Hydrogen Bonding in Aqueous Solution and Peptide Helicity Studied Using Model Systems

by

Evan T. Powers

B.A. Chemistry, Cornell University, 1992

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

June 1999

Signature of Author _____

Department of Chemistry
May 18, 1999

Certified by _____

Daniel S. Kemp
Professor of Chemistry
Thesis Supervisor

Accepted by _____

Dietmar Seyferth
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 Lawrence J. Stern _____

Professor Bruce Tidor _____

Hydrogen Bonding in Aqueous Solution and Peptide Helicity Studied Using Model Systems

by
Evan T. Powers

Submitted to the Department of Chemistry at the
Massachusetts Institute of Technology on May 18, 1999 in partial fulfillment of the
requirements for the degree of Doctor of Philosophy

Abstract

Two subjects are addressed in this thesis. The first is the energetics of hydrogen bonding in aqueous solution. Several types of hydrogen bond donors are placed near the acetamide carbonyl in variants of the reporting conformational template $AcHel_1$. The resulting perturbations of the system's conformational equilibria are used to measure the strength of the donor-acetamide hydrogen bond. Although charged donors interact most strongly with the acetamide, this is attributed mostly to charge-dipole attraction. The strongest hydrogen bonds, with energies up to −2 kcal/mol, are formed by amide NHs. Those formed by other donors, such as alcohols or ammonium ions, have energies around 0 kcal/mol. Implications for protein folding are discussed.

The second subject addressed is peptide helicity. The ribonuclease C-peptide's helicity is investigated to test the novel divided peptide method, in which a helical peptide is divided into fragments, the fragments' helical tendencies are assessed through their conjugates with $AcHel_1$, and this information is synthesized to yield a description of the full peptide's helicity. The results from this method are consistent with a known i to i+8 glutamate / arginine interaction enhancing the C-peptide's helicity. Evidence for previously unknown i to i+1 interactions is found as well.

The new template SO_3Hel_1, in which a sulfamate ($N\text{-}SO_3^-$) replaces $AcHel_1$'s acetyl, is prepared. SO_3Hel_1 is a moderately better helix initiator than $AcHel_1$, but an expected helix-stabilizing interaction between the putative helix dipole and the sulfamate's charge is not observed. The helicities of SO_3Hel_1-peptide conjugates are studied by the amide hydrogen exchange technique, where fractional site helicities are determined by comparing the exchange rate constants of a helical peptide's amides to their intrinsic exchange rate constants. Intrinsic exchange rate constants calculated from a literature model are not accurate enough for these studies. Fractional site helicities for eight SO_3Hel_1-peptide conjugates are therefore calculated using intrinsic exchange rate constants measured in corresponding non-helical $HHel_1$-peptide conjugates. It is shown that the helices induced by SO_3Hel_1 are frayed increasingly toward the C-terminus, and that interactions not considered in current peptide helicity prediction algorithms significantly influence peptide helicity.

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

Acknowledgements

I have walked this Earth for eight and one score years, and have spent a full quarter of that time in graduate school. By daily reinforcement during this rather lengthy undertaking, I have come to understand the clichéd truth about the scientific life: the Universe is wholly unsympathetic toward our little efforts to understand it; comfort and cheer come from the people in one's life. In this section, I would like to thank the people who have comforted and cheered me over the last seven years.

The work in this thesis was carried out in the laboratories of Professor D. S. Kemp. I thank Professor Kemp for the effort he made to be a mentor, rather than just a graduate adviser. He has provided my funding and a fertile environment in which to work, and even an apartment in which to live during my last year. Infinitely more valuable than any of this, however, he has imparted to me a way of thinking about and doing science that has profited me enormously.

I am fortunate for having had talented and agreeable co-workers throughout my graduate career. I thank all of the members of the Kemp group with whom I have been associated, but I am particularly grateful to the following: Jon Zerkowski (rappin' Jonny Z. my first and only baymate, whose tolerance for my taste in music bordered on the saintly), Jeff Rothman (Transdude), Dave Chalfoun, Janette Lee, Linda Shimizu (Disco Grrrl), Sherri Oslick (the Queen of all Baked Goods), Kim McClure (who worked with me on the project described in chapter 1), Sam Tsang (who worked with me on the project described in chapter 3), Peter Renold (who started the project described in chapter 4), Lawrence Williams (the only other member of the group who knew what a manticore was), Songpon Deechongkit, Justin Miller, William Shalongo, and finally Robert Kennedy, Eva Arce, and Peter Rudolf (my current lunching companions, and each of whom proofread portions of this thesis). I also thank Dr. Sue Pochapsky (who measured the critical ROESY spectrum for chapter 4), Dr. Jeff Simpson (for his help with the department spectrometers), and my Father, Professor David Powers (without whose help the analysis in chapter 5 would have been impossible, but more on him later).

The core of my social life at MIT was formed within my first year. During this time, I met Jeff Eckert (the crafty Eck), Dave Wang (who introduced me to volleyball and, less

successfully, to downhill skiing), Bill Kobertz (my fellow fibula-breaker), Martha Rook and John Bearly (through whose generosity we all spent several weekends in a pleasant cabin in Vermont), Dennis Hall (the king of the mellow), and Ben Turk (the only person I know who's had a single played on WFNX, and my apartment-mate for three years- I still owe Ben big for dealing with our frozen toilet while I was out gallivanting in Mexico). Happy additions were made to this core all through the rest of my time here. I had the great good luck of knowing Matt and Becky Bogyo (the Boge and Meggy, my apartment-mates for two years, with whom I enjoyed many evenings watching such artistic triumphs as "Busted on the Job 2" and "Happy Gilmore"), Becky Carazza (I daresay the most energetic person I've ever known), Jill Mello (who introduced me to the 5:00 cookie break), Noah Snyder (another apartment-mate along with Matt and Becky), Marjie Solomon, Seble Wagaw, Bob Murry, Silvia Hoehn, Dana Buske, and I'm sure I'm forgetting others who deserve mention. To paraphrase Lou Grant from the series finale of "The Mary Tyler Moore Show", I treasure these people.

My fixed point over the last year and a half has been Christy Esau. I thank Christy for taking me hiking and to Red Sox games, for deciphering "Myst" and "Riven" with me, and for innumerable other kindnesses. To quote a song (roughly), when I have needed her she has been nearby, whether I was mountain high or valley low and, to quote another song, I am glad for what we've had and what's to come. (I should also thank Christy's apartment-mates, Erin Gray, Tara Schmidt, and Rebecca Horne for tolerating my frequent presence in their apartment.)

It has been said that the existence of a great disproportion between one's intentions and the reality that one is likely to encounter is a demonstration of the absurd. In this sense, it is absurd for me to try to conjure the words in this space that I need to thank my family for all of their love and support. I would probably need another thesis, but here is my attempt. I thank my sister Emily for her empathy. I thank my brother Chris for his exhortations to not fall into the "barbarism of specialization". I thank my mother Lya and my father David for my life and upbringing. Duly noting how these words fall short, I dedicate this thesis to them.

To Mom, Dad, Chris, and Emily

"Music, states of happiness, mythology, faces belabored by time, certain twilights and certain places try to tell us something, or have said something we should not have missed, or are about to say something..."

Jorge Luis Borges

Preface

Four separate projects are reported in this thesis, the common theme of which is the use of simple model systems to study phenomena that usually occur in complicated settings. The project reported in chapter 1 revisits hydrogen bonding in aqueous solution to clarify some unresolved issues, especially as regards protein folding. The remaining chapters address peptide helix formation, a subject that has been well studied but about which there remain many unanswered questions. Two new tools are introduced. The first, reported in chapter 3, uses the templated systems that have been studied for years in the Kemp group. The second, reported in chapter 5, introduces new rigor into the amide hydrogen exchange technique for measuring fractional site helicities to provide insight into the scope of context dependence of residue helix propensities. This project requires a new template for helix initiation, whose preparation and characterization are reported in chapter 4. The background necessary for these three projects on peptide helicity is given in chapter 2.

Table of Contents

Chapter 1. A Determination of the Bounds on Intrinsic Hydrogen Bonding Energies in Aqueous Solution

1.1 Introduction

The idea that the hydrogen bond can mediate substantial and specific interactions has proven to be enormously useful. Hydrogen bonding has been used to explain phenomena from the properties of water[1,2,3] to the crystal structure of ammonium chloride,[4] from the tautomerism of acetoacetate esters[5] to the structures of the biological macromolecules.[6,7,8,9] The ubiquity and importance of the hydrogen bond is demonstrated by the prolific publication on the subject. The first comprehensive treatise on hydrogen bonding,[10] published in 1960, included over two thousand references; it has been estimated[11] that as of 1991 over twenty thousand publications had appeared.

Yet, significant issues pertaining to hydrogen bonding have not been settled. Especially controversial are the thermodynamics of hydrogen bonding between two solutes in aqueous solution. This is a crucial question because of its relevance to biological processes; it is impossible to understand protein folding or protein-ligand

---

[1] Latimer, W. M.; Rodebush, W. H. *J. Am. Chem. Soc.* **1920**, *42*, 1419.
[2] Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933**, *1*, 515.
[3] Huggins, M. L. *J. Phys. Chem.* **1936**, *40*, 723.
[4] Huggins, M. L. *Phys. Rev.* **1922**, *19*, 346.
[5] Huggins, M. L. *J. Org. Chem.* **1936**, *1*, 407.
[6] Huggins, M. L. *Chem. Rev.* **1943**, *32*, 195.
[7] Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1936**, *22*, 439.
[8] Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205.
[9] Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
[10] Pimentel, G. C.; McClellan, A. L. *The Hydrogen Bond*; W. H. Freeman & Co: San Francisco, London, **1960**.
[11] Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, **1991**.

association at a causal, chemical level without understanding the contribution of hydrogen bonding to these events. This thesis introduces a new method for measuring the free energies of hydrogen bond formation in water between amide carbonyls and several biologically relevant hydrogen bond donors.

1.2 Background

A brief summary of the occurrence, geometry, and thermodynamics of hydrogen bonding is appropriate. Experimental results are emphasized, but the theory of hydrogen bonding is briefly discussed at the end of section 1.2.1, and computational investigations of hydrogen bonding thermodynamics are not reviewed.

1.2.1 Occurrence and Geometry of Hydrogen Bonds

Hydrogen bonds are interactions that occur between a donor, D-H, that has a hydrogen covalently bound to an electronegative atom, and an acceptor, A, that has a free lone pair. As established by crystallography,[11] functional groups that are known to donate hydrogen bonds include O-H groups (in water, alcohols, and carboxylic acids), N-H groups (in amines, amides, and heterocycles), N^+-H groups (in ammonium ions or protonated heterocycles), S-H groups (in thiols), and sometimes even C-H groups. Functional groups that are known to accept hydrogen bonds include -O- groups (in water, alcohols, and ethers), $-O^-$ groups (in carboxylates, phosphates, and sulfates), =O groups (in

carboxylic acids, esters, and amides), -N< groups (in amines and heterocycles), and -S- groups (in thiols and thioethers).

Hydrogen bond formation is accompanied by characteristic changes in NMR and IR spectra.[10] The hydrogen atom of the donor is deshielded upon hydrogen bond formation, with changes in chemical shifts of several ppm being commonly observed in NMR spectra. The peak in IR spectra due to stretching of the D-H bond shifts to lower frequencies by up to 300 cm^{-1}, broadens substantially, and becomes more intense. Changes also occur in the D-H bending modes but these are more difficult to identify.

Hydrogen bonds have strong geometric preferences compared to other non-covalent interactions, but they still tolerate more variation in bond distances and angles than covalent bonds. As revealed by surveys of crystal structures,[11] the hydrogen bond length, defined as the distance between the donor H atom and the acceptor, is usually between 1.5 and 2.5 Å (less than the sum of the van der Waals radii of the two atoms), with the stronger bonds tending to be shorter (Badger's rule[12]). The angle formed by the D, H, and A atoms (∠DHA) can be quite variable, but is most often slightly sub-linear, typically between 160° and 170°.

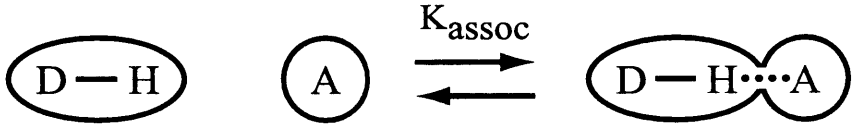
The first model for the hydrogen bond treated it as an electrostatic interaction between the partial positive charge on the donor hydrogen and the partial negative charge on the acceptor atom.[13] Although it is now recognized that factors such as dispersion, induction, and charge transfer have to be considered in addition to electrostatics for a complete understanding of hydrogen bonding,[14] it is clear that electrostatics are by far the most important force in hydrogen bonding. This is demonstrated by the success of extended

[12] Badger, R. M. *J. Chem. Phys.* **1934**, *2*, 128.
[13] Pauling, L. *Proc. Natl. Acad. Sci. USA* **1928**, *14*, 359.

electrostatic models that include multipole interactions in predicting the geometry of hydrogen bonded complexes.[15]

1.2.2 Thermodynamics of Hydrogen Bond Formation in the Gas Phase

Figure 1. Hydrogen bond formation in the gas phase, where DH is the hydrogen bond donor and A is the hydrogen bond acceptor.



The gas phase is the simplest medium in which hydrogen bonding can be studied. Effects like solvation do not complicate the gas phase association between a hydrogen bond donor and an acceptor (illustrated in figure 1), so that the enthalpy change for this reaction is a direct measure of hydrogen bond strength. Such gas phase association enthalpies have been measured for a number of complexes, a selection of which is presented in table 1. These enthalpies are large, especially for the cases in which the donor or the acceptor is charged, where the hydrogen bond enthalpy can be as much as 30% of the enthalpy of covalent carbon-hydrogen bonds.[16] It should be noted, though, that other gas phase non-hydrogen bonding interactions also have large enthalpies. For example, the polar interaction between ethyl acetate molecules results in a dimerization

[14] Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, **1996**.
[15] Buckingham, A. D.; Fowler, P. W. *J. Chem. Phys.* **1983**, *79*, 6426.
[16] *CRC Handbook of Chemistry and Physics, 72nd edition*; Lide, D. R., Ed. CRC Press: Boca Raton, Ann Arbor, Boston, **1991-1992**.

enthalpy of −4.5 kcal/mol.[17] This notwithstanding, it is clear that hydrogen bonds are exceptionally strong non-covalent interactions. If entropy (which is usually between 20 and 30 cal/mol °K for gas phase associations)[17,18,19] were not present as a balancing force, the K_{assoc} for two gas phase water molecules at room temperature would be about 1.5×10^{4}; for a methyl ammonium ion and water, it would be 3.1×10^{13}. These exercises emphasize how the net gain of a hydrogen bond can favor the product of an association reaction. One might thus anticipate that, in solution, where association entropies should be less than in the gas phase, hydrogen bond formation could strongly promote associations. That is, so long as solvation effects do not overly attenuate hydrogen bond enthalpies.

Table 1. Gas phase enthalpies of formation for various hydrogen bonds.

D-H	A	ΔH (kcal/mol)	Method
NH_3	NH_3	-4.4	deviation from the principle of corresponding states
H_2O	H_2O	-5.7	
CH_3COOH	CH_3COOH	-8.0	
H_2O	CH_3COO^-	-16.0	ion cyclotron resonance
$CH_3NH_3^+$	H_2O	-18.4	
$CH_3NH_3^+$	CH_3NH_2	-21.7	
$CH_3NH_3^+$	$HCONH_2$	-30.0	

[17] Lambert, J. D. *Discussions Faraday Soc.* **1953**, *15*, 226.
[18] Meot-Ner, M. *J. Am. Chem. Soc.* **1984**, *106*, 1257.
[19] Meot-Ner, M.; Sieck, L. W. *J. Am. Chem. Soc.* **1986**, *108*, 7525.

1.2.3 Thermodynamics of Hydrogen Bond Formation in Solution with Non-Competitive Solvents

Figure 2. Hydrogen bond formation in solution where the solvent is non-competitive.

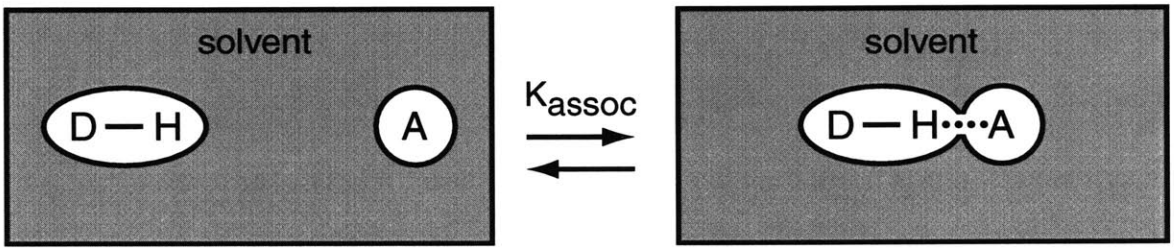


Figure 2 depicts hydrogen bond formation in a non-competitive solvent, that is, a solvent that has no capacity to donate or accept hydrogen bonds such as a liquid alkane or CCl_4. In such a solvent, one would not expect any strong or specific solvent-solute interactions to affect hydrogen bond formation. Although the higher dielectric constants of non-competitive solvents ($\varepsilon \approx 2.0$ for liquid alkane and 2.2 for CCl_4[16]) might influence the electrostatic contributions to the enthalpy of hydrogen bonding, the net gain of a hydrogen bond should still favor the associated state as it does in the gas phase. Furthermore, the entropy of association should be less unfavorable since translation is more restricted in solution than it is in the gas phase,[20] so for many cases one would expect K_{assoc} in non-competitive solvent to be considerable. For hydrogen bonds between neutral species this turns out to be true. Equilibrium constants have been measured for a vast number of complexes encompassing a huge variety of donors and acceptors using IR and NMR spectroscopy to detect and quantify complex formation.

The first systematic hydrogen bonding studies in non-competitive solvents were carried out using calorimetry in tandem with IR and NMR spectroscopy to measure the association constants in carbon tetrachloride of a series of hydrogen bond acceptors with *p*-fluorophenol as the common donor,[21,22,23,24] and subsequent research extended the database of known association constants to other donors and acceptors.[25] This body of data has been used to construct scales of hydrogen bond donor and acceptor strengths. The two most extensive are the α_2^H, β_2^H scale[26] and the C_a, C_d scale[27] (α_2^H and C_d measure hydrogen bond donor strength, β_2^H and C_a measure hydrogen bond acceptor strength). Table 2 lists representative α_2^H and β_2^H values.

Table 2. Hydrogen bond donor and acceptor strengths in carbon tetrachloride (larger numbers indicate stronger donors or acceptors).

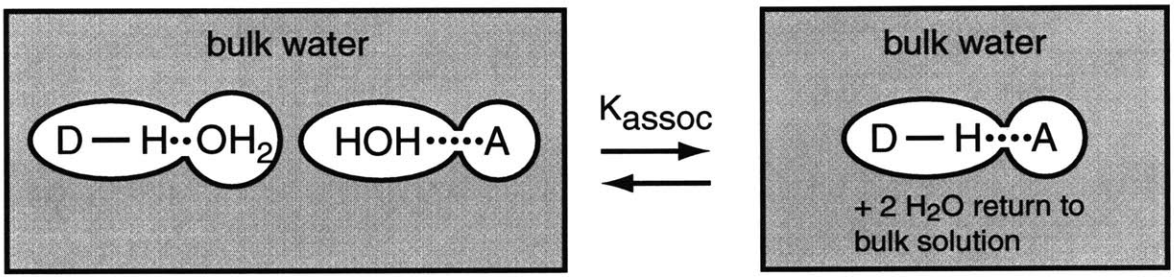
Compound	α_2^H	β_2^H
CH_3OH	0.43	0.47
C_6H_5OH	0.60	0.22
CH_3COOH	0.61	0.44
$(CH_3)_2CO$	-	0.50
CH_3NH_2	0.16	0.58
$CH_3CONHCH_3$	0.40	0.72
CH_3CH_2SH	0.0	0.24
$(CH_3)_2S$	-	0.29

[20] Doig, A. J.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 338
[21] Arnett, E. M.; Murty, T. S. S. R.; Schleyer, P. v. R.; Joris, L. *J. Am. Chem. Soc.* **1967**, *89*, 5955.
[22] Gurka, D.; Taft, R. W.; Joris, L.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1967**, *89*, 5958
[23] Taft, R. W.; Gurka, D. *J. Am. Chem. Soc.* **1969**, *91*, 4794.
[24] Arnett, E. M.; Joris, L.; Mitchell, E.; Murty, T. S. S. R.; Gorrie, T. M.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1970**, *92*, 2365.
[25] Kamlet, M. J.; Abboud, J.-L.; Taft, R. W. *Prog. Phys. Org. Chem.* **1981**, *13*, 485 and references therein.
[26] Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.
[27] Raevsky, O. A.; Grigor'ev, V. Y.; Kireev, D. B.; Zefirov, N. S. *Quant. Struct.-Act. Relat.* **1992**, *11*, 49.

These values demonstrate that hydrogen bond donor and acceptor strengths are not necessarily related to acid and base strengths. For instance, α_2^H (phenol) $\approx$ α_2^H (acetic acid) even though phenol is much less acidic than acetic acid, and β_2^H (N-methylacetamide) > β_2^H (methylamine) even though methylamine is much more basic than N-methylacetamide. The difference in pK_a between two compounds is predictive of their difference in hydrogen bonding donor or acceptor strength only when they belong to the same family of compounds (e.g. carboxylic acids, phenols, amines, etc.).

1.2.4 Thermodynamics of Hydrogen Bond Formation in Aqueous Solution

Figure 3. Hydrogen bond formation in aqueous solution.



Hydrogen bond formation in aqueous solution is shown in figure 3. Unlike non-competitive solvents, water interacts strongly with solutes by accepting hydrogen bonds from donors[28] and donating hydrogen bonds to acceptors.[29,30] Thus, it is not true in water as it is in the gas phase and non-competitive media that there is a net gain of a hydrogen

[28] Eaton, G.; Symons, M. C. R.; Rastogi, P. P. *J. Chem. Soc. Faraday Trans. 1* **1989**, *85*, 3257.
[29] Eaton, G.; Symons, M. C. R. *J. Chem. Soc. Faraday Trans 1* **1988**, *84*, 3459.

bond driving the equilibrium to favor the association products. When solutes associate, they are merely exchanging hydrogen bonds with water for hydrogen bonds with each other. The enthalpy of hydrogen bond formation between D-H and A is spent in displacing the solvating water molecules, which are then free to join the bulk solvent, with whatever enthalpy and entropy changes attend that process. If there is to be a favorable enthalpy change upon hydrogen bond formation, it must result from a selectivity of D-H and A for each other as hydrogen bonding partners over water.[31,32] This selectivity is a second order effect, dependent on the relative strength of different types of hydrogen bonds, and in most cases will be easily overcome by the abundance of water in aqueous solution. Liquid water is 110 M in hydrogen bond donor and acceptor sites, so that by Le Chatelier's principle, bimolecular hydrogen bond equilibria are generally shifted toward the solvated reactants.

Two unique problems exist with hydrogen bonding studies in water. First, the low abundance of hydrogen bonded complexes in aqueous solution makes them difficult to detect. Second, the techniques used to study hydrogen bond formation in non-competitive media are less telling in water.[33] These techniques detect the presence or absence of a hydrogen bond. In water, hydrogen bonds are nearly always present; the relevant issue is the abundance of one hydrogen bonded form in the presence of another. These factors have conspired to make the thermodynamics of hydrogen bonding in water difficult to characterize. Nevertheless, much effort has gone into the study of the thermodynamics of aqueous hydrogen bonding, primarily because of its relevance to

---

[30] Eaton, G.; Symons, M. C. R.; Rastogi, P. P.; O'Duinn, C.; Waghorne, W. E. *J. Chem. Soc. Faraday Trans.* **1992**, *88*, 1137.
[31] Hine, J. *J. Am. Chem. Soc.* **1972**, *94*, 5766.
[32] Stahl, N.; Jencks, W. P. *J. Am. Chem. Soc.* **1986**, *108*, 4196.

biological phenomena. The approaches that have been used can be roughly divided into two categories: the direct method and the incremental energy method.

1.2.4a The Direct Method for Studying Aqueous Hydrogen Bonding

In the direct method for studying aqueous hydrogen bonding, the thermodynamics of association equilibria are measured and attributed directly to hydrogen bond formation. This is essentially the same approach as that used to study hydrogen bonding in non-competitive media, differing mostly in the ease with which complex formation can be quantified.[33] The results from a number of direct method studies are summarized in table 3. This method is typically used to study intermolecular hydrogen bonding between small molecules, although entries 5 and 7 report cases in which intramolecular hydrogen bonding was studied.

---

[33] Moon, A. Y.; Poland, D. C.; Scheraga, H. A. *J. Phys. Chem.* **1965**, *69*, 2960.

Table 3. Hydrogen bond parameters from the direct method (all energies in kcal/mol; a dash indicates that the quantity was not determined).

System	Donor	Acceptor	Technique	ΔH_{hbond}	ΔG_{hbond}	K_{assoc}
urea[34,35]	amide NH	amide CO	heats of solution	−1.5	-	0.042
N-methyl acetamide[36]	amide NH	amide CO	overtone IR	0	-	0.005
δ-valerolactam[37]	amide NH	amide CO	overtone IR	−2.8	-	-
diketopiperazine[38]	amide NH	amide CO	heats of dilution	−2.1	-	0.06
salicylic acid[39] (intramolecular)	phenol OH	carboxylic acid CO	UV and potentiometric titration	−4.9	−1.7	-
	phenol OH	COO^-		−5.0	−3.8	-
phenol-acetate[33]	phenol OH	COO^-	fluorescence	−0.4	-	0.5
hydroxy ethers[40] (intramolecular)	alcohol OH	ether O	coupling constant analysis	-	−0.5	-
phenoxide/ $(CH_2NH_3^+)_2$[32]	NH_3^+	phenoxide	UV	-	-	0.81
formic acid/formate[41]	carboxylic acid OH	COO^-	buffer catalysis	-	-	0.25
formic acid[42]	carboxylic acid OH	carboxylic acid CO	titration	0	-	-

ΔH_{hbond} is negative for most of the cases of intermolecular hydrogen bond formation, indicating that some hydrogen bond donors and acceptors are able to discriminate for each other over water.[43] However, in all these cases K_{assoc} is less than 1, indicating that

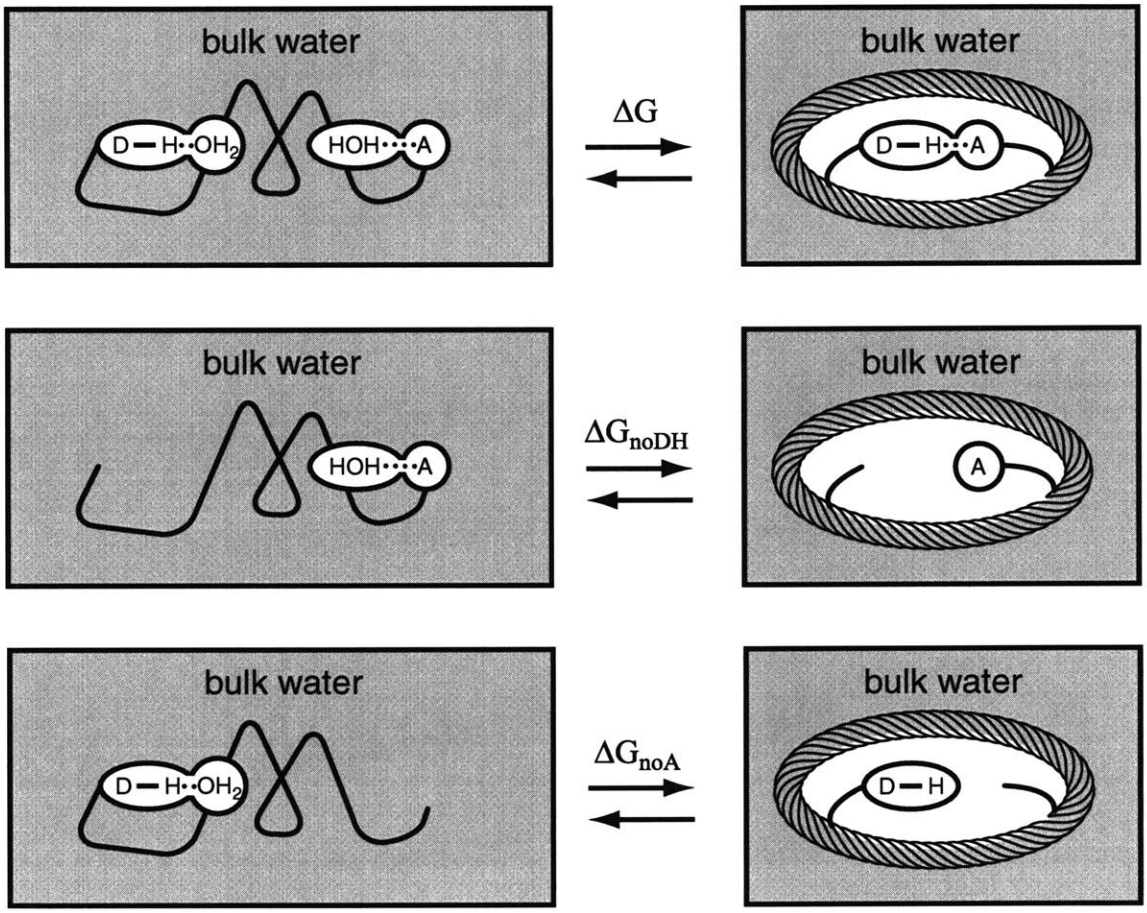
[34] Schellman, J. A. *Compt. rend. trav. lab. Carlsberg* **1955**, *29*, 223.
[35] Kresheck, G. C.; Scheraga, H. A. *J. Phys. Chem.* **1965**, *69*, 1704.
[36] Klotz, I. M.; Franzen, J. S. *J. Am. Chem. Soc.* **1962**, *84*, 3461.
[37] Susi, H.; Timasheff, S. N.; Ard, J. S. *J. Biol. Chem.* **1964**, *239*, 3051.
[38] Gill, S. J.; Noll, L. *J. Phys. Chem.* **1972**, *76*, 3065.
[39] Hermans, Jr. J.; Leach, S. J.; Scheraga, H. A. *J. Am. Chem. Soc.* **1963**, *85*, 1390.
[40] Beeson, C.; Pham, N.; Shipps, Jr. G.; Dix, T. A. *J. Am. Chem. Soc.* **1993**, *115*, 6803.
[41] Hand, E. S.; Jencks, W. P. *J. Am. Chem. Soc.* **1975**, *97*, 6221.
[42] Schrier, E. E.; Pottle, M.; Scheraga, H. A. *J. Am. Chem. Soc.* **1964**, *86*, 3444.
[43] It is remarkable, though, that estimates of the heat of formation of water-water hydrogen bonds in the liquid state are more negative (-3.5 to -4.5 kcal/mol) than those in table 3 (for examples see Scatchard, G.; Kavanagh, G. M.; Ticknor, L. B. *J. Am. Chem. Soc.* **1952**, *74*, 3715 and Nemethy, G.; Scheraga, H. A. *J. Chem. Phys.* **1962**, *36*, 3382).

this selectivity is not enough to make intermolecular complex formation favorable. ΔG_{hbond} is only negative in the two cases in which hydrogen bond formation is intramolecular, where the donor and acceptor being held so that they experience high effective concentrations of each other reduces the entropic cost of hydrogen bond formation.

