformation. For example: Let $P(v, w) = 1 + 3v + v^2(2w + w^2)$; then Coefficient[P, w, 0] = $1 + 3v$; Coefficient[P, w, 1] = $2v^2$; Coefficient[P, w, 2] = v^2. As a result, Coefficient[P, w, 0] + $w(1)$Coefficient[P, w, 1] + $w(2)^2$ Coefficient[P, w, 2] = $1 + 3v + v^2(2w(1) + w(2)^2)$ which is the desired new form of P: P(v, $w(1)$, $w(2)$)]. In general, the symbolic transformation expression for converting the conventional LR state sum SS into a recursive state sum SS' is: given by SS' = Coefficient[SS, w, 0] + Sum[$w(j)^{(j-2)}$ Coefficient[SS, w, j], $\{j, 3, n\}$], where for example Sum[j^2, $\{j, 1, 4\}$] = $1 + 4 + 9 + 16$.

Chapter 2: Lifson-Roig (LR) Algorithms for Fits to t/c, CD, and Model data

Rather than conventional 3x3 or 4x4 matrices, these calculations use the 8x8 matrices that we have introduced previously.[14] All these give equivalent polynomial state sums, with two major differences. First, the state weight for a helical conformation of length k is t^2w^k, and second, in the initial state sum calculation, a conventional v

[14] Miller, J. S.; Kennedy R. J.; Kemp, D. S. "Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid α-Helix Propensities" *J. Am. Chem. Soc.* 2002, *124*, 945-962.

weighting is used for isolated helical conformations, but a t^2 weight is used for helices. At

the evaluation stage of the calculation, we set $t = v = 0.048$, according to common

usage.[15] The double t, v weighting system for helix initiation allows easy partitioning of

the state sum; for example χ_{Hel-k}, the mole fractions of conformations containing

uninterrupted helices of length k, is $(1/SS)$(sum of all polynomial terms containing $t^2 w^k$ as

a coefficient). The partitioning of state sums into such terms is rapid and efficient with

modern PC-based computational algorithms.

The ratio t/c is modeled as in 3), and the t/c ratio intrinsic to the template Ac-Hel,

ts/cs, is set equal to a peptide-invariant parameter $A = 0.76$, as previously modeled.[16]

The template initiation parameter B must be multiplied by the ratio SS_{te}/SS_{cs} of LR

peptide state sums.

3) $t/c = (ts + te)/cs = ts/cs + te/cs = A + B \ SS_{te}/SS_{cs}$

The state sum SS_{cs}, is a normal peptide helical state sum, modified slightly from

that reported previously, [17] $aa \cdot m^{(n-1)} \ mc \cdot ab$, where $aa = (0,0,0,0,0,0,1,1)$, $ab = $

$(0,1,0,1,0,1,0,1)$, and m is the matrix with nonzero elements as follows: $(1,1) = w$; $(1,2) = $

$u \ w$; $(2,3) = 1$; $(2,4) = 1$; $(3,5) = u \ w$; $(3,6) = v$; $(4,7) = 1$; $(4,8) = 1$; $(5,1) = w$; $(5,2) = $

$v^2/(u \ w)$; $(6,3) = 1$; $(6,4) = 1$; $(7,5) = u \ w$; $(7,6) = v$; $(8,7) = 1$; $(8,8) = 1$. Matrix mc

introduces a capping parameter for the ${}^{l}L$ residue that terminates the polyalanine

[15] Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. "Kinetics of Amide Proton Exchange in Helical Peptides of Varying Chain Lengths. Interpretation by the Lifson-Roig Equation" *Biochemistry* **1992**, *31*, 1263-1299.

[16] Kemp, D. S.; Oslick, S. L.; Allen, T. J. "The Structure and Energetics of Helix Formation by Short Templated Peptides in Aqueous Solution. 3. Calculation of the Helical Propagation Constant s from the Template Stability Constants t/c for Ac-Hel$_1$-OH, $n = 1$-6" *J. Am. Chem. Soc.* **1996**, *118*, 4249-4255.

[17] Wallimann, P.; Kennedy R. J.; Kemp, D. S. "Large Circular Dichroism Ellipticities for N-Templated Helical Polypeptides Are Inconsistent with Currently Accepted Helicity Algorithms" *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1290-1292

sequence. As calculated from data reported elsewhere,[18] $c = 0.77$ for measurements at 2 °C, and $c = 1.0$ for measurements at 25 °C. With the exception of the term $(1,2) = u\,w\,c$, all terms of mc equal the corresponding terms of m.

$$
\mathbf{m} = \begin{bmatrix}
w & uw & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & uw & v & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & \dfrac{v^2}{uv} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & uw & v & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\qquad
\mathbf{mc} = \begin{bmatrix}
w & uwc & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & uw & v & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
w & \dfrac{v^2}{uv} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & uw & v & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
$$

The te state sum SS_{te} is calculated from the matrix product $aa \cdot n \cdot nn \cdot m^{(n-3)}\, mc \cdot ab$, where the nonzero coefficients of n are: $(7,5) = w$; $(7,6) = w$; $(8,7) = 1 = (8,8)$, and the nonzero coefficients of nn are $(5,1) = w/u$; $(5,2) = w$; $(6,3) = 1 = (6,4)$, $(7,5) = u\,w$; $(7,6) = v$; $(8,7) = 1 = (8,8)$. These coefficients were determined by the weightings of helical substates for each peptide length n. As an example, the sixteen weightings for $n = 4$ are as follows: hhhh: $w^4\,c$; ;hhhc: w^3; chhh: t^2w^3c; hhch: w^2v; hchh: wv^2; hhcc: w^2; (hchc and hcch): $2w\,v$; (chhc, chch, cchh): $3v^2$; hccc: w; (chcc, cchc, ccch), $3v$, cccc, 1.