The advantage of the direct method lies in its simplicity; the data that result from it can be interpreted directly in terms of hydrogen bonding without correcting for extraneous effects. The disadvantage lies in the association constants being so small that, in some cases, associations can only be observed at solute concentrations high enough to affect the properties of the solvent (particularly for the studies represented in entries 1-3 of table 3). In order to avoid this problem one can study systems where hydrogen bonds form intramolecularly and therefore more abundantly (entries 5 and 7 of table 3), or one can turn to the method described below.

1.2.4b The Incremental Energy Method for Studying Aqueous Hydrogen Bonding

Figure 4. A folding process, such as those studied using the incremental energy method, occurring with and without the possibility of the acceptor (A) or donor (D-H) being hydrogen bonded in the folded product.



All of the examples of the preceding sections studied hydrogen bonding between small, simple molecules in the gas phase or in solution. The incremental energy method studies large biomolecules that undergo binding or folding processes to which hydrogen bonding contributes and for which accurate free energies can be measured. The free

energy of hydrogen bonding (ΔG_{hbond}) is equated with the change in the free energy of the process when a single hydrogen bond donor or acceptor is deleted from a substrate or a particular site of a foldable protein chain ($\Delta\Delta G$).[44,45,46] For the example of a folding protein illustrated in figure 4,

$$\Delta\Delta G = (\Delta G - \Delta G_{noDH}) = (\Delta G - \Delta G_{noA}) = \Delta G_{hbond}$$

where $\Delta\Delta G$ and ΔG_{hbond} are as noted above, ΔG is the free energy of the process when the donor is present and the hydrogen bond can form, ΔG_{noDH} is the free energy of the process when the donor is absent, and ΔG_{noA} is the free energy of the process when the acceptor is absent.

Table 4 lists reported enthalpies and free energies of hydrogen bonding obtained by this method. The data indicate that hydrogen bond formation between neutral donors and acceptors is favorable ($\Delta G_{hbond} < 0$), and particularly so when either or both of the donor and the acceptor are charged.

---

[44] Fersht, A. R. *Biochemistry* **1988**, *27*, 1577.
[45] Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
[46] Pace, C. N. *Methods Enzymol.* **1995**, *259*, 538.

Table 4. Hydrogen bond energies from the incremental energy method (all energies are in kcal/mol).

System	Process	Donor	Acceptor	ΔH_{hbond}	ΔG_{hbond}
lysozyme[47]	folding	Ser OH	Ser O	-	−3.0
Staphlococcal Nuclease[48]	folding	various uncharged		-	−1.4 to −2.7
		various, one member charged		-	−2 to −4.3
Staphlococcal Nuclease[49]	folding	Tyr OH	Asp COO^-	-	−1.8
Carbonic Anhydrase[50]	folding	various uncharged		-	−0.9
Ribonuclease T1[51]	folding	various uncharged		-	−1.3 ± 0.6
T4 Lysozyme[52]	folding	amide NH	amide CO	-	−0.9
Barnase (buried H-bonds)[53]	folding	various uncharged		-	−0.5 to −2.0
		various, one member charged		-	<−3
Barnase (surface H-bonds)[53]	folding	various uncharged		-	~0
		various, one member charged		-	−0.3 to −1.0
i to i+4 side chain interactions in helical peptides[54,55]	folding	Lys NH_3^+	Glu COO^-	-	−0.3
		Gln NH	Glu COOH	-	−0.3
		Gln NH	Glu COO^-	-	−0.3
		Gln NH	Asp COOH	-	−0.4
		Gln NH	Asp COO^-	-	−1.0
Glycogen Phosphorylase[56]	binding	amide NH	alcohol O	-	−1.5
Thermolysin[57]	binding	phosphon-amide NH	amide CO	-	−4.0

[47] Yamada, H.; Kanaya, E.; Ueno, Y.; Ikehara, M.; Nakamura, H.; Kikuchi, M. *Biol. Pharm. Bull.* **1994**, *17*, 612.
[48] Byrne, M. P.; Manuel, R. L.; Lowe, L. G.; Stites, W. E. *Biochemistry* **1995**, *34*, 13949.
[49] Thorson, J. S.; Chapman, E.; Schultz, P. G. *J. Am. Chem. Soc.* **1995**, *117*, 9361.
[50] Mårtensson, L. G.; Jonsson, B.-H.; Andersson, M.; Kihlgren, A.; Bergenhem. N.; Carlsson, U. *Biochem. Biophys. Acta* **1992**, *1118*, 179.
[51] Shirley, B. A.; Stanssens, P.; Hahn, U.; Pace, C. N. *Biochemistry* **1992**, *31*, 725.
[52] Koh, J. T.; Cornish, V. W.; Schultz, P. G. *Biochemistry* **1997**, *36*, 11314.
[53] Serrano, L.; Kellis, Jr. J. T.; Cann, P.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 783.
[54] Scholtz, J. M.; Qian, H.; Robbins, V.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668.
[55] Huyghues-Despointes, B.; Klinger, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *34*, 13267.
[56] Street, I. P.; Armstrong, C. R.; Withers, S. G. *Biochemistry* **1986**, *25*, 6021.
[57] Bartlett, P. A.; Marlowe, C. K. *Science* **1987**, *235*, 569.

Tyrosyl-tRNA Synthetase[58]	binding	various uncharged		-	−0.5 to −1.5
		various charged		-	−3.5 to −4.5
synthetic adenine receptor[59]	binding	imide NH or adenine NH	adenine N or imide CO	−0.8	+0.2
cyclodextrin[60]	binding	phenol OH	alcohol O	−1.3 to −2.1	−0.3 to −0.6
vancomycin[61]	binding	amide NH	amide CO	−0.8	−3.3
ristocetin[61]	binding	amide NH	amide CO	+0.3	−1.6

The incremental energy method permits both easy characterization of energetics and measurements to be made at reasonably low solute concentrations. However, the values for ΔG_{hbond} listed in table 4 must yet be regarded with caution. In order to interpret $\Delta\Delta G$ simply as a hydrogen bonding energy, other interactions must not be affected by the deletion of D-H or A. The network of interactions that contribute to folding or binding phenomena is often exceedingly complicated, and since the measured energy change is small, the sum of errors contributed by other effects could be significant. There have been many attempts to correct $\Delta\Delta G$ for these errors.[44,45,46,62] Account has been taken of changes in hydrophobicity, van der Waals interactions, configurational entropy- the most baroque analysis of this type divides $\Delta\Delta G$ into 18 terms. Usually the largest correction is for the burial within the folded form of the protein or the bound form of the protein-substrate complex of the unhydrated partner of the deleted donor or acceptor. This destabilizing effect invariably makes the $\Delta\Delta G$ from removing a donor or acceptor more

[58] Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235.
[59] Kato, Y.; Conn, M. M.; Rebek, Jr. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1208.
[60] Ross, P. D.; Rekharsky, M. V. *Biophys. J.* **1996**, *71*, 2144.
[61] Williams, D. H.; Cox, J. P.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. *J. Am. Chem. Soc.* **1991**, *113*, 7020.

negative than ΔG_{hbond}.[63] However, if the widowed partner were somehow able to remain hydrogen bonded to solvent, then the destabilization would be ameliorated, and $\Delta\Delta G$ would be more representative of ΔG_{hbond}.[45,62,64,65] The hydrogen bonding status of the widowed partner can sometimes be ascertained when the crystal structure of the altered system is available. Unfortunately, however, the issue is more often left unresolved.

1.2.4c Status of the Problem of Hydrogen Bonding in Water

To understand the contribution of hydrogen bonding to processes that occur in water in general, and to biologically relevant processes such as protein folding in particular, one must know the intrinsic free energy of hydrogen bonding. This has been defined as the free energy change on going from the idealized state in which a donor and acceptor are poised to hydrogen bond but instead are solvated, to the state in which they are hydrogen bonded.[20] Attempts to determine this quantity have not led to conformity of opinion regarding aqueous hydrogen bonding thermodynamics.[20,44,62,65] Over the last decade, reviews have appeared that have claimed that hydrogen bonding is a major driving force for protein folding.[62,66] Others have asserted with equal confidence that it is not.[65,67] It is unlikely that the matter will be settled using the existing pool of data. For progress to be made, data will be required from new systems that combine the incremental energy method's ease of measurement with the direct method's ease of interpretation. We

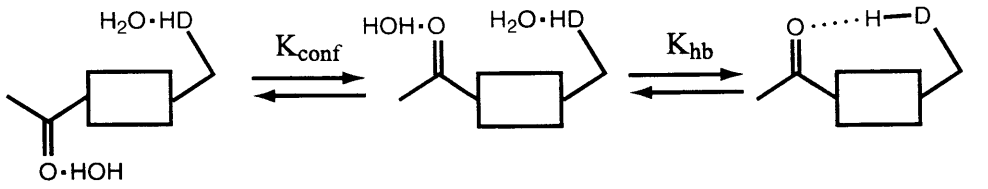
---

[62] Myers, J. K.; Pace, C. N. *Biophys. J.* **1996**, *71*, 2033.
[63] Hendsch, Z. S.; Jonsson, T.; Sauer, R. T.; Tidor, B. *Biochemistry* **1996**, *35*, 7621.
[64] Alber, T.; Dao-pin, S.; Wilson, K.; Wozniak, J. A.; Cook, S. P.; Matthews, B. W. *Nature* **1987**, *330*, 41.
[65] Honig, B.; Yang, A.-S. *Adv. Protein Chem.* **1995**, *46*, 27.
[66] Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
[67] Dill, K. A. *Biochemistry* **1990**, *29*, 7133.

propose that, if properly designed, a small molecule in which intramolecular hydrogen bond formation has an observable effect on the molecule's conformation could provide the necessary system.

1.2.5 A New System for the Study of Aqueous Hydrogen Bonding

Figure 5. A three state system for studying hydrogen bonding.



Consider the system pictured in figure 5, described schematically by the three conformations A, in which a hydrogen bond cannot form between D-H and C=O, B_{nhb}, in which D-H and C=O could hydrogen bond but are instead solvated, and B_{hb} where the D-H···O=C hydrogen bond has formed. The two equilibrium constants in the figure are $K_{conf} = [B_{nhb}]/[A]$, which is for the conformational transition from the A state to the B state, and $K_{hb} = [B_{nhb}]/[B_{hb}]$. We assert that K_{hb} is the intrinsic hydrogen bonding equilibrium constant (as per the definition given above and in ref 20).

Imagine that one could measure the overall A ⇄ B equilibrium constant, $([B_{nhb}]+[B_{hb}])/[A]$, by a reliable physical method; let this equilibrium constant be denoted

the B/A ratio. In terms of the equilibrium constants in figure 5, K_{conf} and K_{hb}, the B/A ratio is given by

$$B/A = K_{conf} + K_{conf}K_{hb}$$

The B/A ratio is thus directly proportional to K_{hb}.

The relationship between the B/A ratio and K_{hb} can be used to quantify hydrogen bonding energetics. Say that one could replace D-H by a stronger hydrogen bond donor, D'-H, without altering K_{conf}. Then the only difference between the system with D-H and the system with D'-H would be the larger K_{hb} of the new hydrogen bond donor. This difference would manifest itself in an increase in the measured B/A ratio. The shift in the overall B/A equilibrium therefore directly reports the relative strengths of the D-H···O=C and D'-H···O=C hydrogen bonds. Can this be taken a step further, to use the B/A ratio to determine the intrinsic hydrogen bonding equilibrium constant? For the system in figure 5, only two variables determine the B/A ratio: K_{conf} and K_{hb}. If K_{conf} can be measured independently of the B/A ratio, then K_{hb} is given immediately by

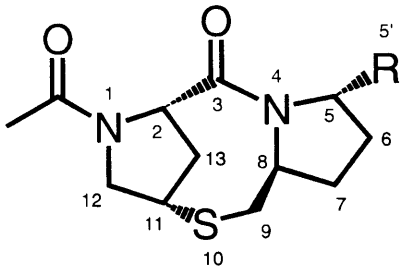
$$K_{hb} = \frac{B/A - K_{conf}}{K_{conf}}$$

Thus, the system of figure 5 can measure not only relative hydrogen bonding energetics, but also intrinsic hydrogen bonding energetics whose determination by other techniques has been so controversial.

Any real, experimental, system that can be used in this way for measurements of intrinsic hydrogen bond free energies must meet two conditions. A sensitive and

unambiguous monitoring of the A ⇄ B equilibrium must be available, and the B state must accommodate an acceptable hydrogen bonding geometry while having a minimum number of alternative, non-hydrogen bonding conformations. Candidates that meet these criteria are not easy to find. Beeson, et. al.[40] have used a rigid oxadecalin system to measure the hydrogen bonding in D_2O for a hydroxyl OH to ether O hydrogen bond (−0.5 kcal/mol; see table 3, entry 7), and Paliwal et. al.[68] have developed a cognate system to measure the strength of aryl-aryl interactions in $CDCl_3$. In 1988, Kemp and co-workers introduced the conformationally restricted Ac-Pro-Pro analog $AcHel_1$, pictured in figure 6, as an N-terminal template for nucleating helices in linked peptides.[69] In 1995, they reported a detailed conformational study of this system that analyzed the hydrogen bonding between the substituent at the 5 position (specifically, R = CONHMe) and the acetamide carbonyl.[70] This laid the foundation necessary for using $AcHel_1$ as a tool for measuring intrinsic hydrogen bond free energies.

Figure 6. The structure of $AcHel_1$ (R is variable).



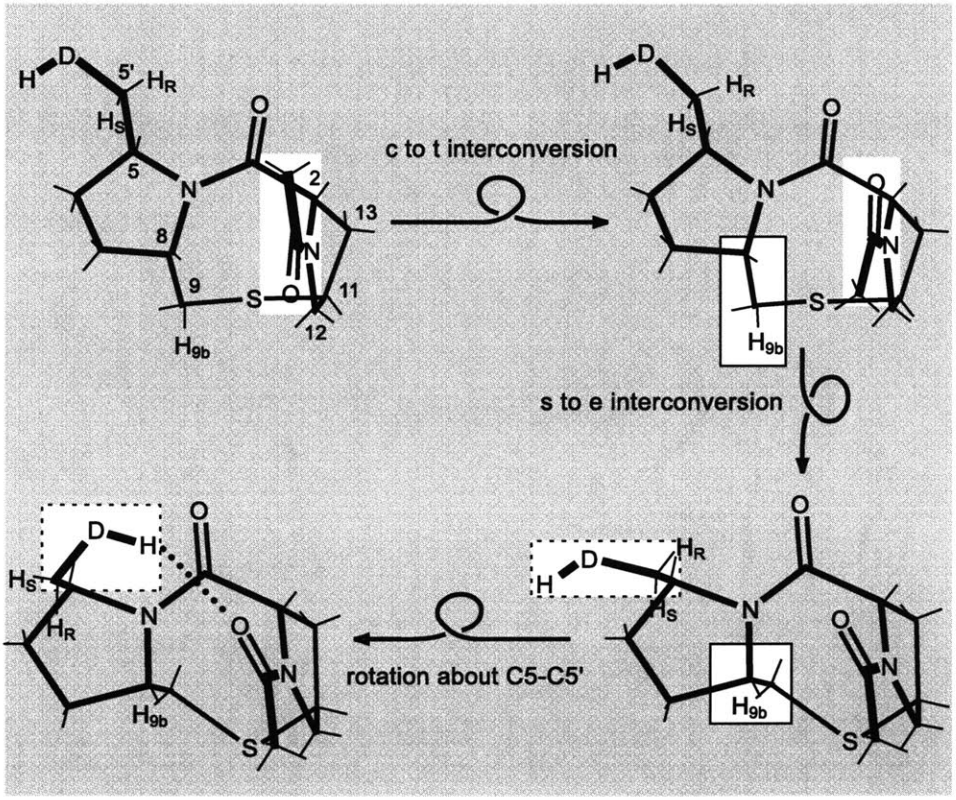
The work reported in this chapter is built upon this example, but varies the structure of the donor at the 5 position. Two points about this system should be noted prior to

[68] Paliwal, S.; Geib, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1994**, *116*, 4497.
[69] Kemp, D. S.; Curran, T. P. *Tetrahedron Lett.* **1988**, *29*, 4931.
[70] Kemp, D. S.; Allen, T. J.; Oslick, S. L. *J. Am. Chem. Soc.* **1995**, *117*, 6642.

developing its analysis. First, since a hydrogen bond is formed intramolecularly, the extent of hydrogen bond formation will be concentration independent, and measurements can be made in dilute solution. Second, the system is rigid enough that the hydrogen bonding geometry is precisely defined by the geometry of the system.

1.3 The Use of $AcHel_1$ to Study Aqueous Hydrogen Bonding

Figure 7. Conformational equilibria of an $AcHel_1$ derivative in which the 5' substituent is a generalized hydrogen bond donor (CH_2DH). The regions where the actual transitions occur are highlighted.



In previous studies, the low energy conformations of the $AcHel_1$ framework were shown to be extremely limited.[70] There are only three degrees of freedom: a cis (c) to trans (t) equilibrium about the N1 acetamide, a staggered (s) to eclipsed (e) equilibrium about the C8-C9 bond, and rotation about the C5-C5' and C5'-DH bonds that takes the hydrogen bond donor from being unoriented (u) to being oriented (o) for hydrogen bonding with the acetamide. This is illustrated in figure 7, where the R group is shown as a generalized donor, R = CH_2DH. Since the c and t state of the acetamide interconvert slowly on the NMR time scale, they are represented by separate NMR resonances. Their relative abundances can be measured by NMR peak integration and expressed as a ratio, the t/c ratio. Although the C8-C9 bond only adopts the s conformation when the acetamide is in the c state (overall state designation: cs), the t state has been shown to consist of a rapid equilibrium between the s and e conformations (the ts and te states, respectively), with the ts state strongly favored over the te state. However, only the te state appears to form strong intramolecular hydrogen bonds when the C5-C5' bond is in its oriented state.

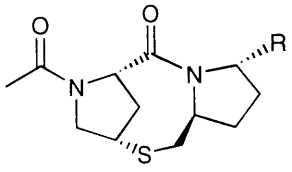
That hydrogen bonds form between the acetamide carbonyl of $AcHel_1$ and donors at the 5 position has been demonstrated by several criteria. For the case where R is an N-methylamide (R = CONHMe) the existence of this hydrogen bond was apparent in the NH stretching mode in an IR spectrum (in $CDCl_3$), in the temperature dependence of the chemical shift in an NMR spectrum (in 9:1 H_2O:D_2O), and nuclear Overhauser enhancements (NOEs; measured in $CDCl_3$) place the NH group in a position close enough to the carbonyl oxygen for a hydrogen bond.[71] Also, for the case where R =

[71] Cammers-Goodwin, A.; Allen, T. J.; Oslick, S. L.; McClure, K. F.; Lee, J. H.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 3082.

CH_2OH, one of the observed peaks corresponding to the OH stretch in CH_3CN is at a frequency characteristic of hydrogen bonded hydroxyl groups (3521 cm^{-1}).[72]

This chapter reports and analyzes the effect on the t/c ratio of placing the hydrogen bond donors shown in figure 8 at $AcHel_1$'s 5 position:

Figure 8. Hydrogen bond donors to be placed at the 5 position of $AcHel_1$.



R =

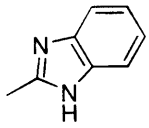

$AcHel_1CH_2OH$

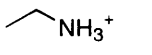

$AcHel_1CH_2SH$


$AcHel_1CH_2NH_2$

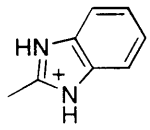

$AcHel_1CH_2NHAc$


$AcHel_1CONH_2$


$AcHel_1benzim$


$AcHel_1CH_2NH_3^+$


$AcHel_1CH_2NMe_2H^+$


$AcHel_1benzim^+$

(It should be noted that several of these compounds ($AcHel_1CH_2OAc$,[73] $AcHel_1CH_2OH$,[73] $AcHel_1CH_2NH_2$,[73] $AcHel_1CH_2NHAc$,[73] $AcHel_1CONH_2$,[71] and $AcHel_1CH_2SH$[74]) were prepared prior to the work in this thesis). The hydrogen bond donors in figure 8 represent most of the common hydrogen bond donors found in proteins. The alcohol and thiol are analogies to the serine and cysteine side chains. The amides stand for the peptide

[72] Bellamy, L. J. *The Infra-Red Spectra of Complex Molecules*; Methuen & Co. Ltd.: London, **1958**.
[73] McClure, K. F. Unpublished results.

backbone and the asparagine and glutamine side chains. The amines portray the α-amino group and the lysine side chain. Finally, the benzimidazole is a more synthetically accessible analog of the histidine side chain. The syntheses of the $AcHel_1$ derivatives in figure 8 are detailed in the experimental section.

1.3.1 Plan for Analyzing t/c Ratios

In the $AcHel_1$ system, a hydrogen bond between a donor at the 5 position and the acetamide carbonyl can only form when the acetamide is in its t rather than its c state. The stability of the trans state is thus tied directly to hydrogen bond formation, and the t ⇄ c equilibrium constant (t/c ratio), which can be measured by integrating the separate peaks the t and c states produce in an NMR spectrum, reports the donor's ability to form a hydrogen bond. The c and t states could correspond to the A and B states of figure 5; however, the t/c ratio by itself is not enough to determine the intrinsic hydrogen bonding energy. An estimate for the equivalent of K_{conf}, the equilibrium constant between the non-hydrogen bonding reference state and the state that is solvated but poised to hydrogen bond, is also required. K_{conf} for the $AcHel_1$ system is more complicated than it was for the ideal system of figure 5, since it encompasses all of the conformational transitions that are shown in figure 7. Furthermore, any dipole-dipole interactions or charge-dipole interactions that might affect K_{conf} independently of hydrogen bonding have to be accounted for.

---

[74] Lee, J. H. Thesis, Massachusetts Institute of Technology, **1997**.

1.3.2 The Mass Action Expression for the $AcHel_1$ System

The three conformational transitions that occur in $AcHel_1$ are the t ⇄ c transition of the acetamide, the s ⇄ e transition of the C8-C9 bond, and the u ⇄ o transition about the C5-C5' and C5'-DH bonds. These transitions are illustrated in figure 7, but not all of the states of the system are represented in this figure. A complete list of states can be made by following the conformational transitions from the starting cs state to the final hydrogen bonded state. In what follows, each state will be denoted by up to four descriptors, c or t to indicate the state of the acetamide, s or e to indicate the state of the C8-C9 torsion, u or o to indicate the state of the C5-C5' and C5-DH torsions, and hb to indicate whether an intramolecular hydrogen bond has formed. For example, the tsu state is the state in which the acetamide is trans, the C8-C9 torsion is staggered, and the C5-C5' and C5'-DH bonds are such that the donor is unoriented for hydrogen bonding.

The cs state is in equilibrium with the tsu state via the t ⇄ c transition. The tsu state is, in turn, in equilibrium with the teu state via the s ⇄ e transition. At this point, equilibria run in parallel from the tsu and teu states. The tsu state is in equilibrium with the tso state, and the teu state is in equilibrium with the teo state, both via the u ⇄ o transition about the C5-C5' and C5'-DH bonds. The hydrogen bond donor and the acetamide carbonyl are positioned properly for hydrogen bonding but are still solvated in the tso and teo states. Hydrogen bonding takes place when the solvating water molecules are shed, and the tso and teo states respectively become the tso,hb and teo,hb states, but it should be emphasized that the hydrogen bond in the teo,hb state is expected to be shorter

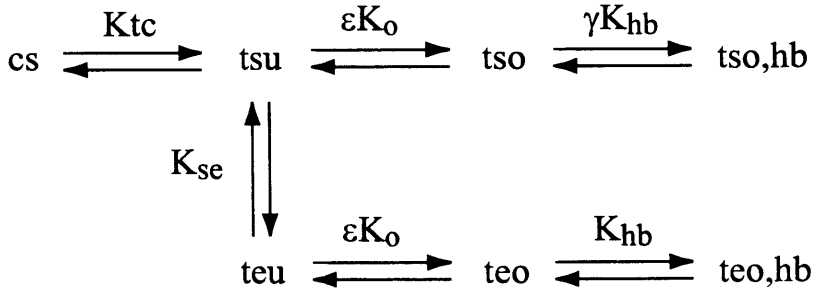
and stronger than that in the tso,hb state.[70] The system's states and their features are summarized in the table 5.

Table 5. States in the hydrogen bonding equilibrium of $AcHel_1R$.

States	Descriptions
cs	hydrogen bond cannot form because the acetamide is in its cis state
tsu and teu	hydrogen bond cannot form because the donor is not oriented
tso and teo	hydrogen bond could form, but donor and acceptor are still solvated
tso,hb and teo,hb	hydrogen bonded (teo,hb hydrogen bond stronger than that of tso,hb)

The equilibria and equilibrium constants among the states in table 5 are mapped out below in figure 9.

Figure 9. The complete set of equilibria that occur in $AcHel_1$ derivatives. The states are denoted by up to 4 descriptors: t or c to indicate the state of the acetamide, s or e to indicate the state of the C8-C9 torsion, u or o to indicate the state of the C5-C5' and C5'-DH torsions, and hb to indicate when there is a hydrogen bond.



All of the conformational transitions in $AcHel_1$ that precede actual hydrogen bond formation, those that lead up to the tso and teo states, are toward the left in figure 9. Their equilibrium constants are as follows. K_{tc} is the equilibrium constant between the cs state and the tsu state, K_{tc} = [tsu]/[cs]. K_{se} is the equilibrium constant between the tsu and teu states, K_{se} = [teu]/[tsu]. εK_o is the equilibrium constant between both the tsu and tso states and the teu and teo states, εK_o = [tso]/[tsu] = [teo]/teu]. This equilibrium constant is assumed to be independent of the s $\rightleftarrows$ e equilibrium, since the conformational energetics about the C5-C5' and C5'-DH bonds should be the same in both the s and the e states. The factor ε multiplies K_o to account for the increase in electrostatic attraction as charged 5 substituents (such as ammonium ions) approach the acetamide carbonyl upon entering the oriented state. For neutral compounds, $\varepsilon = 1$.

The final transitions on the right side of figure 9, tso $\rightleftarrows$ tso,hb and teo $\rightleftarrows$teo,hb, are the two that are most of interest, since they involve hydrogen bond formation. In fact, the equilibrium constants between these two pairs of states are the intrinsic hydrogen bonding equilibrium constants whose determination is the object of this chapter. K_{hb} is the intrinsic hydrogen bonding equilibrium constant between the teo and teo,hb states, K_{hb} = [teo,hb]/[teo]. γK_{hb} is the analogous intrinsic hydrogen bonding equilibrium constant between the tso and tso,hb, γK_{hb} = [tso,hb]/[tso]. The factor γ multiplies K_{hb} when hydrogen bonding occurs in the s state to account for the differing e and s state hydrogen bond strengths. γ is expected to be less than one, since the tso,hb hydrogen bond should be longer, and therefore weaker, than the teo,hb hydrogen bond.

The equilibrium constants from figure 9 are summarized in table 6.

Table 6. Summary of equilibrium constants from figure 7.

Equilibrium Constant	Description
K_{tc} = [tsu]/[cs]	Cost of entering t state from c state, with C8-C9 staggered and C5-C5' and C5'-DH unoriented
K_{se} = [teu]/[tsu]	Cost of entering e state from s state while acetamide is t and with C5-C5' and C5'-DH remaining unoriented
εK_o = [tso]/[tsu] = [teo]/[teu]	Cost of entering the oreinted state from the unoriented state while acetamide is t and C8-C9 torsion either s or e
ε	Factor to account for electrostatic interactions that might occur between charged donor and acetamide carbonyl; ε=1 for neutral compounds
K_{hb} = [teo,hb]/[teo]	Intrinsic hydrogen bonding equilibrium constant when C8-C9 torsion is e
γK_{hb} = [tso,hb]/[tso]	Intrinsic hydrogen bonding equilibrium constant when C8-C9 torsion is s
γ	Factor to account for the weaker hydrogen bond when the C8-C9 torsion is s

It is important to note here that, although hydrogen bond formation must certainly impact the relative populations of the t and c states, the s and e states, and the u and o states, it does not affect K_{tc}, K_{se}, or εK_o. These equilibrium constants represent the energetic cost of aligning the system into its hydrogen bonding conformation. This cost must be paid by hydrogen bonding; it is not lowered by it.

1.3.3 Derivation of the Fundamental Relationship that Equates the Intrinsic Hydrogen Bonding Equiliubrium Constant to Observables

The t/c ratio for an $AcHel_1$ derivative with a hydrogen bond donor is the ratio of the total concentrations of t state species to the total concentration of c state species. It can be expressed as a sum of the equilibrium constants in table 6 as follows:

$$\left(\frac{t}{c}\right)=\frac{[tsu]+[teu]+[tso]+[teo]+[tso,hb]+[teo,hb]}{[cs]} \qquad \textbf{1 - 1}$$

$$= K_{tc} + K_{tc}K_{se} + K_{tc}\varepsilon K_o + K_{tc}K_{se}\varepsilon K_o + K_{tc}\varepsilon K_o \gamma K_{hb} + K_{tc}K_{se}\varepsilon K_o K_{hb}$$

$$= [K_{tc}(1+K_{se})(1+\varepsilon K_o)]+[K_{tc}(\gamma+K_{se})\varepsilon K_o]\times K_{hb}$$

This expression for the t/c ratio is split into two parts, one that does not contain K_{hb} and one that does. The part that does not contain K_{hb} corresponds to the contributions to the t/c ratio from the states up to and including the tso and teo states. Imagine that one had an analog of the $AcHel_1$ donor compound that was identical in every respect, except that it could not hydrogen bond. The t/c ratio for this compound, the reference $AcHel_1$ derivative, would have only contributions up to and including the tso and teo states- it would be equal to the expression in the first set of brackets in equation 1-1:

$$\left(\frac{t}{c}\right)_{ref}=\frac{[tsu]+[teu]+[tso]+[teo]}{[c]}=K_{tc}(1+K_{se})(1+\varepsilon K_o) \qquad \textbf{1 - 2}$$

If the t/c ratio for such a reference compound were known, it could be used to simplify equation 1-1. Substituting $(t/c)_{ref}$ for $K_{tc}(1+K_{se})(1+\varepsilon K_o)$ in equation 1-1 gives

$$\left(\frac{t}{c}\right)=\left(\frac{t}{c}\right)_{ref}+[K_{tc}(\gamma+K_{se})\varepsilon K_o]\times K_{hb}$$

Also, since

$$\left(\frac{t}{c}\right)_{ref} \times \frac{1}{(1+K_{se})} \times \frac{1}{(1+\varepsilon K_o)} = K_{tc}$$

the remaining appearance by K_{tc} in the equation 1-1 can be eliminated by substituting the above for K_{tc}. This yields

$$\left(\frac{t}{c}\right) = \left(\frac{t}{c}\right)_{ref} + \left(\frac{t}{c}\right)_{ref}\left[\frac{\gamma+K_{se}}{1+K_{se}}\right]\left[\frac{\varepsilon K_o}{1+\varepsilon K_o}\right]K_{hb} \qquad \textbf{1 - 3}$$

Finally, isolating K_{hb} gives the master equation that will be used in the rest of the chapter:

$$\frac{\left[\frac{(t/c)-(t/c)_{ref}}{(t/c)_{ref}}\right]}{\left[\frac{\gamma+K_{se}}{1+K_{se}}\right]\left[\frac{\varepsilon K_o}{1+\varepsilon K_o}\right]} = K_{hb} \qquad \textbf{1 - 4}$$

The roles of the factors in brackets in equation 1-4, $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$, $(\gamma + K_{se})/(1 + K_{se})$, and $\varepsilon K_o/(1 + \varepsilon K_o)$, in the computation of K_{hb} is discussed in the following three paragraphs.

The numerator of the left-hand side of equation 1-4, $[(t/c) - (t/c)_{ref}]/(t/c)_{ref}$, consists of the measured t/c ratios of two compounds: the $AcHel_1$ donor compound (whose K_{hb} is of interest), and the analogous $AcHel_1$ reference derivative. With a properly chosen reference compound, $[(t/c) - (t/c)_{ref}]/(t/c)_{ref}$ represents the total increase in the t/c ratio that can be specifically attributed to hydrogen bonding (any extraneous charge-dipole and dipole-dipole effects have been corrected for).

The first factor in the denominator of the left-hand side of equation 1-4, $[(\gamma + K_{se})/(1 + K_{se})]$, consists only of quantities that pertain to the s ⇄ e equilibrium about C8-C9: K_{se}, which measures the system's intrinsic bias toward the s or e states, and γ, which measures

the difference in strength between hydrogen bonds formed in the s and e states. [(γ + K_{se})/(1 + K_{se})] reflects the extent to which the system's s ⇄ e equilibrium diminishes the effect of hydrogen bonding on the t/c ratio. If hydrogen bonds were equally strong in the tso,hb and teo,hb states, then γ would equal one, this factor would itself reduce to one, and the s ⇄ e equilibrium would be irrelevant to the calculation of K_{hb}. If the system were locked in the e state and $K_{se} \rightarrow$ infinity, then the factor would approach one and the s ⇄ e equilibrium would again be irrelevant to the calculation of K_{hb}. However, if K_{se} and γ were both small, then this factor would also be small, and any shift in t/c equilibrium would indicate a proportionately large value of K_{hb}. This last case is the most likely, given that the system is biased against the e state, so K_{se} must be small, and that the hydrogen bond in the teo,hb state is much stronger than that in the tso,hb state, so γ must also be small.[70]

The second factor in the denominator of the left-hand side of equation 1-4, [$\varepsilon K_o/(1+\varepsilon K_o)$], consists only of quantities that pertain to the u ⇄ o equilibrium about C5-C5' and C5'-DH: K_o, which measures the systems intrinsic bias toward one state or the other, and ε, which measures the electrostatic attraction between charged donors and the acetamide carbonyl in the oriented state. This factor reflects the extent to which the system's preferences about the C5-C5' and C5'-DH bonds diminish the effect of hydrogen bonding on the t/c ratio, and it is equal to the fractional population of the oriented state in the absence of hydrogen bonding, f_o:

$$f_o = \frac{\varepsilon K_o}{1+\varepsilon K_o}$$

If the system were locked in the oriented state and $\varepsilon K_o \rightarrow$ infinity, then $f_o \rightarrow 1$ and this factor would be irrelevant to the calculation of K_{hb}. However, if the oriented state were disfavored, then any shift in the t/c ratio would indicate a proportionately large K_{hb}.

The derivation of equation 1-4 reduces the problem of determining intrinsic hydrogen bonding equilibrium constants in the $AcHel_1$ system to a few manageable tasks. For a given $AcHel_1$ derivative with a donor at the 5 position, the following must be accomplished. Its t/c ratio must be measured, an appropriate reference compound must be identified, and the reference compound's t/c ratio measured; these t/c ratios determine $[(t/c)_{donor}-(t/c)_{ref}]/(t/c)_{ref}$. The values of K_{se} and γ must be measured; these equilibrium constants determine $[(\gamma+K_{se})/(1+K_{se})]$. Finally, the value of f_o must be estimated. These steps are followed in the next sections for the set of $AcHel_1$ derivatives shown in figure 8.

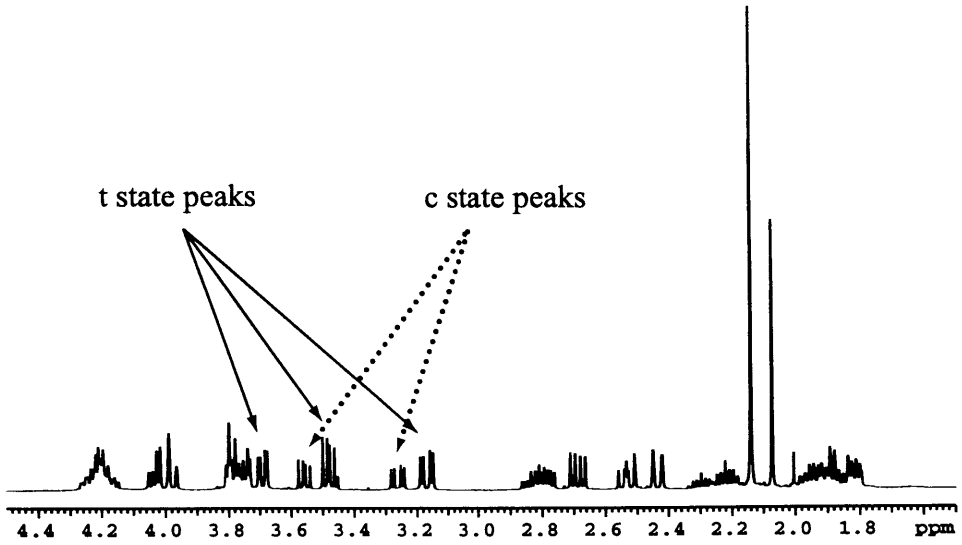
1.4 Determination of K_{hb} for $AcHel_1$ Derivatives with Donors at the 5 position

1.4.1 Measurements of $(t/c)_{donor}$ and $(t/c)_{ref}$

The t/c ratio for any $AcHel_1$ derivative can be measured, as mentioned above, by integrating the separate peaks that these states produce in an ^{1}H-NMR spectrum. An example of one such spectrum, that of $AcHel_1CH_2OH$, is shown in figure 10 with some of the t and c state peaks identified (the method by which t/c ratios are calculated is detailed in the experimental section).[75]

[75] This spectrum was assigned by DQ-COSY and NOESY spectroscopy.[73] For the most part, the resonances of the protons in the $AcHel_1$ core varied little from derivative to derivative. This allowed the assignment of the ^{1}H-NMR spectra of other compounds in this series were by analogy with this spectrum.

Figure 10. ^{1}H-NMR spectrum of $AcHel_1CH_2OH$, showing the different t and c state peaks. Specific examples of t and c state peaks are indicated.



The t/c ratios for the $AcHel_1$ derivatives in figure 8 are listed in increasing order in table 7. All of these t/c ratios were measured at 25°C in D_2O either in unbuffered solution or at the pD indicated (pD was adjusted as necessary by adding trifluoroacetic acid-d_1 (TFA-d_1) or NaOD and checked using pH paper). All quantities reported for hydrogen bonds therefore refer to hydrogen bonds formed with the deuterium isotope. However, solvent isotope effects on hydrogen bonding are small[76,77,78] so the data obtained in this work should be applicable to hydrogen bonds formed with protium as well as deuterium.

[76] Hermans Jr., J.; Scheraga, H. A. *Biochem. Biophys. Acta* **1959**, *36*, 534.
[77] Scheraga, H. A. *Ann. NY Acad. Sci.* **1961**, *608*.
[78] Calvin, M.; Hermans Jr., J; Scheraga, H. A. *J. Am. Chem. Soc.* **1960**, *81*, 5048.

Table 7. The t/c ratios for the $AcHel_1$ derivatives with hydrogen bond donors at the 5 position at 25 °C and the pD indicated. If no pD is given then the t/c ratio was measured in unbuffered D_2O. The error is not reported for the t/c ratio of $AcHel_1benzim^+$ because it could not be determined the way the others were (see experimental section), but had to be computed from the ratio of integrals of just one t state and one c state peak in its NMR spectrum.

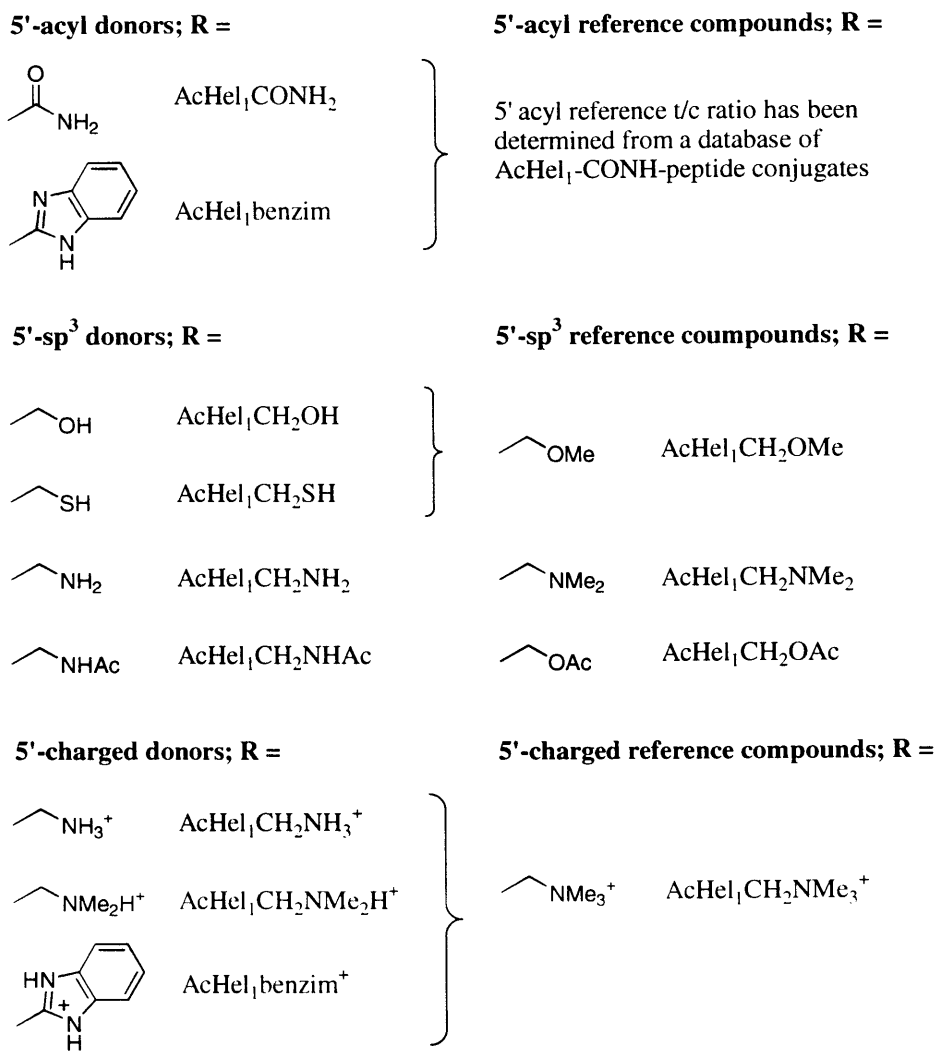
$AcHel_1$ derivative	t/c ratio
$AcHel_1CH_2OH$	1.81 (±0.03)
$AcHel_1CH_2SH$	1.82 (±0.07)
$AcHel_1CONH_2$	1.85 (±0.04)
AcHel1benzim (pD 13)	1.93 (±0.10)
$AcHel_1CH_2NHAc$	2.14 (±0.06)
$AcHel_1CH_2NH_2$ (pD 13)	2.21 (±0.07)
$AcHel_1CH_2NMe_2H^+$ (pD 1)	4.96 (±0.28)
$AcHel_1CH_2NH_3^+$ (pD 1)	5.23 (±0.21)
$AcHel_1benzim^+$ (pD 1)	7.7

One can see from table 7 that the derivatives with positively charged donors (R = $CH_2NH_3^+$, $CH_2NMe_2H^+$ and $benzim^+$), whose t/c ratios range from 4.96 to 7.7, most strongly stabilize the t state relative to the c state. The t/c ratios of the derivatives with neutral donors (R = CH_2OH, CH_2SH, CH_2NHAc, $CONH_2$, and benzim) are all in the range from 1.8 to 2.1. Based solely on these t/c ratios, one would conclude that charged hydrogen bond donors are the best donors, and that all the neutral donors have similar abilities to donate hydrogen bonds, consistent with statements in the literature based on data from the incremental energy method.[48,53,58] However, it is not yet clear how much of the observed differences in t/c ratios can be attributed to charge-dipole or dipole-dipole interactions. These effects have to be assessed using the t/c ratios of reference compounds.

An ideal reference compound would be identical to a given $AcHel_1$ derivative from figure 8, except that the donor's hydrogen bonding potential would be turned off. This is not physically realizable. Nevertheless, the ideal can be approximated using obvious non-hydrogen bond forming analogs for some of the $AcHel_1$ derivatives in figure 8. Hence, $AcHel_1CH_2OMe$ is the reference compound for $AcHel_1CH_2OH$ and $AcHel_1CH_2SH$; $AcHel_1CH_2NMe_2$ is the reference compound for $AcHel_1CH_2NH_2$; and $AcHel_1CH_2OAc$ for $AcHel_1CH_2NHAc$. A natural choice for $AcHel_1CONH_2$'s reference compound would be the methyl ester, $AcHel_1CO_2Me$. However, a more accurate reference t/c ratio for $AcHel_1$ derivatives in which the substituent at the 5 position is an acyl group has been determined from a large number of $AcHel_1CONH$-peptide conjugates.[70] The latter value will be used, but it should be noted that the two possible values of $(t/c)_{ref}$ are similar (the t/c ratio of $AcHel_1CO_2Me$ is 0.70, while the reference t/c ratio determined from $AcHel_1$-peptide derivatives is 0.83).[70] Selecting a reference compound for the $AcHel_1$ derivatives with ammonium ions ($AcHel_1CH_2NH_3^+$ and $AcHel_1CH_2NMe_2H^+$) poses a problem. The obvious choice is the trimethylammonium derivative of $AcHel_1$, $AcHel_1CH_2NMe_3^+$, but the sterically demanding trimethylammonium ion will not be able to approach the acetamide as closely as an ammonium ion. The value of εK_o in the expression $f_o = \varepsilon K_o/(1+\varepsilon K_o)$ is not likely to be the same in $AcHel_1CH_2NMe_3^+$ and $AcHel_1CH_2NH_3^+$, so equation 1-4 may not strictly hold for this pair of compounds. Since the attraction between the acetamide carbonyl and the charged group at the 5 position is probably larger in $AcHel_1CH_2NH_3^+$ than in $AcHel_1CH_2NMe_3^+$, the t/c ratio for $AcHel_1CH_2NMe_3^+$ has to be regarded as a lower limit on the $(t/c)_{ref}$ for the $AcHel_1$derivatives with ammonium ions. Selecting reference

compounds for $AcHel_1benzim$ and $AcHel_1benzim^+$ also poses a challenge. The reference for the unprotonated $AcHel_1benzim$ is taken to be the same as the reference for $AcHel_1CONH_2$, since the dipole of the C=N in the benzimidazole ring is expected to have a similar effect on the t/c ratio as the C=O of the amide. This should not be the case in $AcHel_1benzim^+$ as resonance in the benzimidazolium ion should eliminate the C=N dipole. Since $AcHel_1benzim^+$ is charged, $AcHel_1CH_2NMe_3^+$ will be used as its reference compound. The reference compounds that will be used in this chapter, and how they will be used, is summarized in figure 11. Their synthesis is described in the experimental section.

Figure 11. $AcHel_1$ derivatives with hydrogen bond donors at the 5 position, and their matching $AcHel_1$ reference compounds.



The compounds are grouped in figure 11 according to whether the substituent at the 5 position is neutral and sp^2 hybridized (5'-acyl $AcHel_1$ derivatives), neutral and sp^3 hybridized (5'-sp^3 $AcHel_1$ derivatives), or charged (5'-charged $AcHel_1$ derivatives). The t/c ratios of these compounds are reported in table 8.

Table 8. t/c ratios for reference $AcHel_l$ derivatives. The * in the first row indicates that the reference t/c ratio for the 5'-acyl $AcHel_l$ derivatives was determined from a large database of $AcHel_l$ derivatives attached to peptides.

$AcHel_l$ derivative	t/c ratio	reference for
*	0.83	5'-acyl $AcHel_l$ derivatives
$AcHel_lCH_2OAc$	1.35 (±0.04)	5'-sp^3 $AcHel_l$ derivatives
$AcHel_lCH_2NMe_2$ (pD 13)	1.40 (±0.04)	
$AcHel_lCH_2OMe$	1.60 (±0.06)	
$AcHel_lCH_2NMe_3^+$	3.48 (±0.10)	5'-charged $AcHel_l$ derivatives

The reference t/c ratios for the 5'-sp^3 $AcHel_l$ derivatives are all around 1.5, indicating that the non-hydrogen bonding effects on the t/c ratio within this class of $AcHel_l$ derivatives have little impact on t/c ratios. Looking between classes of $AcHel_l$ derivatives, however, shows that charge-dipole and dipole-dipole interactions can significantly affect the t/c ratio apart from hydrogen bonding. According to the reference t/c ratio for the 5'-acyl $AcHel_l$ derivatives, the t state is about half as stable (t/c = 0.83) in this class of compounds as it is in the 5'-sp^3 $AcHel_l$ derivatives. Furthermore, the reference t/c ratio for the 5'-charged $AcHel_l$ derivatives indicates that the t state is about twice as stable (t/c = 3.48) in this class of compounds as it is in the 5'-sp^3 $AcHel_l$ derivatives. The variation seen in the reference t/c ratios promises that corrections for the dipole-dipole interactions in the 5'-acyl $AcHel_l$ derivatives and for charge-dipole interactions in the 5'-charged $AcHel_l$ derivatives will be significant.

The quantities $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ for all the $AcHel_l$ derivatives are listed in table 9.

Table 9. The t/c ratios and quantities $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ measured in D_2O at 25°C for the 5'-acyl, 5'-sp^3, and 5'-charged $AcHel_1$ derivatives.

Compound	Type	$[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$
$AcHel_1CONH_2$	5'-acyl	1.22 (±0.03)
AcHel1benzim (pD 12)	5'-acyl	1.33 (±0.07)
$AcHel_1CH_2OH$	5'-sp^3	0.13 (±0.04)
$AcHel1CH_2SH$	5'-sp^3	0.14 (±0.06)
$AcHel_1CH_2NHAc$	5'-sp^3	0.59 (± 0.06)
$AcHel_1CH_2NH_2$ (pD 13)	5'-sp^3	0.58 (±0.05)
$AcHel_1CH_2NH_3^+$ (pD 1)	5'-charged	0.50 (±0.07)
$AcHel_1CH_2NMe_2H^+$(pD 1)	5'-charged	0.43 (±0.09)
$AcHel_1benzim^+$ (pD 1)	5'-charged	1.2

The data in table 9 force a re-evaluation of the statement, made after the t/c ratios for the $AcHel_1$ donor derivatives were presented in table 7, that charged donors form stronger hydrogen bonds than neutral donors, and that the hydrogen bonds formed by neutral donors are all comparable. Although the positively charged groups in the 5'-charged derivatives of $AcHel_1$ stabilize the t state much more than any of the neutral groups in the 5'-acyl or 5'-sp^3 $AcHel_1$ derivatives, the values of $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ for the 5'-charged compounds are not larger than those of some of the neutral donors. A significant portion of this t state stabilization can apparently be attributed to charge-dipole interactions that are also present in $AcHel_1CH_2NMe_3^+$, rather than to strong hydrogen bonding interactions. Furthermore, the values of $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ for the neutral 5'-acyl and 5'-sp^3 $AcHel_1$ derivatives are more variable than their t/c ratios are. This results from the correction for the significant stabilization of the c state, possibly by dipole-dipole interactions, in the 5'-acyl derivatives. The t/c ratios for the 5'-acyl donor derivatives of

$AcHel_1$, although they are in the same range as those of the 5'-sp^3 donor derivatives of $AcHel_1$, represent more of a perturbation from their reference t/c ratios.

One step required for determining intrinsic hydrogen bonding equilibrium constants has been completed in this section: the correction of the t/c ratios of $AcHel_1$ donor compounds for non-hydrogen bonding effects. All that is required now is to determine $(\gamma+K_{se})/(1+K_{se})$ and f_o.

1.4.2 Determination of K_{se} and γ

In this section, the quantities relating to the staggered ⇄ eclipsed equilibrium about the C8-C9 bond are determined. K_{se} is dealt with first, followed by γ.

1.4.2a Equilibrium Constant K_{se}

K_{se} is the equilibrium constant between the tsu and teu states. Since the unoriented to oriented transition about the C5-C5' and C5'-DH bonds is independent of the s ⇄ e transition about the C8-C9 bond, the energetic cost of the s ⇄ e conversion should be the same whether it occurs in the unoriented state (tsu ⇄ teu) or the oriented state (tso ⇄ teo). K_{se} is therefore also the equilibrium constant between the tso and teo states, and it can be written that

$$K_{se} = \frac{[teu]+[teo]}{[tsu]+[tso]}$$

While the acetamide is in the t state, the total concentration of species in the e state that are not hydrogen bonded divided by the total concentration of species in the s state that are not hydrogen bonded is equal to K_{se}. The equality between K_{se} and [e state concentration]/[s state concentration] no longer holds if one includes hydrogen bonded states in the ratio, since hydrogen bonds favor the e state. Therefore, one cannot determine K_{se} directly from the ratio [total e state concentration]/[total s state concentration] in $AcHel_1$ derivatives that have hydrogen bond donors. However, one can determine this equilibrium constant from the $AcHel_1$ reference compounds, since these have no hydrogen bonded states.

The e and s state populations in the $AcHel_1$ reference compounds can be determined using the t state chemical shift of the 9b proton (this proton is labeled in figure 7, and its resonance is at 3.17 ppm in the ^{1}H-NMR spectrum of $AcHel_1CH_2OH$ in figure 10). The chemical shift of the 9b proton ($\delta 9b_{obs}$) is a weighted average of the limiting chemical shifts of the pure ts and te states ($\delta 9b_{ts}$ and $\delta 9b_{te}$) so that

$$\delta 9b_{obs} = f_s \times \delta 9b_{ts} + f_e \times \delta 9b_{te} \qquad \textbf{1 - 5}$$

where f_s is the mole fraction of the s state within the t state and f_e is the mole fraction of the e state within the t state. Substituting $f_e = 1 - f_s$ into the equation 1-5 above and rearrangement gives

$$f_s = \frac{\delta 9b_{obs} - \delta 9b_{te}}{\delta 9b_{ts} - \delta 9b_{te}} \qquad \textbf{1 - 6}$$

K_{se} in terms of f_s is simply

$$K_{se} = \frac{\text{concentration of e states}}{\text{concentration of s states}} = \frac{f_e}{f_s} = \frac{1 - f_s}{f_s}$$

1 - 7

To determine K_{se} for a reference compound, one need only determine f_s, and to determine f_s, one only needs $\delta 9b_{obs}$ for the $AcHel_1$ derivative in question and the values of $\delta 9b_{te}$ and $\delta 9b_{ts}$. The limiting values of $\delta 9b_{te}$ and $\delta 9b_{ts}$ have been assigned in previous work. The t state of $AcHel_1CO_2Me$ approaches pure s character, and $\delta 9b$ in this compound is 3.32 ppm.[70] The t states of $AcHel_1$ derivatives with very helical peptides attached to them approach pure e character, and $\delta 9b$ in these compounds is about 2.87 ppm.[70]

For the 5'-acyl $AcHel_1$ derivatives, K_{se} is small and difficult to measure, but it has been placed[70] between 0.05 and 0.20. K_{se} could not be much larger, since the t state of 5'-acyl $AcHel_1$ derivatives that lack hydrogen bond donors (such as $AcHel_1CO_2Me$) have so little e character. It also could not be much smaller, or else unreasonably large hydrogen bonding energies would be required to effect the s $\rightleftarrows$ e transition about the C8-C9 bond.

For the 5'-sp^3 $AcHel_1$ derivatives, K_{se} is larger. The chemical shifts of the 9b protons of the neutral 5'-sp^3 reference compounds are 3.17 ppm for $AcHel_1CH_2OAc$, 3.15 for $AcHel_1CH_2NMe_2$, and 3.15 for $AcHel_1CH_2OMe$. Using these chemical shifts in equations 1-6 and 1-7 yields K_{se} = 0.5, 0.61, and 0.61. The average K_{se} for 5'-sp^3 compounds is 0.57.