---

[18] Maison, W.; Kennedy, R. J.; Miller, J. S.; Kemp, D. S. "C-Terminal Helix Capping Propensities in a Polyalanine Context for Amino Acids Bearing Nonpolar Aliphatic Side Chains" *Tetrahedron Lett.* **2001**, *42*, 4975-4977.

$$m = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w & w & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad mc = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dfrac{w}{u} & w & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & uw & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The above LR state sums were transformed into state sums for the H-bonding cooperativity model as follows: Each noncooperative LR state sum is a polynomial in v, t and m, in which v is the weighting used for the helical residues in nonhelical conformations chc and chhc, t is the helix initiation parameter, and w is the alanine helical propensity. For the state sum SS_{cs}, each term within the polynomial containing the coefficient of form $t^2 w^k$ corresponds to a conformation that contain a single continuous helical sequence of length k. For the cooperative state sum SS_{te}, all w^k terms that lack a t^j coefficient ($j = 2, 4,$ or 6) correspond to helices that extend to the peptide N-terminus and are initiated by Ac-Hel. Applied to a polynomial y, the Mathematica® 3.0 function Coefficient [y, t, 2] yields the coefficient for the term in y that contains t^2. Accordingly, 4) converts the state sum SS_{cs} into its cooperative form, and 5) effects the corresponding conversion for SS_{te}. The effect of these transformations is to convert every w^k term in the original state sum into a coop[k] w^k term. (Note that t^4 and t^6 terms are dropped in these calculations, since for plausible values of w, with $t = v = 0.048$, and with $n \le 14$, these terms make insignificant contributions to the state sum.) The function coop[k] was assigned the value 1 for $1 \le k \le 4$, and $(w\text{·})^k$ for $7 \le k$; coop[5] = $(w\text{·})^{(5*1)/3}$; coop[6] = $(w\text{·})^{(6*2)/3}$

4) $SS_{cs}Coop$ = Coefficient[SS_{cs}, t, 0] + t^2 Sum[coop[k] w^k

      Coefficient[Coefficient[SS_{cs}, t, 2], w, k],$\{k, 3, n\}$]

5) $SS_{te}Coop$ = Coefficient[SS_{te}, w, 0] + Sum[coop[k] w^k Coefficient[Coefficient[SS_{te}, t,

      0], w, k],$\{k, 1, n\}$] + ^{t2} Sum[coop[k] w^k Coefficient[Coefficient[SS_{te}, t, 2], w, k],$\{k$,

      $3, n\}$]

Since t and w are used in iteration and symbolic functional transformations, they are only assigned numerical values when these transformations are complete. The parameter v can be assigned its literature value $v = 0.048$ at the calculation of the state sum. At then end of symbolic transformations, t is assigned the literature value of $v = u = 0.048$ and w is assigned the value 1.0.

6)   ellcs = $(u^2/(n*SScsCp))$Sum[k*theta[k] w^k Coefficient [Coefficient

      [$SScsCp,u,2$],w,k],$\{k,3,n\}$]

7)   ellte = $(u^2/(n*SSteCp))$Sum[k*theta[k] w^k Coefficient

      [Coefficient[$SSteCp,u,2$],w,k],$\{k,3,n\}$] + $(1/(n*SScsCp))$Sum[k*theta[k] w^k

      Coefficient[Coefficient[$SSteCp,u,0$],w,k],$\{k,1,n\}$]

8)   ell$_{Calc}$ = te ellte + (1-te)ellcs

Residue ellipticities for the te and cs conformational substates are calculated from equations 6) and 7); the mole fraction of the te substate is calculated from equation 3) as t/c = A + B ($SS_{te}Coop/SS_{cs}Coop$), and the mole fraction of te state χ_{te} = (t/c − A)/(1 + t/c); the calculated overall ellipticity is then given by equation 8), since 1 = χ_{cs} + χ_{ts} + χ_{te}. In calculating the overall length of the polyalanine region, we regard the ^{f}L residue as a helix

stop signal and count the Ac-Hel length as zero. As noted by Schellman et al[19] and Woody et al,[20] the $n\pi^*$ transition of the tertiary proline amides are expected to fall in the range of 230 nm and contribute weakly to the 222 nm CD transition. FH values were calculated analogously by setting the theta[i] terms in 6) and 7) equal to 1.0 and redefining ellcs, ellte and ellcalc respectively as FHcs, FHte, and FHcalc.

The least squares minimization of $((t/c)_{Exp} - (t/c)_{Exp})^2$ was carried out using 2 °C and 25 °C equations 1), 2), and 3), treating B and w' as variable parameters for each temperature. Gratifyingly the values of B at the two temperatures are identical (0.177, 0.176) and consistent with previously assigned literature values.

In units of deg-cm^2 dmol^{-1} res^{-1} theta[k] values used in equations 5) and 6) for $k > 9$ are: $-61,000(1 - 2.5/k)$, and $-53,000(1 - 2.5/k)$ for $k \leq 9$.

[19] Qian, H.; Schellman, J. A. "Helix-Coil Theories: A Comparative Study for Finite Length Polypeptides" *J. Phys. Chem.* **1992**, *96*, 3987-3998.

[20] Manning, M. C.; Woody, R. W. "Theoretical CD Studies of Polypeptides Helices: Examination of Important Electronic and Geometric Factors" *Biopolymers* **1991**, *31*, 569-586.