The value of K_{se} is assumed to be the same for the 5'-charged $AcHel_1$ derivatives as for the neutral 5'-sp^3 $AcHel_1$ derivatives since $\delta 9b_{obs}$ for $AcHel_1CH_2NMe_3^+$ is 3.14 ppm, close to the chemical shifts for this proton witnessed in the neutral 5'-sp^3 reference compounds.

1.4.2b Factor γ

The factor γ corrects the intrinsic hydrogen bonding equilibrium constant of the s state for its poor geometry. The extent to which hydrogen bonding shifts the s $\rightleftarrows$ e equilibrium determines the value of γ. If the hydrogen bonds in the two states were equally favorable, then hydrogen bonding would not affect the disposition of the C8-C9 bond, and γ would be one. However, since hydrogen bonding favors the e state, γ is less than one.

One would expect that the t $\rightleftarrows$ c and s $\rightleftarrows$ e conformational changes ought to be somehow covariant, since both equilibria are affected by hydrogen bonding. In fact, it has been shown that f_s, and 1/(t/c) are linearly related.[70] This relationship can be cast in terms of the equilibria of figure 9 by noting that the rate of change of f_s as a function of 1/(t/c) is

$$\frac{df_s / dK_{hb}}{d[1/(t/c)]/dK_{hb}} = \frac{df_s / d[\varepsilon K_o]}{d[1/(t/c)]/d[\varepsilon K_o]} = \frac{df_s}{d[1/(t/c)]} = \frac{d\left[\dfrac{\delta 9b_{obs} - \delta 9b_{te}}{\delta 9b_{ts} - \delta 9b_{te}}\right]}{d[1/(t/c)]} = \frac{(1-\gamma)K_{se}(t/c)_{ref}}{(1+K_{se})(\gamma+K_{se})}$$

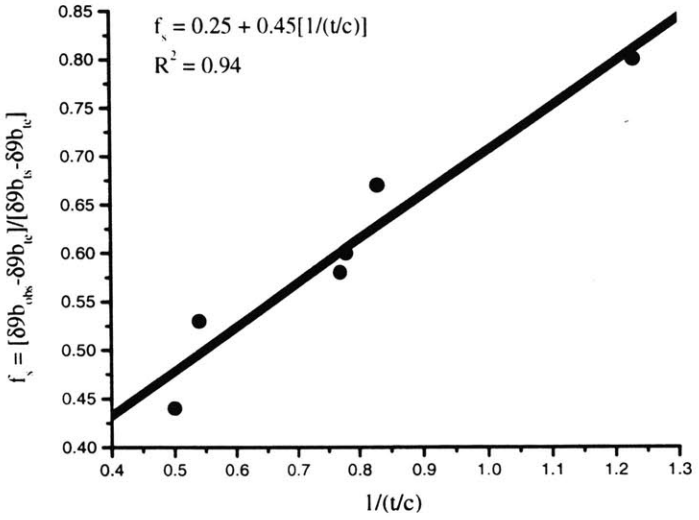
1 - 8

For any set of compounds that have the same γ, K_{se} and $(t/c)_{ref}$, therefore, a plot of $[(\delta 9_{obs} - \delta 9_{te})/(\delta 9_{ts} - \delta 9_{te})]$ vs. 1/(t/c) should be linear with a slope given by equation 1-8. This slope can be determined by linear regression and solved for γ if $(t/c)_{ref}$ and K_{se} are known.

Only two examples of 5'-acyl derivatives are available in this work, one of which, AcHel$_1$benzim, must be excluded from consideration since the benzimidazole's ring

current affects the chemical shifts of nearby protons.[79] To do the necessary linear regression, 1/(t/c) and $\delta 9b_{obs}$ data from derivatives of $AcHel_1$ that have appeared elsewhere[70] (such as $AcHel_1CONHMe$, $AcHel_1COGlyOH$, and $AcHel_1COAlaOH$) or that have not yet been published[73] (such as $AcHel_1CONHCH_2CF_3$ and $AcHel_1CONHOMe$) have to be used. The plot including these data is shown in figure 12 (the data are tabulated in appendix 1). There is a strong correlation between f_s and 1/(t/c) ($R^2 = 0.94$), and the slope of the line of best fit is 0.45.

Figure 12. Plot of f_s vs. 1/(t/c) for 5'-acyl $AcHel_1$ derivatives.



With

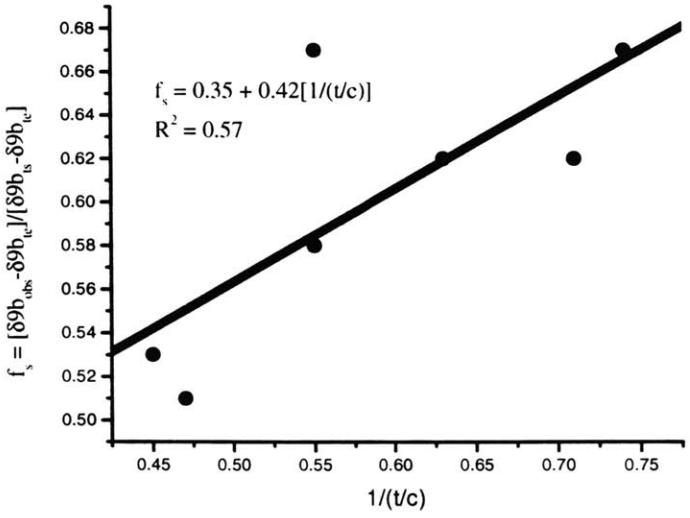
[79] For example, the chemical shift of H8 in $AcHel_1CH_2OH$ (assigned by COSY[73]) is 4.2 ppm. In $AcHel_1$benzim the chemical shift of the only peak that could correspond to the 8 proton (as judged by its coupling pattern) is 4.5 ppm.

$$0.45 = \frac{(1-\gamma)K_{se}(t/c)_{ref}}{(1+K_{se})(\gamma+K_{se})}$$

and $(t/c)_{ref} = 0.83$, if $K_{se} = 0.05$, then $\gamma = 0.04$; if $K_{se} = 0.20$, then $\gamma = 0.09$. So, for 5'-acyl derivatives of $AcHel_1$, $0.04 \leq \gamma \leq 0.09$. This small γ shows that for 5'-acyl donors, the larger donor–acceptor distance in the s state compared to the e state has a very detrimental effect on s state hydrogen bonding.

The plot of f_s vs. 1/(t/c) for the 5'-sp^3 $AcHel_1$ derivatives is shown in figure 13 (the data are, as before, tabulated in appendix 1). These data are not as well correlated as they were for the 5'-acyl cases ($R^2 = 0.57$), probably because of the single outlier point due to $AcHel_1CH_2OH$ (at 1/(t/c) = 0.55 and $f_s = 0.67$ in figure 13). This could indicate that γ is very different for the alcohol than for the other 5'-sp^3 $AcHel_1$ derivatives, but more data would be needed before any conclusions could be drawn. The slope of the line of best fit is 0.42.

Figure 13. Plot of f_s vs. 1/(t/c) for 5' sp^3 $AcHel_1$ derivatives.



With

$$0.42 = \frac{(1-\gamma)K_{se}(t/c)_{ref}}{(1+K_{se})(\gamma+K_{se})}$$

$(t/c)_{ref}$ = 1.45 (the average of the three 5'-sp^3 reference compounds' t/c ratios), and K_{se} = 0.57, then $\gamma = 0.31$ for the 5'-sp^3 derivatives of $AcHel_1$. The value of γ is much larger for 5'-sp^3 donors than for the 5'-acyl donors, perhaps because the longer bonds of the sp^3 hybridized atoms better allow the 5'-sp^3 donors to adapt to the different hydrogen bonding geometries available in the s and e states.

There are not enough data to make a plot such as those in figures 12 and 13 for the 5'-charged $AcHel_1$ derivatives. The value of γ for these compounds has to be assigned by analogy. According to the results for the 5'-acyl and 5'-sp^3 $AcHel_1$ derivatives, the value of γ seems to depend on the hybridization of C5'. Consequently, the γ found for the neutral 5'-sp^3 $AcHel_1$ derivatives is assigned to $AcHel_1CH_2NH_3^+$ and $AcHel_1CH_2NMe_2H^+$ while the limits on γ found for 5'-acyl derivatives are assigned to $AcHel_1benzim^+$.

1.4.2c $(\gamma + K_{se})/(1 + K_{se})$

Using the values for γ and K_{se} obtained above, the quantities $(\gamma + K_{se})/(1 + K_{se})$ for the 5'-acyl, 5'-sp^3, and 5'-charged $AcHel_1$ derivatives can be computed. These are listed in table 10.

Table 10. Values (or ranges) of K_{se}, γ, and $(\gamma+K_{se})/(1+K_{se})$ for $AcHel_1$ derivatives.

$AcHel_1$ derivative	type	K_{se}		γ		$(\gamma + K_{se})/(1 + K_{se})$	
		min	max	min	max	min	max
$AcHel_1CONH_2$	5'-acyl	0.05	0.20	0.04	0.09	0.09	0.24
$AcHel_1$benzim	5'-acyl	0.05	0.20	0.04	0.09	0.09	0.24
$AcHel_1CH_2OH$	5'-sp^3	0.57		0.31		0.56	
$AcHel_1CH_2SH$	5'-sp^3	0.57		0.31		0.56	
$AcHel_1CH_2NH_2$	5'-sp^3	0.57		0.31		0.56	
$AcHel_1CH_2NHAc$	5'-sp^3	0.57		0.31		0.56	
$AcHel_1CH_2NMe_2H^+$	5'-charged	0.57		0.31		0.56	
$AcHel_1CH_2NH_3^+$	5'-charged	0.57		0.31		0.56	
AcHel1benzim$^+$	5'-charged	0.57		0.04	0.09	0.39	0.42

The quantity $(\gamma + K_{se})/(1 + K_{se})$ is smaller for the 5'-acyl $AcHel_1$ derivatives than for either the 5'-sp^3 or 5'-charged. Since K_{hb} is inversely proportional to $(\gamma + K_{se})/(1 + K_{se})$, the same change in t/c ratios of 5'-acyl, 5'-sp^3, and 5'-charged $AcHel_1$ derivatives relative to their reference t/c ratios will indicate a larger K_{hb} for 5'-acyl than for 5'-sp^3 or 5'-charged $AcHel_1$ derivatives.

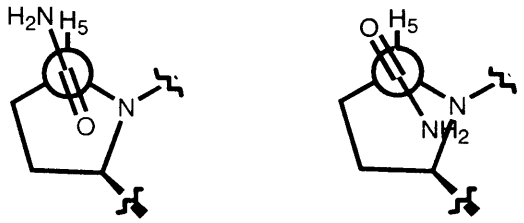
1.4.3 Determination of f_o

The last quantity that has to be determined before K_{hb} can be calculated from t/c ratios is f_o, the fractional population of the oriented state in the absence of hydrogen bonding.

This quantity will be used to account for the unproductive conformations about the C5-C5' and C5'-DH bonds in the calculation of K_{hb}. The values of f_o will be estimated from model systems, or measured in $AcHel_l$ derivatives in situations where hydrogen bonding does not perturb the equilibrium's natural inclinations (reference $AcHel_l$ derivatives and the c states of donor $AcHel_l$ derivatives).

The unoriented ⇄ oriented equilibrium is actually the product of two equilibria: that about the C5-C5' torsion and that about the C5'-DH torsion. For the 5'-acyl derivatives of $AcHel_l$, where the R is typically an amide, the C5'-DH torsion is locked into one conformation by resonance and need not be considered. The C5-C5' torsion has two conformations,[70] shown in figure 14: one in the unoriented state with the N4-C5-C5'-DH dihedral angle ~100° and one in the oriented state with the N4-C5-C5'-DH dihedral angle ~0°.[80]

Figure 14. Newman projections of the unoriented and oriented states for 5'-acyl $AcHel_l$ derivatives.



unoriented oriented

Because there is no proton-proton coupling constant that can indicate the state of the C5-C5' torsion, the equilibrium constant between these two states cannot be

[80] According to vacuum molecular mechanics calculations[70], the hydrogen bonds formed from the oriented state have an NH···O distance of 2.05Å in the e state and 2.22Å in the s state, an NH···O angle of 153° for the e state and 155° for the s state and an H···O=C angle of 125° in the e state and 118° in the s state.

experimentally determined. However, the C5-C5' torsion is analogous to, and ought to have the same conformational preferences as, a proline ψ torsion. According to this assumption, it ought to be possible to estimate f_o for 5'-acyl $AcHel_1$ derivatives from studies of proline's conformation. In this analogy, the oriented state about the C5-C5' bond corresponds to proline's "core α" region (where ψ is roughly between +30° and −45°, as defined by Swindells, et. al.[81]). The fraction of proline residues in coil regions of protein structures (those regions that are not part of regular secondary structures) with core α conformations is 0.32,[82] so f_o for the 5'-acyl $AcHel_1$ derivatives should likewise be about 0.32. $AcHel_1$benzim will be assigned the same value of f_o, since the forces affecting its conformation about the C5-C5' bond ought to be similar to those affecting $AcHel_1CONH_2$'s conformation about the C5-C5' bond.

For 5'-sp^3 derivatives of $AcHel_1$, the C5-C5' torsion has three states: gg (in which DH is gauche to both N4 and C6), ga (in which DH is gauche to N4 and anti to C6), and ag (in which DH is anti to N4 and gauche to C6). These states are illustrated in figure 15. Hydrogen bonds can form in either the gg or ga rotamers[83] but only when the C5'-DH bond is rotated such that the donor proton is directed toward the acetamide (the hydrogen bond in the ga rotamer of the e state is shown in the lower left of figure 7). Hydrogen bonds cannot form in the ag state under any circumstances. Considering both the ag ⇄ (gg, ga) equilibrium of the C5-C5' bond and the (away from acetamide) ⇄ (toward

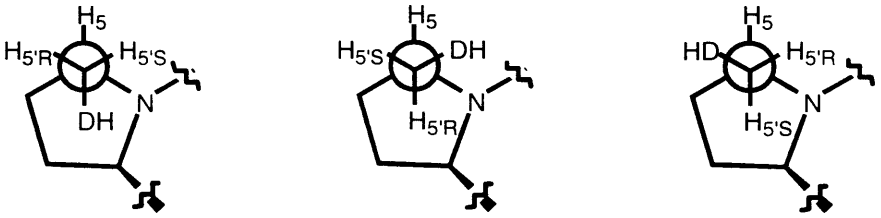
---

[81] Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nature Struct. Biol.* **1995**, *2*, 596.
[82] Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.
[83] According to vacuum molecular mechanics calculations (CHARMm force field), the hydrogen bonds formed from the gg state of the C5-C5' torsion when DH = OH have NH···O distances of 1.89 Å and 2.10 Å, NH···O angles of 169° and 162°, and H···O=C angles of 117° and 117° for the s and e states. The hydrogen bonds formed from the ga state of the C5-C5' torsion have NH···O distances of 1.90 Å and 1.99 Å, NH···O angles of 152° and 167°, and H···O=C angles of 146° and 127°.

acetamide) equilibrium of the C5'-DH bond, f_o is the fraction of the rotamer population within the gg+ga states (f_{gg+ga}) that has the DH proton rotated toward the acetamide (f_{ta}). Thus f_o is given by

$$f_o = f_{gg+ga} \times f_{ta}$$

Figure 15. Newman projections of the gg, ga, and ag rotamers about the C5-C5' bond in 5'-sp^3 $AcHel_1$ derivatives (DH stands for a generalized donor).



gg ga ag

The values of f_{gg+ga} for $AcHel_1$ derivatives can be estimated using the coupling constants between the pro-R 5' proton and the 5 proton ($^3J_{5'R,obs}$) and the pro-S 5' proton and the 5 proton ($^3J_{5'S,obs}$) in the t and c states of the appropriate reference compounds or the c state of the donor itself (in either case the equilibrium should be unperturbed by hydrogen bonding). These two coupling constants are weighted averages of the coupling constants in the pure gg, ga, and ag states:

$$^3J_{5'R,obs} = f_{gg} \times {}^3J_{5'R,gg} + f_{ga} \times {}^3J_{5'R,ga} + f_{ag} \times {}^3J_{5'R,ag} \qquad \textbf{1 - 9}$$

$$^3J_{5'S,obs} = f_{gg} \times {}^3J_{5'S,gg} + f_{ga} \times {}^3J_{5'S,ga} + f_{ag} \times {}^3J_{5'S,ag}$$

where f_{gg}, f_{ga}, and f_{ag} are the mole fractions of the gg, ga, and ag states ($f_{gg+ga} = f_{gg} + f_{ga}$), ${}^3J_{5'R,gg}$, ${}^3J_{5'R,ga}$, and ${}^3J_{5'R,ag}$ are the limiting coupling constants for the pure gg, ag, and ga states between the pro-R 5' and 5 protons, and ${}^3J_{5'S,gg}$, ${}^3J_{5'S,ga}$, and ${}^3J_{5'S,ag}$ are the limiting coupling constants for the pure gg, ga, and ag states between the pro-S 5' and 5 protons. Note that, although it was straightforward to determine which two peaks in the ^{1}H-NMR spectrum of a 5'-sp^3 AcHel$_1$ derivative belong to the 5' protons (either using a 2D-COSY spectrum or by analogy to the spectra of other AcHel$_1$ derivatives that have been fully assigned), it could not be decided which of the two peaks was due to the pro-R and which to the pro-S proton. If the two peaks could have been assigned unambiguously, then the peaks' coupling constants could in turn have been assigned to ${}^3J_{5'R,obs}$ and ${}^3J_{5'S,obs}$, and the two equations 1 - 9 along with the condition $1 = f_{gg} + f_{ga} + f_{ag}$ could have been solved to provide f_{gg}, f_{ga}, and f_{ag}. Since the peaks could not be assigned unambiguously, either of the two peaks' coupling constants could be ${}^3J_{5'R,obs}$ with the other being ${}^3J_{5'S,obs}$. The equations 1-9 and $1 = f_{gg} + f_{ga} + f_{ag}$ therefore have to be solved for both possible assignments, yielding two sets of f_{gg}, f_{ga}, and f_{ag} that lead to two values of f_{gg+ga}. Both of these will be considered simultaneously in further calculations.

For each 5'-sp^3 AcHel$_1$ derivative, the limiting coupling constants between the pro-R and S 5' protons and the 5 proton for the gg, ga, and ag states were calculated as a function of the dihedral angle between the coupled protons and the electronegativity and orientation of the substituents using the Haasnoot-Altona modification of the Karplus equation.[84] Both possible assignments of ${}^3J_{5'R,obs}$ and ${}^3J_{5'S,obs}$ for several of the 5'-sp^3

[84] Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*. 2783. For 5'-sp^3 AcHel$_1$ derivatives with O attached to C5', $J_{R,gg}$ = 1.4 Hz, $J_{R,ga}$ = 10.8 Hz, $J_{R,ag}$ = 4.0 Hz, $J_{S,gg}$ = 2.1 Hz, $J_{S,ga}$ = 3.3 Hz, $J_{S,ag}$ = 10.8 Hz; with N attached to C5', $J_{R,gg}$ = 2.1 Hz, $J_{R,ga}$ = 11.3 Hz, $J_{R,ag}$ = 3.6 Hz, $J_{S,gg}$ = 2.8 Hz, $J_{S,ga}$

$AcHel_1$ derivatives are listed in table 11, along with the resulting populations of the gg, ga, and ag states and, from them, f_{gg+ga}.

Table 11. Coupling constants between the pro-R 5' and 5 protons and the pro-S 5' and 5 protons for the t and c states of reference 5'-sp^3 $AcHel_1$ derivatives and the c states of donor 5'-sp^3 $AcHel_1$ derivatives. Note that the c state coupling constants for $AcHel_1CH_2OAc$ and the t and c state coupling constants for $AcHel_1CH_2NMe_2$ could not be determined due to peak overlap.

Compound	$^3J_{5'R,obs}$, $^3J_{5'S,obs}$ (Hz)	f_{gg}, f_{ga}, f_{ag}	f_{gg+ga}
$AcHel_1CH_2OMe$ (t)	3.2, 8.1 or 8.1, 3.2	0.31, 0.01, 0.68 or 0.27, 0.70, 0.03	0.32 or 0.97
$AcHel_1CH_2OMe$ (c)	3.2, 7.6 or 7.6, 3.2	0.35, 0.02, 0.63 or 0.32, 0.64, 0.04	0.37 or 0.96
$AcHel_1CH_2OAc$ (t)	3.5, 6.3 or 6.3, 3.5	0.44, 0.09, 0.47 or 0.42, 0.49, 0.09	0.53 or 0.91
$AcHel_1CH_2OH$ (c)	3.3, 7.3 or 7.3, 3.3	0.36, 0.04, 0.60 or 0.33, 0.62, 0.05	0.40 or 0.95
$AcHel_1CH_2SH$ (c)	2.7, 9.1 or 9.1, 2.7	0.37, -0.04, 0.67 or 0.32, 0.70, -0.02	0.33 or 1
$AcHel_1CH_2NHAc$ (c)	4.2, 7.9 or 7.9, 4.2	0.27, 0.12, 0.61 or 0.24, 0.61, 0.15	0.39 or 0.85
$AcHel_1CH_2NH_2$ (c)	3.3, 9.4 or 9.4, 3.3	0.22, 0.0, 0.78 or 0.17, 0.79, 0.04	0.22 or 0.96

This method of determining rotameric populations is not without faults, since 1) the limiting coupling constants are calculated from an empirical model and 2) it is assumed that the 5'H-5H dihedral angles in the three staggered states are exactly 60°, 180° or

---

= 3.0 Hz, $J_{S,ag}$ = 11.3 Hz; with S attached to C5', $J_{R,gg}$ = 2.5 Hz, $J_{R,ga}$ = 11.9 Hz, $J_{R,ag}$ = 3.4 Hz, $J_{S,gg}$ = 3.2 Hz, $J_{S,ga}$ = 2.7 Hz, $J_{S,ag}$ = 11.9 Hz

−60°, when steric or electronic factors could cause them to deviate from these idealized geometries. Nevertheless, a pattern is visible in the data of table 11. The values of f_{gg+ga} when $^3J_{5'R,obs}$ is assigned the small and $^3J_{5'S,obs}$ the large coupling constant tend to be small, with an average of 0.37. With the opposite assignment they tend to be large, with an average of 0.94. These average values, 0.37 and 0.94, will be taken to represent the two extreme possibilities for f_{gg+ga}.

The fraction of states with the (toward acetamide) C5'-DH conformation in the gg or ga states, f_{ta}, can be estimated by assuming that this torsion has no preferences among its accessible conformations. Under this assumption, f_{ta} is just the number of hydrogens on the donor divided by the number of accessible conformations. Since the three staggered conformations about the C5'-OH, C5'-SH, and C5'-NH_2 bonds ought to be accessible, this is 1/3, 1/3, and 2/3 respectively for $AcHel_1CH_2OH$, $AcHel_1CH_2SH$, and $AcHel_1CH_2NH_2$. The situation is less clear cut for $AcHel_1CH_2NHAc$, since primary amides have weak conformational preferences about the C(O)N-C bond.[85] It is taken to be 1/3 since sp^2 atoms (like the amide N in $AcHel_1CH_2NHAc$) bonded to sp^3 atoms (like C5') usually have 4 accessible conformations (two eclipsing and two bisected), but for amides one of the eclipsing conformations (where the C5-C5'-N-C(O) dihedral angle is 0°) is highly disfavored.[85]

The lack of a good reference compound complicates the determination of f_{gg+ga} for $AcHel_1CH_2NH_3^+$ and $AcHel_1CH_2NMe_2H^+$. As noted previously, the equilibrium about the C5-C5' bond is probably different in $AcHel_1CH_2NMe_3^+$ and $AcHel_1CH_2NH_3^+$ because of the size of the trimethylammonium group. Furthermore, both of the t state H5' resonances are overlap with other peaks in the ^{1}H-NMR spectrum of $AcHel_1CH_2NMe_3^+$,

and coupling constants cannot be obtained from them. A different tactic must be employed to determine f_{gg+ga} for the 5'-charged $AcHel_1$ derivatives.

The electrostatic interactions between the positively charged donors in the 5'-charged $AcHel_1$ derivatives and the acetamide carbonyl should make them more prone to adopting the gg or ga conformations, so that f_{gg+ga} for the 5'-charged $AcHel_1$ derivatives should be greater than f_{gg+ga} for the 5'-sp^3 $AcHel_1$ derivatives. The value of f_{gg+ga} for the 5'-sp^3 $AcHel_1$ derivatives could therefore be taken as a lower limit on f_{gg+ga} for the 5'-charged $AcHel_1$ derivatives. Under this assumption, f_{gg+ga} for the 5'-charged $AcHel_1$ derivatives is at least 0.37, the lower limit of the range determined for 5'-sp^3 $AcHel_1$ derivatives.

The value of f_{ta} should be one for $AcHel_1CH_2NH_3^+$ because of the symmetry of the ammonium ion. The same fraction for $AcHel_1CH_2NMe_2H^+$ is more difficult to determine because the bulk of the dimethylammonium ion makes some of the staggered conformations about the C5'-NMe_2H^+ bond inaccessible. In particular, two conformations about the C5'-NMe_2H^+ bond within the (gg, ga) state of the C5-C5' torsion can be immediately eliminated from consideration because of severe steric repulsion. In neither the gg nor the ga state can the C5-C5'-N-H dihedral angle be 180° because in the gg state this places both of the aminomethyl groups close to the N4-C5-C6-C7-C8 pyrrolidine ring while in the ga state this places both of the aminomethyl groups close to the C3 carbonyl. The remaining conformations will be regarded as energetically equivalent, so that f_{ta} will be set at 1/2 for this compound.

Finally, because the benzimidazolium ion has an axis of symmetry, there are only two equivalent conformations about the C5-C5' torsion for $AcHel_1benzim^+$ (according to molecular mechanics calculations). The value of f_o for this derivative ought to be 1.

---

[85] Chakrabarti, P.; Dunitz, J. D. *Helv. Chim. Acta* **1982**, *65*, 1555.

The f_o values (or ranges) for the 5'-acyl, 5'-sp^3, and 5'-charged $AcHel_l$ derivatives are listed in table 12.[86]

Table 12. Values (or ranges) of f_{gg+ga}, f_{ta}, and f_o for $AcHel_l$ derivatives. Dashes indicate that the quantity was not explicitly determined.

$AcHel_l$ derivative	type	f_{gg+ga}		f_{ta}	f_o	
		min	max		min	max
$AcHel_lCONH_2$	5'-acyl	-		-	0.32	
$AcHel_l$benzim	5'-acyl	-		-	0.32	
$AcHel_lCH_2OH$	5'-sp^3	0.37	0.94	1/3	0.12	0.31
$AcHel_lCH_2SH$	5'-sp^3	0.37	0.94	1/3	0.12	0.31
$AcHel_lCH_2NH_2$	5'-sp^3	0.37	0.94	2/3	0.25	0.63
$AcHel_lCH_2NHAc$	5'-sp^3	0.37	0.94	1/3	0.12	0.31
$AcHel_lCH_2NMe_2H^+$	5'-charged	0.37	-	1/2	0.19	-
$AcHel_lCH_2NH_3^+$	5'-charged	0.37	-	1	0.37	-
AcHel1benzim$^+$	5'-charged	-		-	1	

The value of f_o for the 5'-sp^3 $AcHel_l$ derivatives depends on how the coupling constants $^3J_{5'R}$ and $^3J_{5'S}$ are assigned. At one extreme, f_o is about the same for 5'-sp^3

[86] It has been implicitly assumed in the above method for determining f_o for the 5' acyl and 5' sp^3 $AcHel_l$ derivatives that the hydrogen bond donors cannot interact with the C3 carbonyl to form a γ-turn-like hydrogen bond. To the extent that this hydrogen bond occurred it would decrease f_o, since formation of this hydrogen bond would interfere with the orientation of the donor toward the acetamide carbonyl. This hydrogen bond, however, has a poor geometry (the D-H···O angle is 147° and the H···O=C angle is 107°), and it has been shown not to occur in acetyl proline N-methylamide in aqueous solution (Madison, V.; Kopple, K. D. *J. Am. Chem. Soc.* **1980**, *102*, 4855). γ-turn hydrogen bonding probably does not strongly influence f_o.

$AcHel_1$ derivatives as for the 5'-acyl $AcHel_1$ derivatives. At the other extreme, f_o for 5'-sp^3 $AcHel_1$ derivatives is about 1/3 of what it is for 5'-acyl $AcHel_1$ derivatives. In any case, f_o only varies over a factor of three, so it should not broaden the range calculated for the intrinsic hydrogen bonding energies overmuch. The minimum values for f_o for 5'-charged $AcHel_1$ derivatives are generally higher than both f_o for 5'-acyl $AcHel_1$ derivatives and the upper bounds on f_o for the 5'-sp^3 $AcHel_1$ derivatives. This can be attributed to the symmetry of the ammonium ion in $AcHel_1CH_2NH_3^+$, for which all conformations about the C5'-NH_3^+ bond are equivalent, and the benzimidazolium ion in $AcHel_1benzim^+$, for which both conformations about C5-C5' are equivalent. The influence of f_o on the calculated value of K_{hb} roughly increases in the order 5'-charged < 5'-acyl < 5'-sp^3.

1.5 Discussion of Intrinsic Hydrogen Bonding Equilibrium Constants and Free Energies

All three of the quantities needed to calculate intrinsic hydrogen bonding equilibrium constants using equation 1-4 have now been determined. The values of $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ from table 9, divided by $(\gamma + K_{se})/(1 + K_{se})$ from table 10, and f_o from table 12 yield the values of K_{hb} listed below in table 13. The relation $\Delta G_{hb} = -RTlnK_{hb}$ gives the corresponding intrinsic free energies of hydrogen bonding (T = 25 °C).

Table 13. Upper and lower bounds on the intrinsic hydrogen bonding equilibrium constants and free energies of hydrogen bond formation in D_2O at 25 °C for 5'-acyl, 5'-sp^3, and 5'-charged $AcHel_1$ derivatives.

$AcHel_1$ derivative	K_{hb} in D_2O at 25 °C		ΔG_{hb} in D_2O at 25 °C	
	min	max	max	min
$AcHel_1CONH_2$	16	42	-1.6 kcal/mol	-2.2 kcal/mol
$AcHel_1$benzim	17	46	-1.7 kcal/mol	-2.3 kcal/mol
$AcHel_1CH_2OH$	0.75	1.9	+0.2 kcal/mol	-0.4 kcal/mol
$AcHel_1CH_2SH$	0.81	2.1	+0.1 kcal/mol	-0.4 kcal/mol
$AcHel_1CH_2NH_2$	1.6	4.1	-0.3 kcal/mol	-0.8 kcal/mol
$AcHel_1CH_2NHAc$	3.4	8.8	-0.7 kcal/mol	-1.3 kcal/mol
$AcHel_1CH_2NH_3^+$	-	2.4	-	-0.5 kcal/mol
$AcHel_1CH_2NMe_2H^+$	-	4.0	-	-0.8 kcal/mol
$AcHel_1$benzim$^+$	-	3.1	-	-0.7 kcal/mol

Before the data are discussed, two questions will be addressed. The first is, how reliable are the data in table 13? Three quantities had to be determined to arrive at the values of K_{hb} in the table: $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$, $(\gamma + K_{se})/(1 + K_{se})$, and f_o. For the cases where closely related $AcHel_1$ reference derivatives were available ($AcHel_1CONH_2$, $-CH_2OH$, $-CH_2SH$, $-CH_2NH_2$, $-CH_2NHAc$), the first of these quantities probably represents the t/c ratio corrected for the intrinsic t/c bias, charge-dipole and dipole-dipole interactions very accurately. Even if errors as large as a factor of two occurred in the estimation of K_{se}, γ, or f_o, the intrinsic free energies of hydrogen bond formation reported for these compounds would still be correct to within about 0.4 kcal/mol. For the other

cases ($AcHel_1$benzim, $-CH_2NH_3^+$, $-CH_2NMe_2H^+$, and -benzim$^+$), there is more chance for a significant error. However, for the 5'-charged $AcHel_1$ derivatives, no error should carry the actual intrinsic hydrogen bonding energies higher than the maximum values quoted for them. These are reliable upper limits; the actual values could only be lower. The second question is, how applicable are these data to biological folding and recognition problems? The hydrogen bonds that form in the $AcHel_1$ derivatives are held between $AcHel_1$'s relatively hydrophobic core and the solvent, an environment that is probably similar to a protein surface. What is learned about hydrogen bonding from $AcHel_1$ derivatives is at least pertinent to the thermodynamics of hydrogen bonds that reside at protein surfaces. However, applying these data to buried hydrogen bonds would require a correction term for the transfer of the hydrogen bond from a partially solvent exposed state to the interior of a folded protein or a protein-substrate complex.

The most common, and therefore most important, hydrogen bonds in proteins occur between amide NHs and amide carbonyls (these account for 40% – 55%of all hydrogen bonds formed in proteins).[87] According to table 13, this type of hydrogen bond is the strongest studied in this chapter. Its intrinsic energy apparently depends on the amide's orientation (perhaps because of subtle differences in geometry, or because of orientation dependent carbonyl-carbonyl interactions[88,89]), with -0.7 kcal/mol $\geq \Delta G_{hb} \geq -1.3$ kcal/mol for $AcHel_1CH_2NHAc$ and -1.6 kcal/mol $\geq \Delta G_{hb} \geq -2.2$ kcal/mol for $AcHel_1CONH_2$. Either way, the data indicate that amide-amide hydrogen bonds favor folded states when they occur on protein surfaces, and, unless desolvating the hydrogen

---

[87] Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Molec. Biol.* **1984**, *44*, 97.
[88] Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R. *Acta Cryst. B*, **1998**, *54*, 320.
[89] Maccallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol.* **1995**, *248*, 361.

bonded NH···O=C unit is very unfavorable, they probably favor folded states when they occur in protein interiors as well. Thus, around half of the hydrogen bonds formed in protein folding favor the folded state.

Hydrogen bonds between alcohols or thiols and amide carbonyls are less common in proteins,[87] and are weaker than amide-amide hydrogen bonds according to the intrinsic hydrogen bonding free energies for $AcHel_1CH_2OH$ and $AcHel_1CH_2SH$. This is not remarkable, as there is no reason to believe that a hydroxyl group would be able to out-compete the solvent as a donor to the acetamide carbonyl, and thiols are known to be poor hydrogen bond donors from studies in non-competitive media.[26,27] It is surprising, however, that their intrinsic hydrogen bonding free energies should be the same. This could be due to the leveling effect of aqueous solvents on hydrogen bonding. If the hydrogen bonding of both alcohols and thiols were at the limit of detection, then whether or not alcohols are actually better donors than thiols, they would appear to be the same. The relative donor strengths of alcohols and thiols can be clarified using media in which effects due to hydrogen bonding are more pronounced, such as water-trifluoroethanol (TFE) mixtures. It has been shown that hydrogen bonding between donors and amide carbonyls is enhanced when TFE is added to water, because the solvation of amide carbonyls worsens with increasing TFE concentration.[71] Based on this mechanism, one would expect any interaction with the amide carbonyl dipole, hydrogen bonding as well as pure electrostatic, to be augmented by TFE. The response of the t/c ratios of the $AcHel_1$ derivatives relative to the t/c ratios of their reference $AcHel_1$ derivatives is examined in appendix 2. Pertinent to the relative hydrogen bonding strengths of alcohols

and thiols, it is shown that alcohols form stronger hydrogen bonds than thiols in 10 mol% TFE in D_2O.

Unprotonated amines do not often occur at the pHs typically encountered by proteins. Still, the intrinsic hydrogen bonding energy found in $AcHel_1CH_2NH_2$ merits a comment. Given the low donor strengths found for amines in non-competitive solvents,[26,27] it is surprising that they seem to form stronger hydrogen bonds with amide carbonyls than alcohols and thiols do. This observation can be rationalized by supposing that the amine's ability to donate a hydrogen bond is increased by its accepting a hydrogen bond from the solvent. Such cooperativity in hydrogen bonding is a well-known phenomenon,[11,90] and it has been shown to account for substantial fractions of the total binding energy in water clusters.[91,92]

Hydrogen bonds between charged donors, such as ammonium ions, and amide carbonyls are not uncommon in proteins (in about half of hydrogen bonds formed by lysine side chains, the acceptor is an amide carbonyl).[87] According to results from the incremental energy method, such hydrogen bonds should be much stronger than those in which the donor is neutral.[48,53,58] However, the intrinsic hydrogen bonding energies listed in table 13 for the 5'-charged $AcHel_1$ derivatives indicate that they do not form strong hydrogen bonds with the acetamide carbonyl. In fact, considering that these intrinsic energies are upper bounds, the charged donors are probably not significantly better donors than the alcohol, thiol, or free amine. This seems to contradict what is believed about hydrogen bonding based on the incremental energy method. However, it must be

---

[90] Kleeberg, H. in *Intermolecular Forces*; Huyskens, P. L.; Luck, W. A. P.; Zeegers-Huyskens, T., Eds.; Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, **1991**, Chapter 10, pp. 251-280.

[91] Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.

stressed that this does not mean that the interactions between the charged 5' donors and the acetamide carbonyl are weak. In fact, their t/c ratios show that the t state is more strongly stabilized by far in the 5'-charged $AcHel_1$ derivatives than in any of the neutral compounds. It is just that, since the non-donor $AcHel_1CH_2NMe_3^+$ shares the most part of this stabilization, the majority of the effect must be attributed to simple charge-dipole interactions rather than hydrogen bonding. It has been suggested that desolvation of ions is a sufficiently unfavorable step as to make the overall process of salt bridge formation in protein interiors unfavorable.[93] Perhaps hydrogen bonding with the acetamide carbonyl cannot compensate for the desolvation of the donors in the 5'-charged $AcHel_1$ derivatives, and the majority of the donor-acetamide interaction is purely electrostatic as a result.

This chapter has addressed the hydrogen bonding of various donors with an acetamide carbonyl. In some applications, for example, the design of small molecules that bind tightly to proteins, the intrinsic hydrogen bonding energetics are not of interest. The relevant issue is how an overall interaction can be maximized. We note in closing that the overall interactions of the donors in this study with the acetamide carbonyl can be evaluated by comparing the stability of the t states in their $AcHel_1$ derivatives. Taking the t state of $AcHel_1CH_2OMe$ as the standard, the free energies of the overall 5-substituent-acetamide interactions, $\Delta\Delta G_{overall}$ are compared in table 14 (the entries in the table are $\Delta\Delta G_{overall} = -RTln[t/c\ ratio] + RTln[t/c\ ratio\ of\ AcHel_1CH_2OMe]$).

---

[92] Cruzan, J. D.; Braly, L. B.; Liu, K.; Brown, M. G.; Loeser, J. G.; Saykally, R. J. *Science* **1996**, *271*, 59.
[93] Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211.

Table 14. Energetics of overall donor-acetamide interactions in $AcHel_1$ derivatives.

$AcHel_1$ derivative	$\Delta\Delta G_{overall}$
$AcHel_1CONH_2$	−0.09 kcal/mol
$AcHel_1$benzim	−0.11 kcal/mol
$AcHel_1CH_2OH$	−0.07 kcal/mol
$AcHel_1CH_2SH$	−0.08 kcal/mol
$AcHel_1CH_2NH_2$	−0.19 kcal/mol
$AcHel_1CH_2NHAc$	−0.17 kcal/mol
$AcHel_1CH_2NMe_3^+$	−0.46 kcal/mol
$AcHel_1CH_2NH_3^+$	−0.70 kcal/mol
$AcHel_1CH_2NMe_2H^+$	−0.67 kcal/mol
$AcHel_1$benzim$^+$	−0.93 kcal/mol

The data in the table show that the interactions between the substituents at the 5 position and the acetamide in the 5'-acyl and 5'-sp^3 $AcHel_1$ derivatives do not stabilize the t state much beyond its inherent stability in $AcHel_1CH_2OMe$. Only the charge-dipole interactions in the 5'-charged $AcHel_1$ derivatives, in particular that involving the benzimidazolium ion in $AcHel_1$benzim$^+$, have a large impact on the t state's stability. Thus, the optimal partner for an amide CO in a binding interaction is probably a benzimidazolium or imidazolium ion, even though these are not optimal hydrogen bond donors.

1.6 Experimental

Equipment. ^{1}H-NMR spectra were measured on Varian VXR500S and 501S spectrometers and processed using the Varian Instruments VNMR 3.1 software. Chemical Shifts are reported relative to the reference signal of (trimethylsilyl)propionic-2,2,3,3-d_4 acid (TMSP) for spectra obtained in D_2O or relative to the residual solvent signal for spectra acquired in other solvents. The pH of samples in D_2O was adjusted as required by the addition of either TFA-d_1 or 40 wt% NaOD in D_2O. Analytical high performance liquid chromatography (HPLC) was performed on a Waters system consisting of two 501 pumps, a rheodyne injector, a model 660 automated gradient controller, a model 740 data module, a model 484 detector, and a Vydac 0.46 × 25 cm (218TP54) C_{18} reverse phase column. Flow rates were 1.0 mL/min. Preparative scale HPLC was performed on a Waters system consisting of a model 590 pump fitted with preparative heads, an Autochrome DPG/S pre-pump solvent mixer, a Rheodyne injector, a model 484 variable wavelength detector, and a Waters 2.5 × 10 cm radial compression column housed in a PrepLC 2.5 cm radial compression module (RCM). Flow rates for preparative HPLC were 12 mL/min. Detection in all uses of HPLC was carried out at 214 nm unless otherwise specified. Mass spectra were measured by the Massachusetts Institute of Technology Department of Chemistry Instrumentation Facility on a Finnegan MAT 8200 mass spectrometer using a glycerol/methanol matrix. Thin layer chromatography (TLC) was performed using Analtech Uniplate 2000 μ silica gel TLC plates, while flash chromatography was carried out using Merck Kieselgel 60. Solvents were generally taken from newly opened bottles of low moisture content without further

purification, except for THF, which was distilled from sodium/benzophenone. Molecular mechanics calculations were carried out using the CHARMm force field as implemented in Quanta 97 (Molecular Simulations Inc., 1997) on a Silicon graphics O_2.

Measurement of t/c Ratios. The t/c ratio is the ratio of the integrals of the peaks due to a single t state proton and a single c state proton in a ^{1}H-NMR spectrum. Accurate values for the single proton integrals were obtained as follows. The NMR spectrum of an $AcHel_1$ derivative was first phased to pure absorption and the integration line was adjusted using the level and tilt parameters to minimize baseline roll. Integrals were measured for each peak or group of peaks for which the contributing proton or protons could be assigned (the integration line was cut at the point where the peak reached baseline on the downfield and upfield sides of the peak). Peaks were assigned either by analogy with the spectra of similar compounds or, if necessary, using a 2D-COSY spectrum. Peaks that were very close to the HDO signal in water spectra (since they were often distorted) and peaks that were very much larger than the average in the spectrum (since they would have too much leverage in the regression) were excluded from the analysis. A two variable linear regression was applied to the data according to the model below.

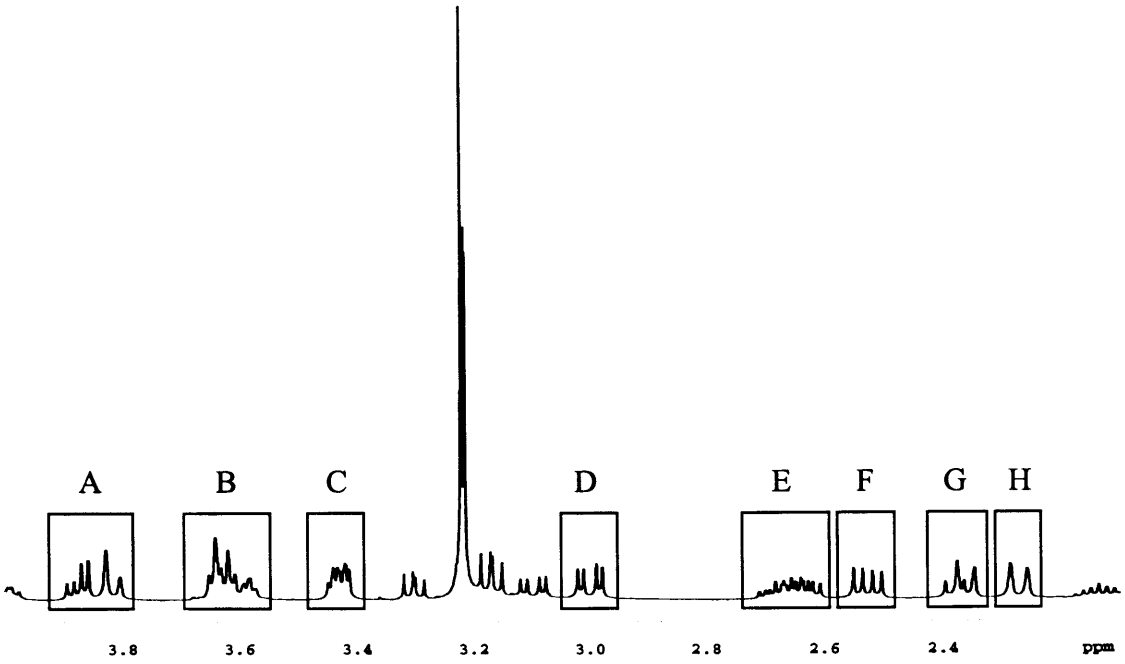
Integral = (# of t state protons)×(integral of a single t state proton) +

(# of c state protons)×(integral of a single c state proton)

In this model, the integral is the dependent variable and the numbers of t and c state protons under the integral are the independent variables. The integrals of single t and c state protons are the parameters to be determined (a constant has been left out of the model, since the integral should be 0 when there are no t or c state protons; also, in

practice, the constants were rarely statistically significant). An example of this integration procedure is shown in figure 13, using the spectrum of $AcHel_1CH_2OMe$ in D_2O. The data are summarized in table 15.

Figure 16. An example of a spectrum of an $AcHel_1$ derivative ($AcHel_1CH_2OMe$ in this case) with several regions integrated. The assignment of each region and the values of the integrals are listed below the spectrum.



A: H12a and 12b t state signals, integral = 28.97 for 2 t protons
B: H12a and 12b c state signals and H11 t and c state signals integral = 22.76 for 1t and 3c protons
C: H5'a t and c state signals integral = 14.68 for 1t and 1c proton
D: H9a t state signal integral = 14.68 for 1t proton
E: H13a t and c state signals integral = 22.54 for 1t and 1c proton
F: H9b t state signal integral = 13.91 for 1t proton
G: H9b c state and H13b c state signals integral = 16.63 for 2c protons
H: H13b t state signal integral = 13.68 for 1t proton

Table 15. Summary of data from figure 16.

Region	Integral	# of t state protons	# of c state protons
A	28.97	2	0
B	41.52	1	3
C	22.76	1	1
D	14.68	1	0
E	22.54	1	1
F	13.91	1	0
G	16.63	0	2
H	13.68	1	0

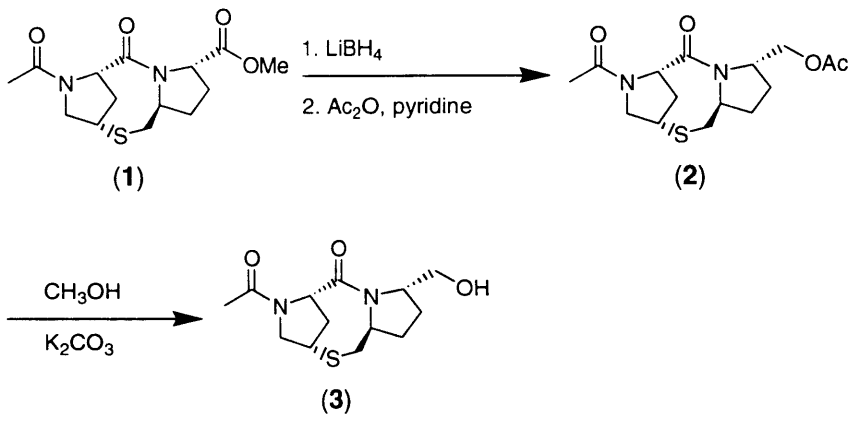
The regression of the integral vs. the numbers of t and c state protons yields 14.319 (standard error = 0.317) for a single t state proton and 8.968 (standard error = 0.259) for a single c state proton. This yields a t/c ratio of 1.597 with a standard error of 0.058 (the error has been propagated in the standard way for quotients).

Synthesis: Description

It was noted previously that several of the $AcHel_1$ derivatives included in this thesis ($AcHel_1CO_2Me$, $AcHel_1CO_2H$, $AcHel_1CH_2OH$, $AcHel_1CH_2OAc$, $AcHel_1CH_2NH_2$, $AcHel_1CH_2NHAc$, $AcHel_1CONH_2$, and $AcHel_1CH_2SH$) had been synthesized prior to the work in this thesis. However, many of these syntheses (those for $AcHel_1CH_2OH$, $AcHel_1CH_2OAc$, $AcHel_1CH_2N_3$, $AcHel_1CH_2NH_2$, $AcHel_1CH_2NHAc$) have not yet appeared in the literature and are therefore presented here. References have been given for the others. All other compounds ($AcHel_1CH_2OMe$, $AcHel_1CH_2NMe_2$, $AcHel_1$benzim, $AcHel_1CH_2NMe_3^+$) were prepared in this thesis. All new compounds were characterized by ^{1}H-NMR, ^{13}C-NMR and high-resolution mass spectroscopy.

The syntheses of all compounds began with the methyl ester, $AcHel_1CO_2Me$ (**1**) whose preparation is described in the literature[94]. Reduction with $LiBH_4$[95] provided a mixture of the alcohol $AcHel_1CH_2OH$ and the amino alcohol $HHel_1CH_2OH$ (in which the acetamide group has been removed from N1). This product mixture was peracetylated with acetic anhydride without purification to yield the single product $AcHel_1CH_2OAc$ (**2**), which after methanolysis gave the alcohol $AcHel_1CH_2OH$ (**3**) (scheme 1).

Scheme 1.

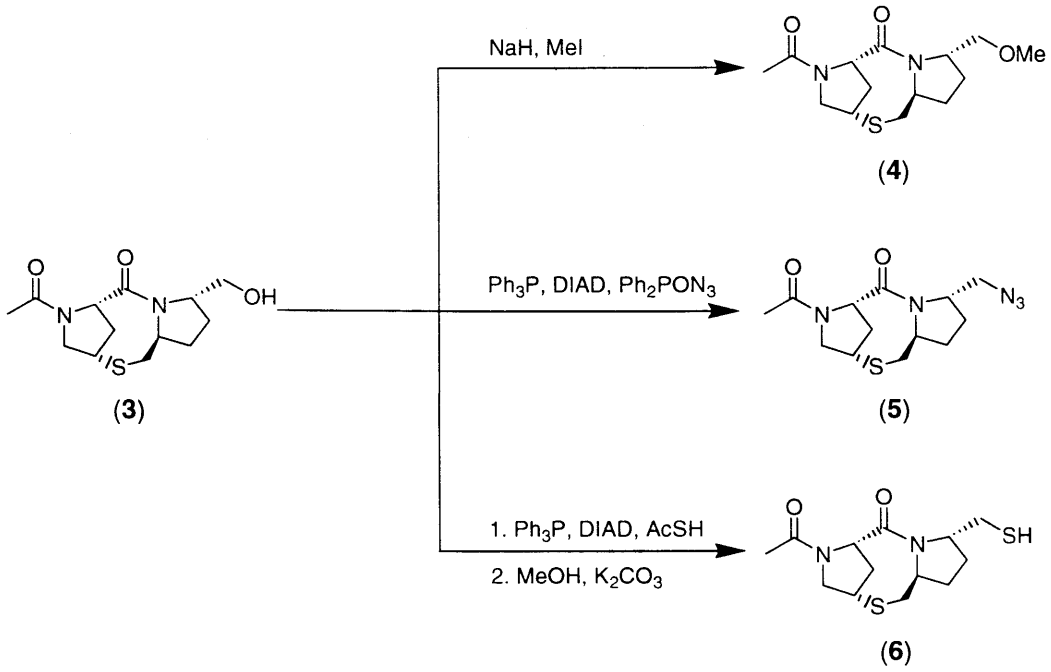


As shown in scheme 2, deprotonation of $AcHel_1CH_2OH$ and treatment with MeI yielded the methyl ether $AcHel_1CH_2OMe$ (**4**) (no alkylation of the thioether was observed). The azide **5**, the precursor to all the amino $AcHel_1$ derivatives, was produced by treatment of **3** with diphenylphosphoryl azide under Mitsonobu conditions.[96] The thiol $AcHel_1CH_2SH$ (**6**) was prepared by displacing the alcohol with thioacetic acid under Mitsonobu conditions followed by methanolysis of the resulting thioacetate.[74]

---

[94] McClure, K. F.; Renold, P.; Kemp, D. S. *J. Org. Chem.* **1995**, *60*, 454.
[95] Brown, H. C.; Narasimhan, S.; Choi, Y. M. *J. Org. Chem.* **1982**, *47*, 4702.

Scheme 2.

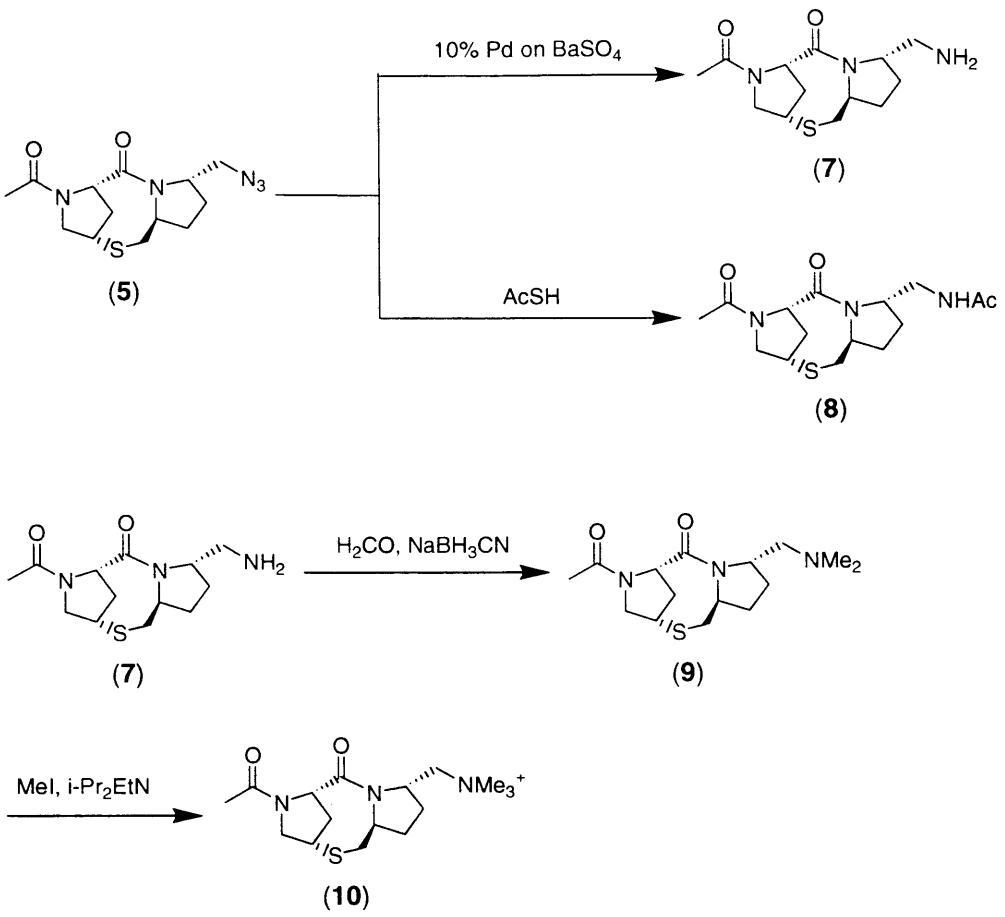


As shown in scheme 3, hydrogenation of the azide **5** with 10% Pd on $BaSO_4$ as catalyst yielded the amine $AcHel_1CH_2NH_2$ (**7**), while reduction of the azide with thioacetic acid[97] yielded the acetamide $AcHel_1CH_2NHAc$ (**8**). Treatment of $AcHel_1CH_2NH_2$ with formaldehyde and $NaBH_3CN$[98] in water at pH 7.4 yielded the dimethylamine $AcHel_1CH_2NMe_2$ (**9**). The trimethylammonium ion $AcHel_1CH_2NMe_3^+$ (**10**) was produced by the action of methyl iodide on $AcHel_1CH_2NMe_2$ (again, no alkylation of the thioether was observed).

---

[96] Lal, B.; Pramanik, B. N.; Manhas, M. S.; Bose, A. K. *Tet. Lett.* **1977**, 1977.
[97] Rosen, T.; Lico, I. M.; Chu, D. T. W. *J. Org. Chem.* **1988**, *53*, 1580.
[98] Lane, C. F. *Synthesis* **1975**, 137.

Scheme 3.

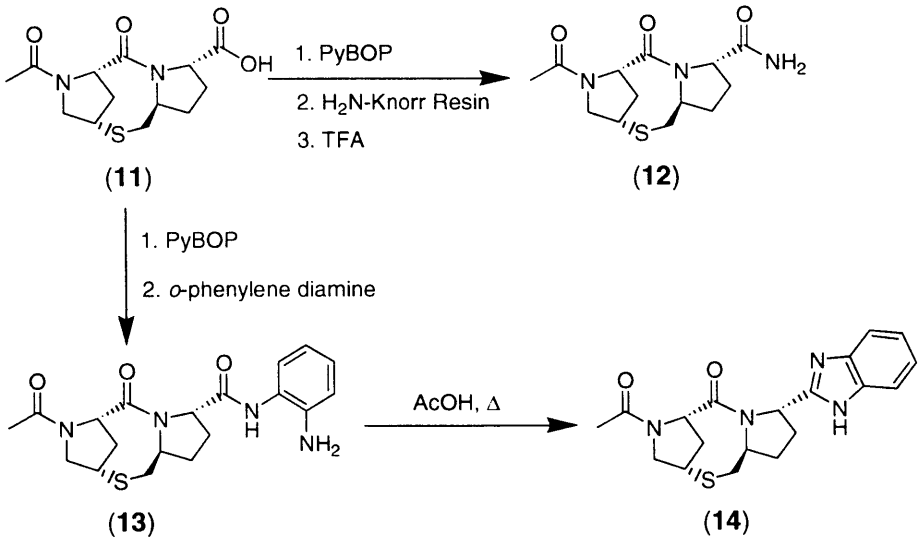


As shown in scheme 4, the simple amide derivative of $AcHel_1$ was prepared by activation of the acid $AcHel_1CO_2H$ (**11**) (prepared as previously described[99]) with PyBOP and coupling to Knorr peptide synthesis resin (2,4-dimethoxy-4'-(carboxymethyloxy)-benzhydrylamine linked to polystyrene). Treatment of the resin with TFA liberated the desired amide $AcHel_1CONH_2$ (**12**). The benzimidazole derivative of $AcHel_1$ was prepared by coupling $AcHel_1CO_2H$ with *o*-phenylene diamine. Heating the resulting *o*-

[99] Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd, J. G.; Muendel, C. *J. Org. Chem.* **1991**, *56*, 6672.

amino anilide (**13**) in acetic acid[100] yielded AcHel$_1$benzim (**14**). This reaction had to be timed carefully because a by-product, possibly epimerized at the 5 position, built up as the reaction progressed. This by-product was, fortunately, separable by preparative HPLC.

Scheme 4.



Synthesis: Procedures and Characterization

AcHel$_1$CH$_2$OAc (**2**). To a stirred solution of AcHel$_1$CO$_2$Me (0.18g, 0.58 mmol) in dry THF were added 0.055 g (2.5 mmol) of $LiBH_4$. The reaction was stirred at 25°C for 19h, then filtered through a fritted glass funnel, rinsing with THF. The filtrate was concentrated in vacuo, dissolved in MeOH and concentrated again. After repeating this twice more, the residue was taken up in 6 mL of pyridine and treated with 4 mL of acetic

[100] Grimmet, M. R. in *Comprehensive Heterocyclic Chemistry*; Katritzky, A.; Rees, C. W., Eds. in chief; Pergamon Press: Oxford, New York, Toronto, Sydney, Paris, Frankfurt, **1984**; vol. 5, Chapter 4, pp. 457-498.

anhydride. After 2h, the reaction was again concentrated and the residue was purified by flash chromatography (9:1 EtOAc: MeOH) to yield 150 mg (0.49 mmol, 79%) of $AcHel_1CH_2OAc$ as a lightly colored oil. R_f (8:2 EtOAc:MeOH): 0.33. ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.81 (d, *J* = 8.9 Hz, 0.43 H), 4.69 (d, *J* = 9.8 Hz, 0.57 H), 4.4 (m, 1.43 H), 4.27 (m, 0.53 H), 4.23 (m, 0.43 H), 4.18 (dd, *J* = 6.1, 11.0 Hz, 0.53H), 4.15 (dd, *J* = 2.9, 11.0 Hz, 0.43 H), 4.12 (dd, *J* = 6.1, 12.2 Hz, 0.57 H), 3.98 (d, *J* = 12.0 Hz, 0.57 H), 3.78 (m, 1.86 H), 3.28 (dd, *J* = 5.6, 15.6 Hz, 0.43 H), 3.17 (dd, *J* = 4.9, 15.9 Hz, 0.57 H), 2.82 (m, 1.0 H), 2.69 (dd, *J* = 7.81, 15.87 Hz, 0.57 H), 2.53 (m, 0.86 H), 2.43 (d, *J* = 14.2 Hz, 0.57 H), 2.33 (m, 1.0 H), 2.14 (s, 1.71 H), 2.11 (two s, 3.0 H), 2.06 (s, 1.29 H), 2.0 (m, 1.0 H), 1.85 (m, 2.0 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present): δ 173.3, 172.3, 171.9, 170.8, 170.7, 64.5, 64.1, 59.22 59.18, 58.8, 58.4, 56.5, 52.7, 44.8, 41.7, 40.0, 39.6, 38.6, 37.7, 34.4, 33.8, 24.6, 24.5, 23.0, 22.9, 21.5, 21.4. HR-FAB MS calcd for $C_{15}H_{22}N_2O_4S$ (M^+): 326.1300. Found: 326.1301.

<u>$AcHel_1CH_2OH$ (3).</u> To a solution of $AcHel_1CH_2OAc$ (0.038 g, 0.12 mmol) in 4.5 mL MeOH was added 0.5 mL of 10% Na_2CO_3 in water. After 8 h, the reaction was filtered and the solvent removed. The residue was taken up in CH_2Cl_2 and purified by flash chromatography (8:2 EtOAc:MeOH) to yield 0.02g (0.07 mmol, 58%) of $AcHel_1CH_2OH$ as an oil. R_f (8:2 EtOAc:MeOH): 0.25. ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.69 (d, *J* = 4.7 Hz, 0.36 H), 4.23 (m, 2.0 H), 4.04 (dd, *J* = 5.9, 12.1 Hz, 0.64 H), 3.98 (d, *J* = 12.1 Hz, 0.64 H), 3.77 (m, 2.08 H), 3.69 (dd, *J* = 3.2, 11.1 Hz, 0.64 H), 3.56 (dd, *J* = 7.3, 11.0 Hz, 0.36 H), 3.48 (dd, *J* = 7.5, 11.2 Hz, 0.64 H), 3.26 (dd, *J* = 5.6, 15.6 Hz, 0.36 H) , 3.17 (dd, *J* = 5.0, 15.9 Hz, 0.64 H), 2.81 (m, 1.0 H), 2.69 (dd, *J* = 7.8, 16.0 Hz, 0.64 H), 2.53 (m, 0.72 H), 2.44 (d, *J* = 14.0 Hz, 0.64 Hz), 2.26 (m, 1.0 H), 2.14

(s, 1.92 H), 2.07 (s, 1.08 H), 2.0 – 1.8 (m, 3.0 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present): δ 172.6, 172.0, 171.6, 170.9, 64.2, 64.0, 63.6, 63.4, 62.5, 62.4, 60.1, 58.5, 57.3, 53.1, 44.3, 41.7, 39.9, 38.9, 38.7, 37.8, 35.0, 33.6, 25.5, 24.3, 23.1, 23.0. EI-MS: 284.2 (M^+)

AcHel$_1$CH$_2$OMe (**4**). To a solution of AcHel$_1$CH$_2$OH (0.01 g, 0.035 mmol) in 2 mL of dry THF at 25 °C were added 0.01 g of NaH (80% dispersion in mineral oil, 0.33 mmol) followed by 5 μL of CH_3I (0.011 g, 0.08 mmol). After 4h, the residual NaH was quenched by the addition of 2 mL of MeOH and the solvent was removed. The residue was taken up in MeOH and purified by preparative HPLC (gradient: 5-100% CH_3CN over 20 min, remainder H_2O; retention time = 10.3 min) to yield 0.008 g (0.026 mmol, 74%) of AcHel$_1$CH$_2$OMe. R_f = 0.2 (9:1 EtOAc:MeOH). ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.67 (d, *J* = 9.8 Hz, 0.62 H), 4.34 (ddd, *J* = 3.2, 7.6, 15.9 Hz, 0.38 H), 4.27 (ddd, *J* = 2.9, 8.1, 15.9 Hz, 0.62 H), 4.18 (m, 1 H), 4.03 (dd, *J* = 5.9, 12.0 Hz, 0.62 H), 3.97 (d, *J* = 12.0 Hz, 0.62 H), 3.78 (m, 1.76 H), 3.59 (m, 1 H), 3.46 (dd, *J* = 7.6, 9.8 Hz, 0.38 H), 3.373 (s, 1.86 H), 3.368 (s, 1.14 H), 3.32 (dd, *J* = 8.3, 10.0 Hz, 0.62 H), 3.25 (dd, *J* = 5.6, 15.6 Hz, 0.38 H), 3.15 (dd, *J* = 4.9, 15.9 Hz, 0.62 H), 2.81 (m, 1 H), 2.68 (dd, *J* = 7.6, 15.9 Hz, 0.62 H), 2.51 (m, 0.76 H), 2.42 (d, *J* = 14.2 Hz, 0.62 H), 2.25 (m, 1 H), 2.13 (s, 1.86 H), 2.06 (s, 1.14 H), 1.9 (m, 3 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present): δ 172.8, 172.4, 171.2, 171.0, 72.8, 72.4, 64.2, 63.1, 60.0, 59.8, 59.5, 59.4, 58.9, 58.3, 56.0, 53.2, 44.3, 41.6, 40.2, 40.0, 38.5, 38.0, 34.2, 33.8, 24.7, 24.4, 22.9, 22.8. HR-FAB MS calcd for $C_{14}H_{22}N_2O_3S$ (M^+): 298.1351. Found: 298.1350.

AcHel$_1$CH$_2$N$_3$ (**5**). To a solution of solution of Ph_3P (0.036 g, 0.137 mmol) in 0.4 mL THF at 0 °C under an Ar atmosphere were added 27 μL of DIAD (0.137 mmol) with

vigorous stirring. The resulting colorless paste was stirred for 20 min at 0 °C before a solution of $AcHel_1CH_2OH$ (0.026 g, 0.091 mmol) in 0.6 mL of THF was added, followed by 30 μL of $(PhO)_2PON_3$. The reaction (still a suspension at this point) was warmed to 25 °C. After 3h, the reaction had become a lightly yellow solution. Water was added at this point (100 μL) to quench the remaining reagents, and after 15 min the solvent was removed. The residue was purified by flash chromatography (85:15 EtOAc:MeOH) to yield 0.019 g (0.063 mmol, 69%) of $AcHel_1CH_2N_3$. R_f = 0.40 (8:2 EtOAc:MeOH). ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.82 (d, *J* = 8.5 Hz, 0.36 H), 4.70 (d, *J* = 9.8, 0.64 H), 4.26 (m, 2 H0. 4.04 (dd, *J* = 5.9, 12.0 Hz, 0.64 H), 3.98 (d, *J* = 12.0 Hz, 0.64 H), 3.8 (m, 1.72 H), 3.60 (m, 0.72 H), 3.52 (dd, *J* = 3.2, 12.5 Hz, 0.64 H), 3.40 (dd, *J* = 7.3, 12.5 Hz, 0.64 H), 3.27 (dd, *J* = 5.7, 15.6 Hz, 0.36 H), 3.17 (dd, *J* = 4.9, 15.9, 0.64 H), 2.81 (m, 1 H), 2.68 (dd, *J* = 7.8, 15.9 Hz, 0.64 H), 2.52 (m, 0.72 H), 2.43 (d, *J* = 14.2, 0.64 H), 2.32 (m, 1 H), 2.14 (s, 1.92 H), 2.10 (s, 1.08 H), 2.0 (m, 1 H), 1.84 (m, 2 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present): δ 64.2, 63.1, 59.7, 59.6, 59.2, 58.5, 56.1, 53.2, 52.72, 52.69, 44.8, 41.7, 40.1, 40.0, 39.4, 37.6, 34.2, 33.5, 24.9, 23.0 (the carbonyl resonances were too weak to be identified). HR-FAB MS calcd for $C_{13}H_{19}N_5O_2S$ (M^+): 309.1259. Found: 309.1260.

<u>$AcHel_1CH_2NH_2$ (7).</u> To a solution of $AcHel_1CH_2N_3$ (0.007 g, 0.022 mmol) in MeOH with 1% diisopropylethyl amine were added 0.02 g of 10% Pd on $BaSO_4$. The solution was placed under a hydrogen atmosphere (15 psi). After 2h, analytical HPLC showd the reaction to be complete (gradient: 9 to 99% CH_3CN over 30 min, remainder H_2O (0.1% TFA); product retention time = 5.44 min), so the reaction mixture was transferred to a 15 mL Falcon tube and centrifuged to separate the catalyst. The supernatant was decanted

and concentrated, and the residue was purified by preparative HPLC (gradient: 5 to 100% CH_3CN over 20 min, remainder H_2O (0.1% TFA); retention time = 7.7 min) to yield 0.006 g of $AcHel_1CH_2NH_2$ as the TFA salt (0.015 mmol, 69%). ^{1}H-NMR (500 MHz, D_2O at pH 1, 2 conformations present): δ 4.70 (d, *J* = 9.3 Hz, 0.84 H), 4.49 (dd, *J* = 6.6, 8.1 Hz, 0.84 H), 4.4 (m, 0.16 H), 4.23 (m, 1 H), 4.03 (dd, *J* = 5.6, 12.2 Hz, 0.84 H), 3.98 (d, *J* = 12.2 Hz, 0.84 H), 3.77 (m, 1.16 H), 3.73 (m, 0.16 H), 3.3 (m, 0.32 H), 3.13 (m, 1.68 H), 3.03 (m, 1 H), 2.82 (m, 1 H), 2.73 (dd, *J* = 7.1, 16.1 Hz, 0.84 H), 2.51 (m, 0.32 H), 2.44 (d, *J* = 14.4 Hz, 0.84 H), 2.24 (m, 1 H), 2.15 (s, 2.52 H), 2.05 (m, 1.48 H), 1.90 (m, 1 H), 1.76 (m, 1 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present, but only the peaks of the major conformer are reported): δ 174.7, 173.0, 63.9, 62.6, 60.52, 60.48, 56.9, 46.3, 41.8, 39.7, 35.2, 34.4, 26.1, 23.1. HR-EI MS calcd for $C_{13}H_{21}N_3O_2S$ (M^+): 283.1354. Found: 283.1353.

<u>$AcHel_1CH_2NHAc$ (**8**).</u> The azide $AcHel_1CH_2N_3$ (0.005 g, 0.016 mmol) was dissolved at 23 °C under an argon atmosphere in 0.3 μL of thiolacetic acid. After 18h the solvent was removed and the residue purified by preparative HPLC (gradient: 10 to 100% CH_3CN over 40 min (remainder H_2O); retention time = 10.8 min) to yield 0.003 g (0.009 mmol, 58%). ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.68 (d, *J* = 10.0 Hz, 0.68 H), 4.25 (m, 2 H), 4.04 (dd, *J* = 5.9, 12.2 Hz, 0.68 H), 3.98 (d, *J* = 12.2 Hz, 0.68 H), 3.78 (m, 1.32 H), 3.72 (m, 0.32 H), 3.44 (dd, *J* = 8.1, 13.7 Hz, 0.32 H), 3.36 (m, 1 H), 3.25 (m, 1 H), 3.10 (dd, *J* = 4.9, 16.1 Hz), 2.81 (m, 1 H), 2.70 (dd, *J* = 7.3, 16.1 Hz, 0.68 H), 2.49 (m, 0.64 H), 2.42 (d, *J* = 14.4 Hz, 0.68 H), 2.22 (m, 1 H), 2.14 (s, 2.04 H), 2.06 (s, 0.96 H), 1.99 (s, 2.04 H), 1.98 (s, 0.96 H), 1.91 (m, 1 H), 1.77 (m, 2 H). HR-FAB MS calcd for $C_{15}H_{23}N_3O_3S$ (M^+): 325.1460. Found: 325.1461.

AcHel$_1$CH$_2$NMe$_2$ (**9**). To a solution of AcHel$_1$CH$_2$NH$_2$ (as the TFA salt; 0.024 g, 0.06 mmol) in 2 mL of 50 mM, pH 7.4 aqueous phosphate buffer were added 50 µL (0.6 mmol) formaldehyde followed by 0.038 g (0.6 mmol) $NaBH_3CN$. After 2h analytical HPLC showed the reaction to be complete (gradient: 9 to 99% CH_3CN over 30 min, remainder H_2O (0.1% TFA); product retention time = 6.16 min). The product was isolated by direct purification of the reaction mixture by preparative HPLC (gradient: 5 to 100% CH_3CN over 20 min (remainder H_2O, 0.1% TFA); retention time = 8.2 min) to yield 10 mg of AcHelCH$_2$NMe$_2$ as the TFA salt (0.024 mmol, 40%). ^{1}H-NMR (500 MHz, D_2O at pH 1, 2 conformations present): δ 4.71 (d, *J* = 10.3 Hz, 0.83 H), 4.56 (m, 0.83 H), 4.49 (m, 0.17 H), 4.23 (m, 1 H), 4.02 (dd, *J* = 5.4, 12.0 Hz, 0.83 H), 3.98 (d, *J* = 12.2 Hz, 0.83 H), 3.77 (m, 1.17 H), 3.73 (m, 0.17 H), 3.32 (m, 1.17 H), 3.21 (dd, *J* = 7.3, 13.2 Hz, 0.17 H), 3.15 (m, 1.66 H), 3.00 (s, 2.49 H), 2.97 (s, 1.02 H), 2.94 (s, 2.49 H), 2.82 (m, 1.0 H), 2.70 (dd, *J* = 7.6, 15.9 Hz, 0.83 H), 2.50 (m, 0.34 H), 2.43 (d, *J* = 14.7 Hz, 0.83 H), 2.14 (s, 2.49 H), 2.11 (m, 1 H), 2.06 (s, 0.51 H), 1.91 (dd, *J* = 5.6, 12.7 Hz, 1 H), 1.84 (m, 0.17 H), 1.76 (dd, *J* = 6.1, 13.2 Hz, 0.83 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present, but only the peaks from the major conformation are reported): δ 175.4, 172.3, 64.5, 63.9, 62.5, 60.1, 55.9, 46.9, 44.7, 41.8, 39.6, 35.3, 34.8, 35.6, 23.1. HR-FAB MS calcd for $C_{15}H_{25}N_3O_2S$ (M^+): 311.1667. Found: 311.1667.

AcHel$_1$CH$_2$NMe$_3^+$ (**10**). To a solution of 0.01 g (0.024 mmol) of AcHel$_1$CH$_2$NMe$_2$ (as the TFA salt) in 1 mL acetonitrile were added 16 µL of diisopropyl ethylamine (0.012 g, 0.092 mmol) and 6 µL of methyl iodide (0.014 g, 0.096 mmol). After 48 h, the reaction was complete, although it was difficult to judge by HPLC since the starting material and product had about the same retention times. The product was isolated by

direct purification of the reaction mixture by preparative HPLC (gradient: 5 to 100% CH_3CN over 40 min (remainder H_2O, 0.1% TFA); retention time = 8.7 min) to yield 6 mg of $AcHelCH_2NMe_3^+$ as the TFA salt (0.014 mmol, 57%). ^{1}H-NMR (500 MHz, D_2O at pH 1, 2 conformations present): δ 4.67 (d, *J* = 9.8 Hz, 0.78 H), 4.56 (t, 1 H), 4.20 (m, 1 H), 4.03 (dd, *J* = 5.8, 11.9 Hz, 0.78 H), 3.97 (d, *J* = 11.9 Hz, 0.78 H), 3.78 (m, 1.44 H), 3.44 (dd, *J* = 8.9, 13.1 Hz, 0.22 H), 3.31 (m, 2 H), 3.15 (m, 1.66 H), 3.23 (s, 1.98 H), 3.21 (s, 7.02 H), 3.16 (dd, *J* = 5.2, 16.2 Hz, 0.78 H), 2.81 (m, 1.0 H), 2.70 (dd, *J* = 7.6, 16.2 Hz, 0.78 H), 2.49 (m, 0.44 H), 2.42 (d, *J* = 14.0 Hz, 0.78 H), 2.17 (m, 2 H), 2.13 (s, 2.34 H), 1.99 (s, 0.66 H), 1.93 (m, 2 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present): δ 172.7, 171.4, 63.9, 63.1, 58.8, 57.5, 56.5, 55.5, 55.2, 54.96, 54.93, 54.90, 51.8, 45.8, 41.8, 40.5, 39.5, 38.1, 36.8, 34.4, 33.3, 28.1, 27.8, 23.3, 23.0 (some of the carbonyl resonances were too weak to be identified). HR-FAB MS calcd for $C_{16}H_{28}N_3O_2S$ (M^+): 326.1902. Found: 326.1902.

<u>$AcHel_1CONH_2$ (**12**).</u> To 0.14g of Knorr resin (0.84 mmol/g of Fmoc-protected amino groups, 0.12 mmol) in a peptide synthesis vessel (equipped with a frit at the bottom) were added 2 mL of 30% piperidine in DMF. After 15 min of agitation the solvent was filtered off (by applying a positive N_2 pressure at the top of the vessel), leaving the resin in the vessel. The resin was washed several times with DMF (by adding the DMF to the vessel, agitating, then filtering off the solvent again). A solution of 0.021 g (0.07 mmol) $AcHel_1CO_2H$ in 1 mL of DMF was prepared. To this solution were added 37 μL of diisopropylethyl amine (0.21 mmol), 0.022 g HOBt (0.14 mmol), and 0.074 g PyBOP (0.14 mmol). After 10 min, this solution was added to the resin in the vessel, agitated for 2h, and the solution was removed. The resin was washed as before, then dried under high

vacuum, then treated with 2 mL of 95% TFA / 5% thioanisole. After 2 more hours of agitation, the solvent was again filtered off, this time collecting the filtrate. The filtrate was concentrated, and the product was purified by preparative HPLC (gradient 5% to 100% CH_3CN over 60 min, remainder H_2O (0.1% TFA); retention time = 7.2 min) to yield 0.01g (0.04 mmol, 59%) of $AcHel_1CONH_2$. ^{1}H-NMR (500 MHz, D_2O, 2 conformations present): δ 4.68 (d, *J* = 9.5 Hz, 0.65 H), 4.54 (m, 1 H), 4.33 (m, 1 H), 4.06 (dd, *J* = 5.6, 12.0 Hz, 0.65 H), 4.01 (d, *J* = 12.2 Hz, 0.65 H), 3.81 (m, 1.35 H), 3.76 (m, 0.35 H), 3.32 (dd, *J* = 5.9, 15.4 Hz, 0.35 H), 3.11 (dd, *J* = 4.6, 16.1 Hz, 0.65 H), 2.86 (m, 1 H), 2.79 (dd, *J* = 7.3, 15.9 Hz, 0.65 H), 2.48 (m, 1.35 H), 2.24 (m, 2 H), 2.18 (s, 1.95 H), 2.16 (s, 1.05 H), 1.93 (m, 2 H).

<u>$AcHel_1CONHC_6H_4$-*o*-NH_2 (**13**).</u> To a solution of $AcHel_1CO_2H$ (0.01 g, 0.034 mmol) in 0.5 mL of DMF were added 6 μL of DIEA (0.034 mmol) and 0.018 g (0.034 mmol) of PyBOP. After 10 min, a solution of 0.011 g (0.06 mmol) of *o*-phenylene diamine dihydrochloride and 20 μL diisopropylethyl amine (0.11 mmol) in 0.5 mL of DMF was added. After 3h the solvent was removed from the reaction mixture and the product was purified by preparative HPLC (gradient: 5 to 100% CH_3CN over 30 min, remainder H_2O (0.1% TFA); retention time = 7.8 min) to yield 0.013 g (0.026 mmol, 76%) of $AcHel_1CONHC_6H_4$-*o*-NH_2 as the TFA salt. ^{1}H-NMR (500 MHz, CD_3CN, 2 conformations present but only the peaks of the major conformation are reported): δ 9.19 (s, 1 H), 7.43 (m, 4 H), 4.78 (d, *J* = 8.5 Hz, 1 H), 4.62 (d, *J* = 11.3 Hz, 1 H), 4.3 (m 1 H), 4.02 (dd, *J* = 5.8, 11.6 Hz, 1 H), 3.94 (d, *J* = 11.6 Hz, 1 H), 3.74 (m, 1 H), 3.00 (dd, *J* = 5.5, 16.2 Hz, 1 H), 2.89 (m, 2 H), 2.42 (d, *J* = 14.3 Hz, 1 H), 2.25 (m, 1 H), 2.15 (s, 3 H), 2.12 (m, 1 H), 1.93 (m, 1 H). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present but

only the peaks of the major conformation are reported): δ 173.7, 172.9, 160.5, 130.0, 128.9, 127.9, 125.2, 63.8, 63.5, 62.6, 59.9, 42.0, 38.8, 35.5, 27.0, 23.3. HR-FAB MS calcd for $C_{19}H_{24}N_4O_3S$ (M^+): 388.1569. Found: 388.1567.

AcHel$_1$benzim (**14**). The *o*-amino anilide AcHel$_1$CONHC$_6$H$_4$-*o*-NH$_2$ (as the TFA salt; 0.015 g, 0.03 mmol) was dissolved in 1 mL of AcOH and heated to reflux. After 2h the reaction was cooled to room temperature; analytical HPLC showed the reaction to be complete, but that a side product had accumulated (gradient 9 to 99% CH3CN over 30 min, remainder H2O (0.1% TFA); retention time = 15.0 for main product, 14.2 for contaminant). The solvent was removed from the reaction mixture and the product separated from the contaminant by preparative HPLC (gradient 0 to 100% CH_3CN over 40 min, remainder H_2O (0.1% TFA)). The purification had to be repeated to ensure that the product was pure, eventually yielding 0.004 g (0.008 mmol, 28%) of AcHel$_1$benzim as the TFA salt. ^{1}H-NMR (500 MHz, D_2O at pH 1, 2 conformations present but only the peaks of the major conformation are reported): δ 7.78 (m, 2 H), 7.60 (m, 2 H), 5.66 (d, *J* = 8.6 Hz, 1 H), 4.71 (d, *J* = 9.5 Hz, 1 H), 4.55 (m , 1 H), 4.09 (m, 2 H), 3.85 (m, 1 H), 3.23 (dd, *J* = 4.39, 15.9 Hz), 2.94 (dd, *J* = 7.1, 16.1 Hz, 1 H), 2.88 (ddd, *J* = 6.3, 10.5, 14.7 Hz, 1 H), 2.56 (m, 1 H), 2.48 (d, *J* = 14.4 Hz, 1 H), 2.39 (m, 1 H), 2.26 (s, 3 H), 2.20 (dd, *J* = 7.3, 14.7 Hz, 1 H), 2.09 (dd, *J* = 6.1, 13.4 Hz). ^{13}C-NMR (125 MHz, CD_3CN, 2 conformations present but only the peaks of the major conformation are reported): δ 174.4, 172.0, 155.6, 132.4, 127.6, 115.3, 63.5, 62.7, 60.6, 57.1, 42.0, 38.9, 35.5, 35.0, 29.2, 23.7. HR-FAB MS calcd for $C_{19}H_{22}N_4O_2S$ (M^+): 370.1463. Found: 370.1464.

Appendix 1. Data for Plots of f_s vs. 1/(t/c)

Table 16. f_s and 1/(t./c) data for 5'-acyl compounds for the plot in section 1.4.2b, figure 12.

$AcHel_1$ derivative	t/c ratio	1/(t/c)	$\delta 9b_{obs}$	f_s
$AcHel_1CONH_2$	1.85	0.54	3.11 ppm	0.53
$AcHel_1CONHMe$	2.01	0.50	3.07 ppm	0.44
$AcHel_1CONHOMe$	1.29	0.78	3.14 ppm	0.60
$AcHel_1CONHCH_2CF_3$	1.20	0.83	3.17 ppm	0.67
$AcHel_1COAlaOH$	0.81	1.23	3.23 ppm	0.80
$AcHel_1COGlyOH$	1.30	0.77	3.13 ppm	0.58

Table 17. f_s and 1/(t./c) data for 5'-sp^3 compounds for the plot in section 1.4.2b, figure 13.

$AcHel_1$ derivative	t/c ratio	1/(t/c)	$\delta 9b_{obs}$	f_s
$AcHel_1CH_2OAc$	1.35	0.74	3.17 ppm	0.67
$AcHel_1CH_2NMe_2$	1.40	0.71	3.15 ppm	0.62
$AcHel_1CH_2OMe$	1.60	0.63	3.15 ppm	0.62
$AcHel_1CH_2OH$	1.81	0.55	3.17 ppm	0.67
$AcHel_1CH_2SH$	1.82	0.55	3.13 ppm	0.58
$AcHel_1CH_2NHAc$	2.14	0.47	3.10 ppm	0.51
$AcHel_1CH_2NH2$	2.21	0.45	3.11 ppm	0.53

Appendix 2 Further Studies of Hydrogen Bonding Using Trifluoroethanol-Water Mixtures

It has been known for over 30 years that peptides in aqueous solution usually become more helical upon addition of trifluoroethanol (TFE).[101] The mechanism by which this effect occurs has been debated, with some authors favoring a direct mechanism in which TFE interacts with and stabilizes helical states[102,103] and others favoring an indirect mechanism in which TFE causes a change in the structure of liquid water that destabilizes the non-helical states.[71,104,105,106] Recently, definitive evidence has been presented in favor of the latter mechanism.[71] Using the t/c ratios of derivatives of $AcHel_1$ (much like those in this present study) and the rate of trans ⇄ cis interconversion of the acetyl group in acetyl N-methyl prolinamide, it was demonstrated that TFE acts by destabilizing the hydrogen bonds formed by the solvent to amide carbonyls. Since the stability of hydrogen bonds not involving solvent molecules is unaffected, the net result is a shift in the hydrogen bonding equilibrium to favor interactions between non-solvent donors and acceptors such as those that occur between backbone amides in peptide helices.

Given this indirect mechanism, one would expect that, just as the amide-amide hydrogen bonds in peptide helices are favored by addition of TFE to water, the hydrogen bonds of the $AcHel_1$ derivatives in this study should be favored by TFE. That is, K_{hb}

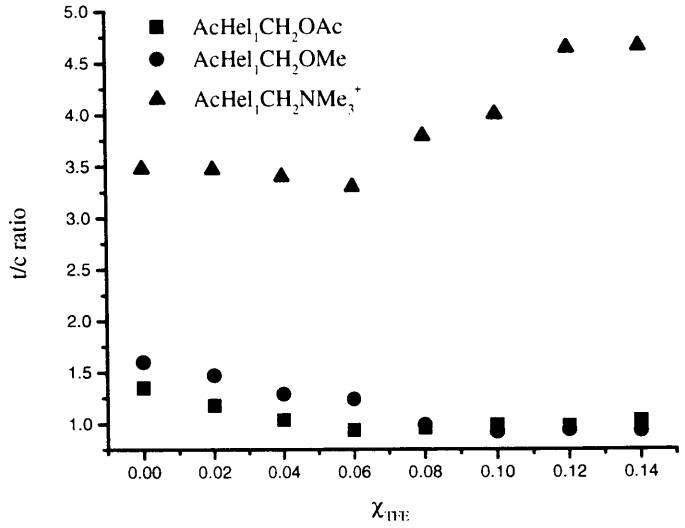
---

[101] Goodman, M.; Listowsky, I.; Masuda, Y.; Boardman, F. *Biopolymers* **1963**, *1*, 33.
[102] Jasanoff, A.; Fersht, A. R. *Biochemistry* **1994**, *33*, 2129.
[103] Rajan, R.; Balaram, P. *Int. J. Peptide Protein Res.* **1996**, *48*, 328.
[104] Conio, G.; Patrone, E.; Brighetti, S. *J. Biol. Chem.* **1970**, *245*, 3335.
[105] Storrs, R. W.; Truckses, D.; Wemmer, D. E. *Bioploymers* **1992**, *32*, 1695.
[106] Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.

values should increase regularly with the mole fraction of TFE (χ_{TFE}) in TFE-water mixtures. This effect creates an opportunity to better differentiate the strengths of the hydrogen bond donors in this study. The data from section 1.5 obtained in D_2O are flattened- that is, the hydrogen bonding abilities of the weaker donors, for example, the thiol and the alcohol, appear similar because they cause such small changes in the t/c ratios of the $AcHel_1$ derivatives in which they reside. Data obtained in TFE-D_2O mixtures should increase all of the K_{hb}s and permit distinctions to be made.

In order to know K_{hb} values as a function of χ_{TFE}, one must first know the t/c ratios of the reference compounds as a function of χ_{TFE}. For the neutral 5'-acyl $AcHel_1$ derivatives, $(t/c)_{ref}$ has been taken to be constant.[71] The responses of the neutral 5'-sp^3 and 5'-charged $AcHel_1$ reference compounds' t/c ratios (those of $AcHel_1CH_2OAc$, $AcHel_1CH_2OMe$, and $AcHel_1CH_2NMe_3^+$) to increasing concentrations of TFE are shown in figure 17. (The t/c ratios of all compounds as functions of χ_{TFE} are shown in table 20 below). It should be noted that deuterated TFE (TFE-d_3) is used in the titration so that its signal does not overwhelm the NMR spectrum. Also, the NMR spectra at varying TFE concentrations from which the t/c ratios were calculated for several compounds ($AcHel_1CH_2OH$,[73] $AcHel_1CH_2NHAc$,[73] $AcHel_1CONH_2$,[73] and $AcHel_1CH_2SH$[74]) were measured prior to the work in this thesis. The t/c ratio vs. χ_{TFE} curves for the two neutral reference $AcHel_1$ derivatives initially decline from χ_{TFE} = 0 to 0.10 then level off, showing that the t state is generally disfavored by increasing χ_{TFE}, possibly because as χ_{TFE} increases the dielectric of the medium decreases.

Figure 17. Plot of t/c ratio vs χ_{TFE} for reference $AcHel_1$ derivatives.

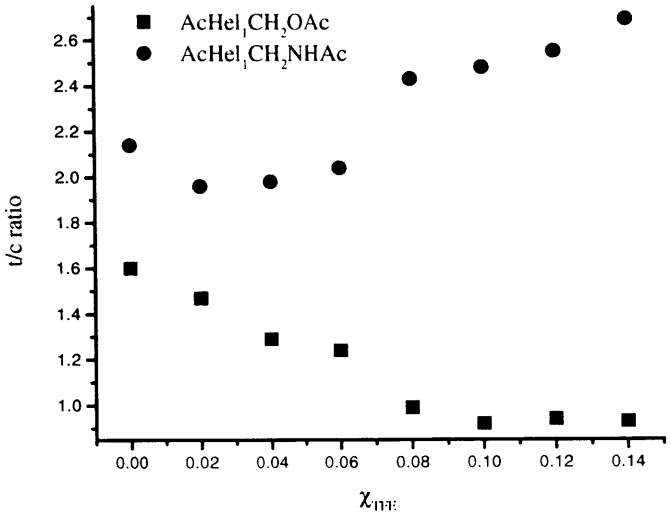


In contrast, the t/c ratio for $AcHel_1CH_2NMe_3^+$ decreases until χ_{TFE} = 0.06 and then rapidly increases until χ_{TFE} = 0.12 at which point it levels off. This can be interpreted as a superposition of effects: first, the destabilization of the t state apparent in the neutral reference compounds and second, an apparent increase in the magnitude of the NMe_3^+-acetamide charge-dipole interaction as χ_{TFE} increases. The latter effect could also be due to the decreasing dielectric of the medium as χ_{TFE} increases or it could be that the desolvation of the acetamide carbonyl that favors hydrogen bonding also favors less specific charge-dipole interactions.

The t/c ratios for $AcHel_1$ derivatives with hydrogen bond donors, like the t/c ratio of $AcHel_1CH_2NMe_3^+$, decrease until χ_{TFE} ~ 0.04 to 0.08, then increase, and finally begin to level off after χ_{TFE} = 0.12 (see table 20). This indicates a similar competition between two effects as was observed for $AcHel_1CH_2NMe_3^+$, but for the neutral hydrogen bond donors, the force that stabilizes the t state cannot be an increase in the strength of a

charge-dipole interaction. It has to be an increase in the strength of the intramolecular hydrogen bond that occurs for all of the neutral donors, including the thiol. For the charged donors, the force that stabilizes the t state with increasing χ_{TFE} could be either an increase in K_{hb} or the charge-dipole interaction or both. The change in t/c ratio with χ_{TFE} of $AcHel_1CH_2NHAc$ is compared to that of its reference compound, $AcHel_1CH_2OAc$, in figure 18 to illustrate the difference in behavior.

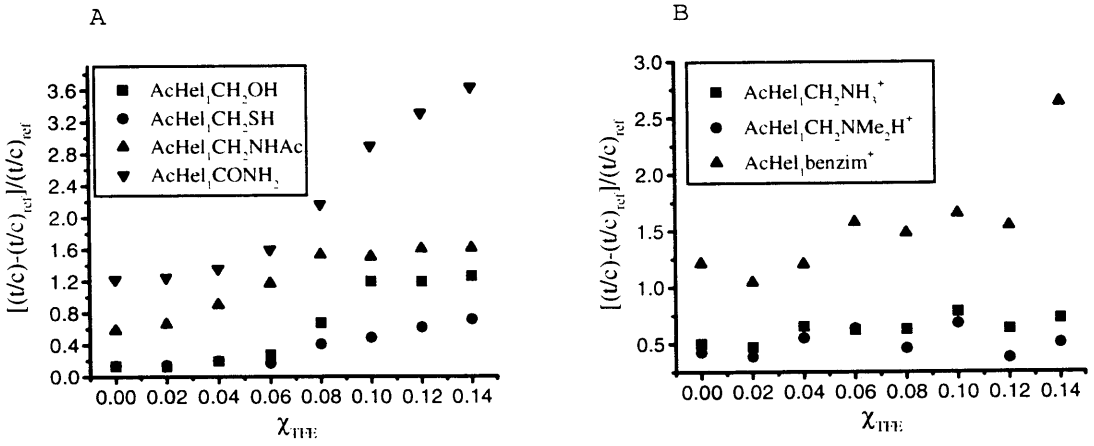
Figure 18. Response to increasing TFE concentration of the t/c ratios of an $AcHel_1$ derivative capable of hydrogen bonding ($AcHel_1CH_2NHAc$) and an $AcHel_1$ derivative not capable of hydrogen bonding ($AcHel_1CH_2OAc$).



With the t/c ratios of both the reference compounds and the hydrogen bond donor compounds known as functions of χ_{TFE}, an idea for how K_{hb} values change as functions of χ_{TFE} can be obtained from plots of $[(t/c) - (t/c)_{ref}]/(t/c)_{ref}$ vs. χ_{TFE}. These are shown in figures 19A and 19B for the neutral and the charged donors respectively. (Note that the free amines $AcHel_1CH_2NH_2$ and $AcHel_1CH_2NMe_2$ and the benzimidazole $AcHel_1$benzim were not titrated with TFE. At the pH's used to ensure that these compounds were

neutralized, TFE (pK_a = 12.4) would be substantially ionized and the TFE effect would likely be much altered.) The values of $[(t/c) - (t/c)_{ref}]/(t/c)_{ref}$ show that the hydrogen bond strengths increase roughly monotonically with χ_{TFE}, as expected for the neutral donors. For the charged donors, however, they have only a slight upward trend (see figure 19B). The only data point that could indicate a TFE effect of the same magnitude as those observed for the neutral donors is that for $AcHel_1benzim^+$ at χ_{TFE} = 0.14. However, at χ_{TFE} = 0.14 the t/c ratio of benzimidazole is very large (>10; see table 20), and it is difficult to measure t/c ratios accurately when they are this large.[71] Given this and how far out of line this datum is with the rest of the data, it can safely be ignored.

Figure 19A. Plots of $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ as a function of χ_{TFE} for neutral donors. **Figure 19B.** The same for charged donors.



Assuming that K_{se}, γ, and f_o are constant with respect to χ_{TFE} for all of the hydrogen bond donors, limits on K_{hb} at any χ_{TFE} can be calculated using equation 1-4. The factors $[(\gamma + K_{se})/(1 + K_{se})]$ and f_o in these relations are listed for the $AcHel_1$ derivatives in tables

10 and 12 respectively. The factors $[(t/c) - (t/c)_{ref}/(t/c)_{ref}]$ at $\chi_{TFE} = 0.10$ are listed in the second column of table 18 below, along with the resulting limits on K_{hb}.

Table 18. Values of $[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$ for $AcHel_1$ derivatives in D_2O with 10 mol % TFE and the consequent K_{hb} values computed using the limits on $[(\gamma + K_{se})/(1 + K_{se})]$ and f_o in tables 10 and 12 and equation 1-4. All data are for 25 °C.

$AcHel_1$ derivative	$[(t/c)-(t/c)_{ref}]/(t/c)_{ref}$	K_{hb}	
		min	max
$AcHel_1CONH_2$	2.89	38	100
$AcHel_1CH_2OH$	1.20	6.9	18
$AcHel_1CH_2SH$	0.49	2.8	7.3
$AcHel_1CH_2NHAc$	1.51	8.7	22
$AcHel_1CH_2NH_3^+$	0.79	-	3.8
$AcHel_1CH_2NMe_2H^+$	0.68	-	6.4
$AcHel_1benzim^+$	1.65	-	4.2

The addition of TFE to water has the expected effect of making intrinsic hydrogen bonding equilibrium constants larger for the neutral donors. K_{hb} for the two amides and even for the thiol are larger by factors of around 3 in 10 mol % TFE, while K_{hb} for the alcohol is more larger by a factor of 5. This demonstrates how sensitive to medium hydrogen bonding can be, with low mole fractions of additives being able to effect substantial changes in the energetics of the process. Note also that with the larger hydrogen bonding effects observed at $\chi_{TFE} = 0.1$, the superiority of alcohols over thiols as hydrogen bond donors becomes evident.

In contrast to what is observed with neutral donors, the intrinsic hydrogen bonding energies do not change very much for the charged 5' donors. However, just as the small upper bounds on ΔG_{hb} in table 13 do not indicate a weak overall interaction between these donors and the acetamide carbonyl, this does not indicate a lack of response to TFE. The t/c ratios of $AcHel_1CH_2NH_3^+$, $AcHel_1CH_2NMe_2H^+$, and $AcHel_1benzim^+$ all increase with increasing χ_{TFE} (see table 20), so the overall interactions between the charged 5' donors and the acetamide carbonyls in these compounds must therefore intensify with increasing χ_{TFE}. However, the t/c ratio of $AcHel_1CH_2NMe_3^+$ also increases with increasing χ_{TFE}, so it must be that hydrogen bonding adds little to the 5'-charged $AcHel_1$ derivatives responses to TFE.

TFE titrations. The $AcHel_1$ derivative of interest was dissolved in 0.7 mL of D_2O and the volumes of TFE-d_3 listed in the second column of table 19 were added serially to the sample to give the required mole fractions of TFE. NMR spectra were measured and t/c ratios calculated at each concentration of TFE.

Table 19. Amounts of TFE-d_3 needed to achieve a given mole fraction of TFE-d_3 in D_2O when starting with 0.7 mL of D_2O.

χ_{TFE}	volume of TFE-d_3 to add	total volume of TFE-d_3
0	0	0
0.02	56 µL	56 µL
0.04	58.4 µL	114.4 µL
0.06	60.8 µL	175.2 µL
0.08	63.5 µL	238.7 µL
0.10	66.3 µL	305.0 µL
0.12	69.4 µL	374.4 µL
0.14	72.5 µL	446.9 µL

The t/c ratio vs. χ_{TFE} data are listed in table 20. As mentioned, some of the TFE titration experiments were performed prior to the work in this thesis and unfortunately data were not always collected at the same set of TFE concentrations. For the two cases where this occurs ($AcHel_1CH_2OH$ and $AcHel_1CH_2SH$), the data are linearly interpolated so that t/c ratio data are available for all of the compounds at χ_{TFE} = 0, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, and 0.14. The interpolated data points in table 13 are indicated with an asterisk.

Table 20. $AcHel_1$ derivatives' t/c ratios vs. χ_{TFE} (the first column indicates the 5 position substituent).

χ_{TFE}	0	0.02	0.04	0.05	0.06	0.075	0.08	0.10	0.12	0.125	0.14	0.15
CH_2OMe	1.60 ±0.04	1.47 ±0.04	1.29 ±0.05	-	1.24 ±0.08	-	0.99 ±0.10	0.92 ±0.08	0.94 ±0.08	-	0.93 ±0.08	-
CH_2OAc	1.35 ±0.02	1.18 ±0.02	1.04 ±0.03		0.94 ±0.08		0.96 ±0.02	0.99 ±0.01	0.98 ±0.05		1.03 ±0.01	
CH_2OH	1.81 ±0.03	1.65 ±0.04	1.54 ±0.05		1.58 ±0.05		1.66 ±0.08	2.02 ±0.08	2.06*		2.10*	2.12 ±0.06
CH_2SH	1.82 ±0.07	1.68*	1.55*	1.48 ±0.03	1.45*	1.40 ±0.05	1.39*	1.37 ±0.07	1.52*	1.56 ±0.06	1.60*	1.62 ±0.07
CH_2NHAc	2.14 ±0.06	1.96 ±0.09	1.98 ±0.05		2.04 ±0.09		2.43 ±0.09	2.48 ±0.06	2.55 ±0.12		2.69 ±0.09	
$CONH_2$	1.85 ±0.04	1.87 ±0.04	1.96 ±0.05		2.16 ±0.06		2.63 ±0.05	3.24 ±0.07	3.58 ±0.11		3.85 ±0.07	
$CH_2NMe_2H^+$	4.96 ±0.28	4.81 ±0.51	5.28 ±0.40		5.41 ±0.75		5.54 ±0.53	6.73 ±0.99	6.39 ±0.64		7.02 ±0.70	
$CH_2NH_3^+$	5.23 ±0.21	5.11 ±0.24	5.63 ±0.24		5.36 ±0.23		6.17 ±0.24	7.15 ±0.62	7.59 ±0.89		8.02 ±0.95	
$benzim^+$	7.7	7.1	7.5		8.5		9.4	10.6	11.8		16.9	
$CH_2NMe_3^+$	3.48 ±0.10	3.47 ±0.18	3.40 ±0.22		3.30 ±0.23		3.79 ±0.25	4.00 ±0.25	4.64 ±0.29		4.65 ±0.30	

Chapter 2. Introduction to Peptide Helicity and the Zimm-Bragg Model for the Helix-Coil Transition

2.1 Introduction

The remainder of this thesis addresses problems in peptide helicity. The purpose of this chapter is to provide the background on the structure of helices and the models used to describe peptide helicity that will be necessary to understand the results of the subsequent chapters.

2.2 Types and Geometries of Peptide Helices

As experimental evidence that polypeptides and proteins adopted regular and repeating structures accumulated, so too did proposals for what the geometry of these structures might be.[1,2,3,4,5,6] When the first protein crystal structure became available,[7] that of myoglobin, it was found that one of the most important of these structures were helices. As the database of protein crystal structures grew, it became apparent that only two forms of helices occurred with notable frequency in proteins[8]: the α helix[4] and the

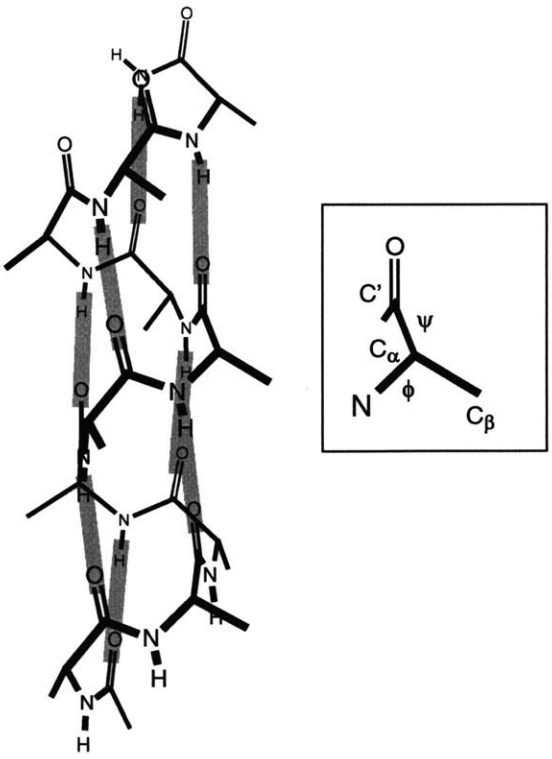
---

[1] Astbury, W. T.; Bell, F. O. *Nature* **1941**, *147*, 696.
[2] Huggins, M. L. *Chem. Rev.* **1943**, *32*, 195.
[3] Bragg, L.; Kendrew, J. C.; Perutz, M. F. *Proc. Roy. Soc.* **1950**, *A203*, 321.
[4] Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205.
[5] Low, B. W.; Baybutt, R. B. *J. Am. Chem. Soc.* **1952**, *74*, 5806.
[6] Donohue, J. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 470.
[7] Kendrew, J. C.; Watson, H. C.; Strandberg, B. E.; Dickerson, R. E.; Phillips, D. C.; Shore, V. C. *Nature* **1961**, *190*, 666.
[8] Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988**, *201*, 601.

less common 3_{10} helix.[6] Over one third of all amino acids in proteins occur in one of these two conformations.[8]

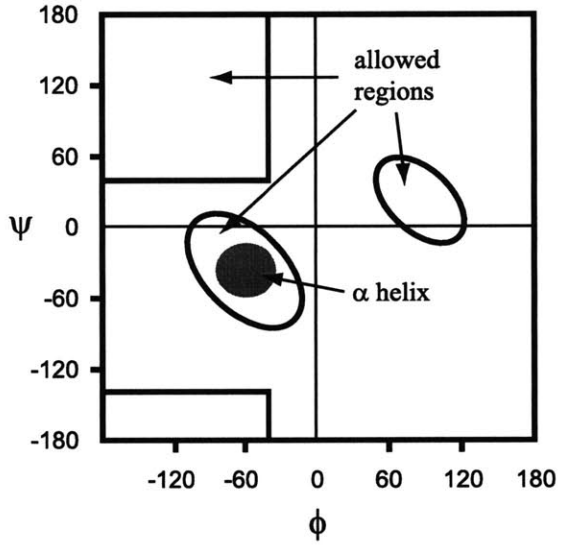
The α helix is pictured in figure 1. Its most distinguishing feature is the string of i to i+4 amide CO to amide NH hydrogen bonds (where i represents the position of the residue in the sequence) that form along the helix. All residues except the three most N-terminal NH groups and the three most N-terminal CO groups are hydrogen bonded in this way. In order for this hydrogen bonding pattern to occur, the backbone torsions φ (corresponding to the C'-N-C_α-C' dihedral angle) and ψ (corresponding to the N-C_α-C'-N dihedral angle) of the intervening residues must be approximately −57° and −47° respectively.

Figure 1. A peptide ($AcAla_{10}NH_2$) in an α helix.



These are the angles for an ideal helix. In reality, they can vary somewhat while still maintaining the α form[8] as illustrated in figure 2, where the ϕ,ψ torsions that have been observed in protein α-helices are shown as a region, rather than a point, in a Ramachandran diagram.[9] Still, the α helix region of this plot is small compared to the total of the sterically allowed regions, which are also shown in figure 2.

Figure 2. Ramachandran diagram showing the ϕ and ψ that have been observed for residues in protein crystal structures (the allowed regions). The region corresponding to protein α helices is shaded gray. Adapted from Swindells, et. al. (ref. 9)

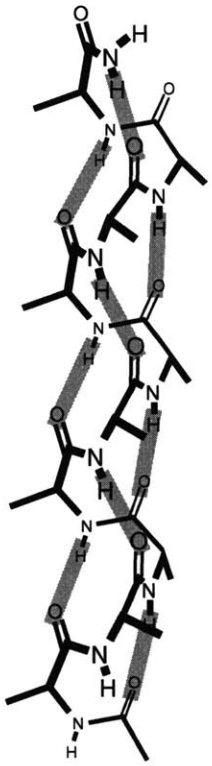


There are approximately 3.7 residues per turn of the α helix, and its length increases by 1.5 Å per residue. The β carbons of the amino acids in helices project outward from the helix barrel, and are inclined toward the N-terminus of the helix. For residues other than alanine and glycine (which do not have χ_1 torsions), the χ_1 torsion of the side chain

[9] Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nature Struct. Biol.* **1995**, *2*, 596.

(corresponding to the C'-C_α-C_β-C_γ dihedral angle) is limited to two of its three rotamers. The g^- rotamer (dihedral angle = 60°) is disfavored because the C_γ of the side chain clashes with the carbonyl oxygen of residue i-3 in this conformation,[10] leaving only the t (dihedral angle = 180°) and g^+ (dihedral angle = -60°) rotamers substantially populated. This restriction is even more severe for β branched residues, for which only one rotamer of the χ_1 torsion is significant. Finally, it has been postulated that a dipole moment is associated with peptide helices, usually attributed to all of the carbonyls being pointed in the same direction.[11] The positive pole of the dipole is at the N-terminus and the negative pole is at the C-terminus.

Figure 3. A peptide ($AcAla_{10}NH_2$) in a 3_{10} helix.



[10] McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295.

The 3_{10} helix is pictured in figure 3. It is pitched somewhat higher and wound somewhat tighter than the α helix. The 3_{10} helix has i to i+3 instead of i to i+4 hydrogen bonds, requiring φ and ψ angles of around -49° and -26° in the ideal case but, as with α helices, these angles can vary somewhat.[8] The 3_{10} helix has 3.0 residues per turn and a length of 1.8 Å per residue. The φ and ψ torsions in the 3_{10} helix are somewhat strained suggesting that the 3_{10} helical form should be less stable than the α form.[6] In fact, residues in proteins are observed in the α helical conformation almost ten times more frequently than in the 3_{10} conformation.[8] Also, when short, monomeric peptides form helices in solution they are almost always assumed to be in the α form[12] (although there is evidence for at least partial 3_{10} character in these helices[13]). The remainder of this chapter will focus on the characteristics of α helices.

2.3 Helix Formation by Short Peptides in Solution.

From x-ray crystallographic evidence, the helices that occur in proteins are known to be structurally well defined.[8] They have distinct starting and ending points, and they are, to a very good approximation, the same in every protein molecule in a sample. Such is not the case for the helices that occur in free peptides in solution. In free peptides, there is no all or none transition to a unique state. Many states of varying and intermediate helicity are populated in addition to the fully helical and fully random coil states. The

---

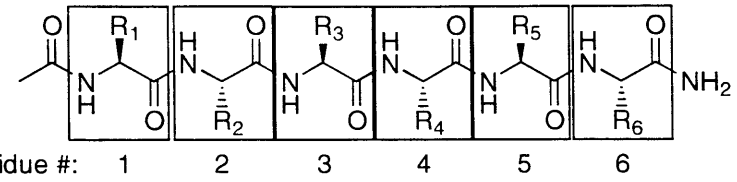
[11] Hol, W. G. J. *Prog. Biophys. Molec. Biol.* **1985**, *45*, 149.
[12] Chakrabartty, A.; Baldwin, R. L. *Adv. Prot. Chem.* **1995**, *46*, 141.
[13] Millhauser, G. L.; Stenland, C. J.; Hanson, P.; Bolin, K. A.; van de Ven, F. J. M. *J. Mol. Biol.* **1997**, *267*, 963.

first experience with non-protein helix formation came in the 1950's, when it was observed that long polymers of amino acids underwent a transition from a random coil state (in which there is no regular hydrogen bonding and the φ and ψ torsions fluctuate among the sterically allowed values) to a helical state.[14,15,16,17,18,19] For polypeptides that were long enough, this transition occurred over very narrow ranges of temperature, solvent composition, or pH. Because this behavior called phase transitions to mind, it was natural that the several mathematical models that appeared to describe this phenomenon[20,21,22,23,24,25] were adapted from models for other order-disorder transitions. In particular, the two most successful descriptions of the heilx-coil transition, one due to the Zimm and Bragg,[23] the other to Lifson and Roig,[24] were based on work that was originally intended to describe magnetization.[26] Although the Lifson-Roig model is used more widely, it is easy to translate the results from one framework to the other[27] and the results that will be presented in the following chapters are more easily understood in the context of the Zimm-Bragg model. Therefore, the Zimm-Bragg model is developed in this section.

---

[14] Blout, E. R.; Idelson, M. *J. Am. Chem. Soc.* **1956**, *78*, 497.
[15] Doty, P.; Yang, J. T. *J. Am. Chem. Soc.* **1956**, *78*, 498.
[16] Doty, P.; Bradbury, J. H.; Holtzer, A. M. *J. Am. Chem. Soc.* **1956**, *78*, 947.
[17] Blout, E. R.; Asadourian, A. *J. Am. Chem. Soc.* **1956**, *78*, 955.
[18] Doty, P.; Wada, A.; Yang, J. T.; Blout, E. R. *J. Polymer Sci.* **1957**, *23*, 851.
[19] Doty, P.; Imahori, K.; Klemperer, E. *Proc. Natl. Acad. Sci USA* **1958**, *44*, 424.
[20] Schellman, *J. Phys. Chem.* **1958**, *62*, 1485.
[21] Peller, L. *J. Phys. Chem.* **1959**, *63*, 1194 and 1199.
[22] Hill, T. *J. Chem. Phys.* **1959**, *30*, 383.
[23] Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
[24] Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.
[25] Gibbs, J. H.; DiMarzio, E. A. *J. Chem. Phys.* **1959**, *30*, 271.
[26] Ising, E.; *Z. Physik* **1925**, *31*, 253.
[27] Qian, H.; Schellman, J. A. *J. Phys. Chem.* **1992**, *96*, 3987.

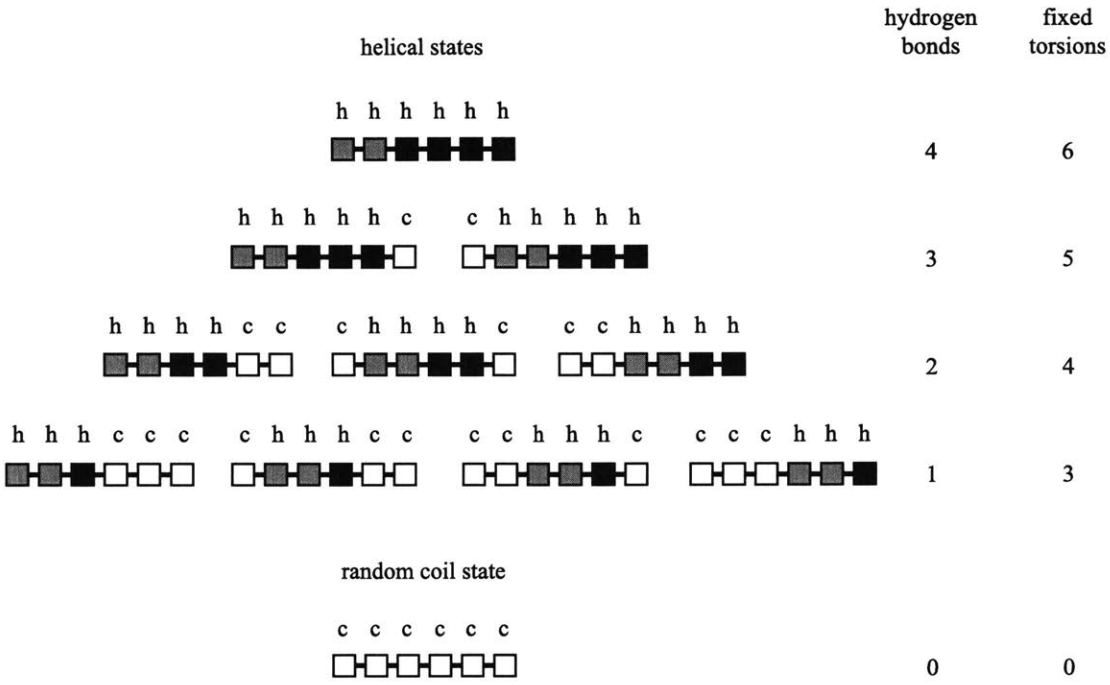
Figure 4. The general peptide AcX1X2X3X4X5X6-NH_2 divided into residues. Residue X1 has side chain R_1, residue X2 has side chain R_2, etc.

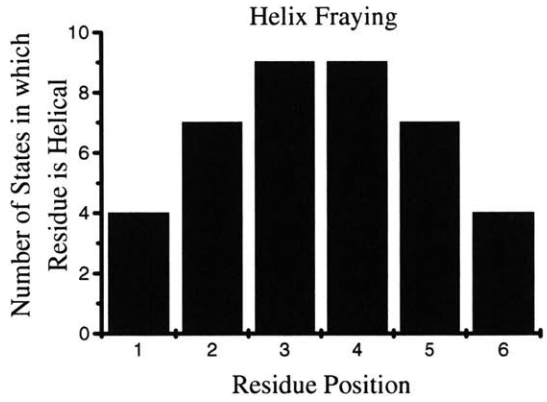


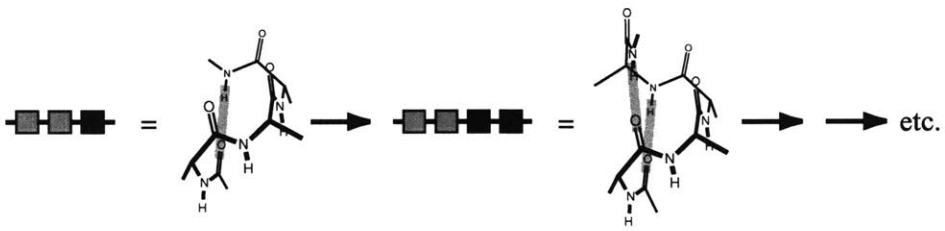
In order to describe a peptide's helix-coil equilibrium completely, all of its possible states have to be enumerated and the mole fraction of each state has to be assigned. The means by which the Zimm-Bragg model accomplishes this is developed below using a generalized acetylated and amidated six residue peptide, AcX1X2X3X4X5X6-NH_2 (shown in figure 4), as a heuristic example. According to the Zimm-Bragg model, the residues (which are delimited in figure 4 by boxes and numbered from N to C-terminus) can each be in one of two conformations, h (helical) or c (coil). A residue is said to be in the h conformation when it is part of a string of at least three consecutive residues with ϕ and ψ torsions in the helical range. A residue is said to be in the c conformation when either its ϕ and ψ torsions are not in the helical range, or when they are in the helical range but not part of a string of at least three like residues. The state of a peptide can thus be represented as a linear sequence of h's and c's, one for each residue, and every sequence of symbols that is consistent with the naming rules above represents a viable state for the peptide. Since many sequences of symbols can be constructed that are consistent with the naming rules, there must be many states that have to be considered in order to completely describe the helix-coil equilibrium for free peptides in solution. The states that make up the helix-coil equilibrium for the six residue peptide of figure 4 are

shown in figure 5 on the following page. One important point is immediately apparent from the figure. There are more states in which the central residues are in the h conformation than states in which the terminal residues are in the h conformation. The likelihood of a residue being helical is therefore maximal for residues in the center of a peptide and diminishes toward either terminus. This phenomenon, called helix fraying, is illustrated by the bar graph toward the bottom of figure 5 that plots the number of states in which a given residue is helical against the residue's position.

Figure 5. The states that make up the helix-coil equilibrium of a six residue peptide (such as AcX1X2X3X4X5X6-NH_2) in cartoon form, and a bar graph showing the dependence of the number of states in which a residue is helical on the residue's position. The structural equivalents of the cartoons are shown at the bottom. See text on following page for an explanation of the shading.







In figure 5, the states of the peptide are represented by strings of boxes as well as strings of symbols. The boxes are not filled if they correspond to a residue that is in the c conformation, but if they correspond to a residue in the h conformation, they can be filled in two different ways according to their position in the helix. If they are either the first or second h, and therefore have fixed ϕ and ψ torsions but do not directly precede a hydrogen bonded NH, they are filled with gray. Otherwise, they are filled with black. The structural translation of this coding is provided underneath the bar graph in figure 5. The formulation of the helical states in terms of the boxes is meant to facilitate understanding of how the energy of each states is assigned. One of the fundamental assumptions of the Zimm-Bragg model is that the energy of a given state is the sum of independent contributions by each residue, or equivalently, that the equilibrium constant of a given state relative to the random coil state (the arbitrary reference state) is the product of independent contributions by each residue. If a residue is in the c conformation (unfilled box), it contributes a factor of 1 to the state's equilibrium constant. If a residue is in the h conformation and directly precedes a hydrogen bonded NH (black filled box), it contributes s_{res} to the state's equilibrium constant, a factor that depends on the identity of the residue. The pair of h residues that do not directly precede a hydrogen bonded NH (gray filled box) together contribute σ to the state's equilibrium constant. For the purposes of this thesis, and consistent with convention, σ will be assumed to be independent of the identity of the residues.

The value of the s_{res} factor reflects a balance between the favorable intrinsic energy of formation of an amide NH - amide CO hydrogen bond (up to -2.2 kcal/mol, according to the results of the chapter 1) and the unfavorable entropy of fixing the residue's ϕ and ψ

torsions, so that s_{res} is usually near 1 (see table 2 below). The s value is also known as a helix propensity. In contrast to the situation for s_{res}, there are no opposing forces contributing to σ. There is only the unfavorable entropy due to fixing two sets of ϕ and ψ torsions, so that σ is usually on the order of 10^{-3}. The σ parameter is also known as an initiation constant, since it is contributed by the residues at the site where the helix begins.[28] Since σ is so small compared to s, initiation (as opposed to propagation) is the difficult step in helix formation.

With the guidelines above, an equilibrium constant for each of the states in figure 5 can be written. These will always consist of a product of σ and the s values for all the h residues after the second. For a helix that starts at the i^{th} residue of a peptide and ends at the $i+n^{th}$,

$$\text{equilibrium constant} = \sigma \prod_{j=i+2}^{n} s_j \qquad \textbf{2 - 1}$$

The equilibrium constants calculated according to this formula for all the states of AcX1X2X3X4X5X6-NH_2 are listed in table 1.

[28] The Zimm-Bragg parameters can be translated into the corresponding parameters of the Lifson-Roig model. In the Lifson-Roig model, the initiation parameter is v^2, and the propensity is w. According to Schellman and Qian (ref. 27), $s = w/(1 + v)$ and $\sigma = v^2/(1+v^4)$. The first and last residues in a helix are assigned a v weight (as opposed to the first two), while all other helical residues are assigned a w weight.

Table 1. Equilibrium constants for the states shown in figure 5 of the peptide AcX1X2X3X4X5X6-NH_2.

	State	Equilibrium Constant
4 h-bonds, 6 fixed ϕ and ψ torsions	hhhhhh	$\sigma s_{X3} s_{X4} s_{X5} s_{X6}$
3 h-bonds, 5 fixed ϕ and ψ torsions	hhhhhc	$\sigma s_{X3} s_{X4} s_{X5}$
	chhhhh	$\sigma s_{X4} s_{X5} s_{X6}$
2 h-bonds, 4 fixed ϕ and ψ torsions	hhhhcc	$\sigma s_{X3} s_{X4}$
	chhhhc	$\sigma s_{X4} s_{X5}$
	cchhhh	$\sigma s_{X5} s_{X6}$
1 h-bond, 3 fixed ϕ and ψ torsions	hhhccc	σs_{X3}
	chhhcc	σs_{X4}
	cchhhc	σs_{X5}
	ccchhh	σs_{X6}
	cccccc	1

(Note that the first two residues never contribute an s to the equilibrium constants. This is because they never immediately precede a hydrogen bonded NH in any of the helical states, so the fixing of their ϕ and ψ angles is never rewarded directly by hydrogen bond formation.) The equilibrium constants in the table can be used to determine the mole fraction of each state, and thus to obtain a complete description of the helix-coil equilibrium for AcX1X2X3X4X5X6-NH_2. Let Z be the sum of all of the equilibrium constants. The mole fraction of a given state is then

$$\chi_{state} = \frac{\text{state's equilibrium constant}}{Z} \qquad \boxed{2-2}$$

These mole fractions can be used to calculate other quantities of interest. The fractional helicity of a peptide is the fraction of residues in the total peptide sample in the h conformation. This is equal to the sum of a set of terms, each of which is the mole fraction of a state multiplied by the fraction of residues that are helical in that state:

$$\text{fractional helicity} = \sum_{\text{all states}} \chi_{state} \left(\frac{\#\text{ of helical residues in the state}}{\text{total }\#\text{ of residues in the peptide}} \right) \qquad \boxed{2-3}$$

The fractional hydrogen bonding of a given residue i, $f_{hb,i}$, is the mole fraction of the states in which this residue is hydrogen bonded. Since residue i is hydrogen bonded when residue i-1 is at least the third in a sequence of h residues, this is equal to the sum of the mole fractions of all of the individual states in which residue i-1 contributes an s value to the equilibrium constant:

$$f_{hb,i} = \frac{\text{sum of eq. constants in which i-1 contributes an s value}}{\text{sum of all equilibrium constants}} = \sum_{\substack{\text{j = all states in which} \\ \text{i is hydrogen bonded}}} \chi_j \qquad \boxed{2-4}$$

The fractional helicity and $f_{hb,i}$ are particularly useful to be able to calculate from the model because they are believed to correlate with parameters measurable by circular dichroism spectroscopy and the amide hydrogen exchange technique respectively (see section 2.5).

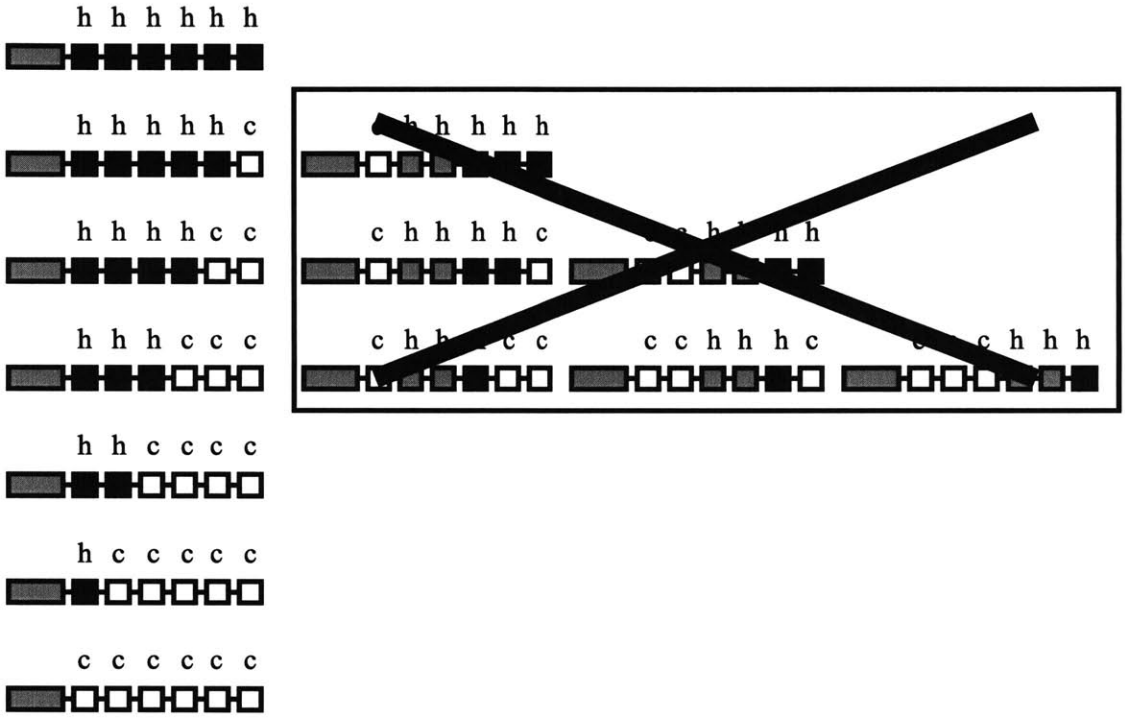
The example that has been used thus far to elucidate the Zimm-Bragg model for peptide helicity is a simple one. For a six residue peptide, all of the possible states can easily be enumerated and their equilibrium constants assigned. However, the complexity

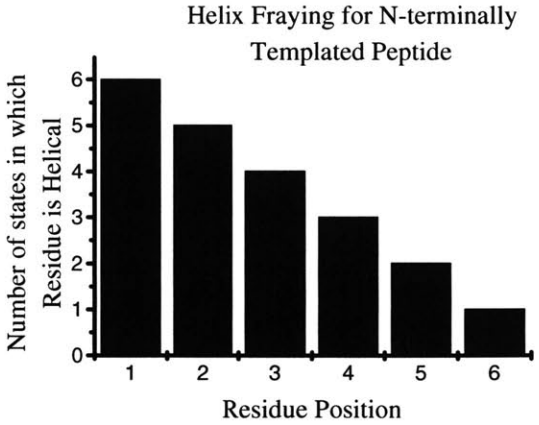
of the equilibrium expression increases exponentially as the size of the peptide increases, especially as one must begin to consider states that have multiple distinct helices. There are 11 terms in the sum of equilibrium constants for the six residue peptide AcX1X2X3X4X5X6-NH_2. For a 10 residue peptide, there are 72. For a 20 residue peptide there are 8,855. Even considering only the states in which there is only a single helix in the peptide (the one-helix approximation), the number of states increases very rapidly with peptide length. A 10 residue peptide has 22 such states and a 20 residue peptide has 172. To handle the massive increase in complexity that is seem in medium length peptides, a method based on matrix multiplication has been introduced into the Zimm-Bragg model to calculate the sum of equilibrium constants. This method is used in chapter 3 to calculate the helicity of a 13 residue peptide, but it is not required to understand the basic features of the model or the vast majority of the work in this thesis. The development of the matrix method is therefore relegated to appendix 1.

2.4 The Advantages of N-terminally Templated Systems

An N-terminal template is a structure that induces helicity when attached to a peptide by mimicking a helical structure to which subsequent residues can conform. If the Zimm-Bragg postulate of a difficult initiation step is correct, then such a template should greatly enhance the helicity of a linked peptide and thereby simplify the study of peptide helicity in two powerful ways. First, helicity could be studied in very short peptides that would never adopt a perceptible amount of helical structure under normal conditions.

Figure 6. The states that make up the helix-coil equilibrium of a N-templated, six residue peptide in cartoon form, and a bar graph showing the dependence of the number of states in which a residue is helical on the residue's position. The states that are crossed out are much less stable than the others because they contain helices that do not initiate at the template.





The second is illustrated in figure 6, where the helical states in an N-templated six residue peptide are shown schematically. Since the helical states that initiate at the template

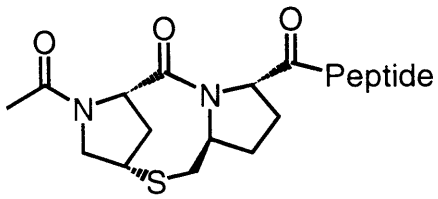
would be vastly more populated than those that initiate elsewhere, all of the helical states that do not initiate at the template could be ignored without compromising accuracy. This second feature of N-templated peptides also has an effect on the fraying of the helix. Whereas untemplated peptides fray bidirectionally from the center outward, N-templated peptides fray unidirectionally from the N-terminus to the C-terminus. This is illustrated in the bar graph at the bottom of the figure, where the number of states in which a residue is helical is plotted against residue position (compare this bar graph to the one in figure 5).

The concept of an N-terminal template was introduced independently by Kemp and Curran,[29] and by Satterthwait and Lerner[30] in 1988. Subsequently, three more N-terminal templates have appeared in the literature.[31,32,,33,3435] The next chapter of this thesis uses the particular template $AcHel_1$, some variants of which were used in the preceding chapter to study hydrogen bonding, to study peptide helicity. The general structure of $AcHel_1$-peptide conjugates is shown in figure 7.

---

[29] Kemp, D. S.; Curran, T. P. *Tetrahedron Lett.* **1988**, *29*, 4931.
[30] Satterthwait, A. C.; Arrhenius, T.; Hagopian, R. H.; Zavala, F.; Nussensweig, V.; Lerner, R. A. *Vaccine* **1988**, *6*, 99.
[31] Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd, J. G.; Muendel, C. C. *J. Org. Chem.* **1991**, *56*, 6672.
[32] Kemp, D. S.; Curran, T. P.; Boyd, J. G.; Allen, T. *J. Org. Chem.* **1991**, *56*, 6683.
[33] Müller, K.; Obercht, D.; Knierzinger, A.; Stankovic, C.; Spiegler, C.; Bannwarth, W.; Trzeciak, A.; Englert, G.; Labhardt, A. M.; Scoenholzer, P. *Perspect. Med. Chem.* **1993**, 513.
[34] Kemp, D. S.; Rothman, J. H. *Tet. Lett.* **1995**, *36*, 4023.

Figure 7. The N-terminal helix initiating template, $AcHel_1$.



$AcHel_1$

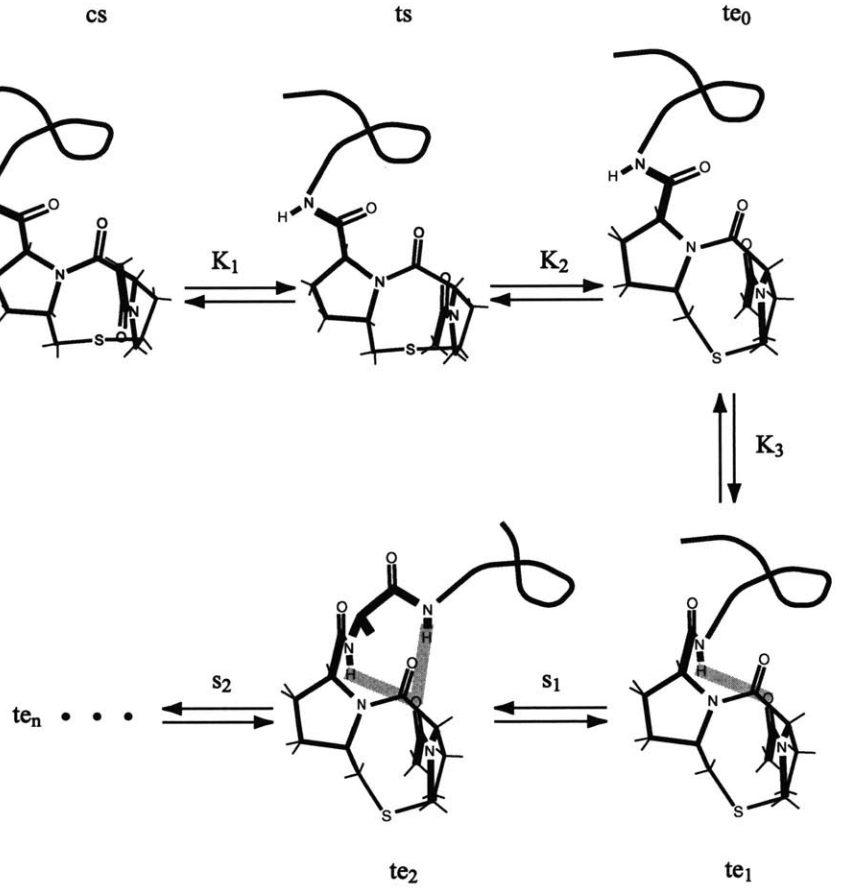
Helix nucleation by $AcHel_1$ was demonstrated and characterized by Kemp and co-workers in a series of papers in 1995 and 1996.[36,37,38] $AcHel_1$ initiates helices by providing carbonyl groups that are spatially disposed so that they can accept hydrogen bonds from the three most terminal amide NH groups that would normally not have hydrogen bonding partners in an untemplated helix. Initiation in this way by $AcHel_1$ has been shown to be about 50 times more efficient than spontaneous initiation (σ for the template is on the order of 10^{-1} compared to ~ 2×10^{-3} for spontaneous initiation).[38] $AcHel_1$ is thus in the category of N-terminal templates that can simplify the study of peptide helicity. Indeed, $AcHel_1$ has been used extensively to probe the helicity of alanine rich peptides,[38,39,40,41] and the key feature of monotonic helix fraying has been demonstrated experimentally by placing glycine at each position from the template-peptide junction to the C-terminus. Since glycine is a helix-breaking residue, its presence

---

[35] Austin, R. E.; Maplestone, R. A.; Sefler, A. M.; Liu, K.; Hruzewicz, W. N.; Liu, C. W.; Cho, H. S.; Wemmer, D. E.; Bartlett, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6461.
[36] Kemp, D. S.; Allen, T. J.; Oslick, S. O. *J. Am. Chem. Soc.* **1995**, *117*, 6641.
[37] Kemp, D. S.; Allen, T. J.; Oslick, S. O.; Boyd, J. G. *J. Am. Chem. Soc.* **1996**, *118*, 4240.
[38] Kemp, D. S.; Oslick, S. O.; Allen, T. J. *J. Am. Chem. Soc.* **1996**, *118*, 4249.
[39] Kemp, D. S.; Boyd, J. G. Muendel, C. C. *Nature* **1991**, *352*, 451.
[40] Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2025.
[41] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 12234.

near the N-terminus arrests helix propagation near the template, and overall helicity of the peptide is low. The effect is seen to decrease as glycine is moved closer to the C-terminus, indicating that there is less helicity at the C-terminus to disrupt.[39,42]

Figure 8. Template and helix-coil equilibria for an n residue $AcHel_1$-peptide conjugate.



The equilibria that occur in an n residue $AcHel_1$-peptide conjugate are illustrated in figure 8. The first three conformational transitions in this figure take place within the template alone, with the peptide remaining as a random coil throughout. These are

[42] Lee, J. H. Thesis, Massachusetts Institute of Technology, 1997.

similar to the conformational transitions that were characterized in chapter 1, but not exactly the same since they are in an abbreviated form (helix initiation from the ts state is not considered, since the NH group at the template-peptide junction has been shown not to form a detectable hydrogen bond in this state,[36] and the unoriented ⇄ oriented equilibrium about the C5-C5' bond is not considered explicitly). The first of the template equilibria is the cis (cs) to trans (ts) equilibrium of the acetamide where the C8-C9 bond remains in its staggerred conformation. The second is the staggered (ts) to eclipsed (te_0) equilibrium about the C8-C9 bond within the trans state. The last is the non-hydrogen bonded (te_0) to hydrogen bonded (te_1) equilibrium within the te state (subscripts within the e state indicate the number of hydrogen bonds that have been formed). Their equilibrium constants are K_1, K_2, and K_3 respectively. The equilibria that follow involve helix formation to progressively greater extent, and their equilibrium constants are the s values of the residues that are joining the helix. Taking the cs-random coil state of the template-peptide conjugate as the reference state, the sum of all of the states' equilibrium constants for an n residue $AcHel_1$-peptide conjugate is

$$Z = 1 + \frac{[ts]}{[cs]} + \frac{[te_0]}{[cs]} + \frac{[te_1]}{[cs]} + \frac{[te_2]}{[cs]} + \cdots + \frac{[te_n]}{[cs]}$$

$$= 1 + \frac{[ts]}{[cs]} + \frac{[ts]}{[cs]}\frac{[te_0]}{[ts]} + \frac{[ts]}{[cs]}\frac{[te_0]}{[ts]}\frac{[te_1]}{[te_0]} + \frac{[ts]}{[cs]}\frac{[te_0]}{[ts]}\frac{[te_1]}{[te_0]}\frac{[te_2]}{[te_1]} + \cdots + \frac{[ts]}{[cs]}\cdots\frac{[te_n]}{[te_{n-1}]}$$

$$= 1 + K_1 + K_1K_2 + K_1K_2K_3 + K_1K_2K_3s_1 + K_1K_2K_3s_1s_2 + \cdots + K_1K_2K_3s_1s_2\cdots s_n$$

Making the substitutions $A = K_1 + K_1K_2$ and $B = K_1K_2K_3$ yields the simpler expression

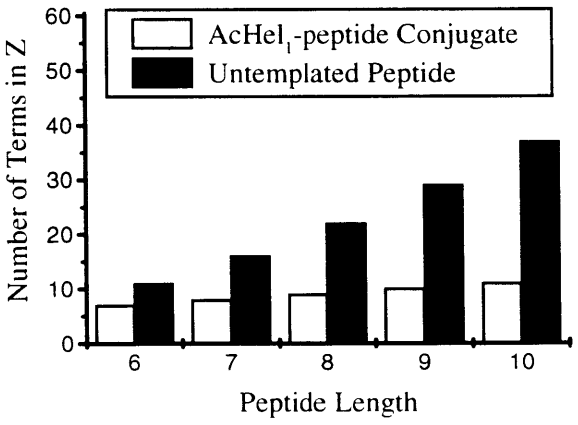
$$Z = 1 + A + B + B(s_1 + s_1 s_2 + \cdots + s_1 s_2 \cdots s_n) \qquad \textbf{2 - 5}$$

The constants A and B have been experimentally determined,[41,43] with A = 0.832 and B = 0.156, so that

$$Z = 1.988 + 0.156 \times (s_1 + s_1 s_2 + \cdots + s_1 s_2 \cdots s_n)$$

The only unknowns in the sum of equilibrium constants for $AcHel_1$-peptide conjugates are the s values of the residues.

Figure 9. Comparison of the number of states that have to be considered (the number of terms in Z) in the helix-coil $AcHel_1$-peptide conjugates and untemplated peptides of the same length.



An n residue $AcHel_1$-peptide conjugate's sum of helix-coil equilibrium constants has the term 1.988 plus one term for each N-terminally initiated helical state for a total of 1+n terms. The total number of terms for a six residue peptide attached to $AcHel_1$ is therefore 7, compared to 11 for an untemplated six residue peptide, as determined in section 2.3. This does not seem such a vast improvement, but the difference becomes larger as the

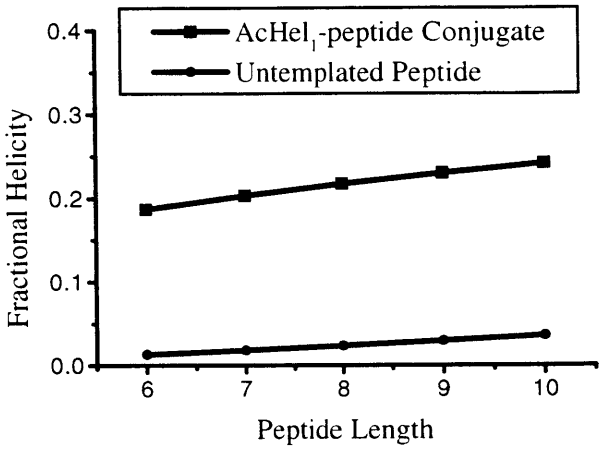
[43] Note that 0.832 is the value that was used in chapter 1 for $(t/c)_{ref}$ for the 5' acyl $AcHel_1$ derivatives. This is because $A = K_1 + K_1K_2 = ([ts]+[te_0])/[cs]$, which is the t/c ratio for $AcHel_1$-peptide derivatives excluding the hydrogen bonding states.

peptide becomes longer. This is illustrated in figure 9, where the number of terms in the sums of equilibrium constants for $AcHel_1$-peptide conjugates and untemplated peptides for increasing peptide lengths is plotted in a bar graph. As shown in the figure, the sum of equilibrium constants for an eight residue $AcHel_1$-peptide conjugate has only half as many terms as that of an untemplated peptide of the same length. For a twelve residue peptide this figure is down to one quarter. This advantage of studying the helix-coil equilibrium in $AcHel_1$ templated systems should be considered together with the much greater helicities of $AcHel_1$-peptide conjugates compared to untemplated peptides. The fractional helicity of an n residue $AcHel_1$-peptide conjugates according to relation 2-3 is given by

$$\text{fractional helicity} = \frac{B\left(\frac{1}{n}s_1 + \frac{2}{n}s_1s_2 + \cdots + \frac{n}{n}s_1s_2\cdots s_n\right)}{Z} \qquad \boxed{2\text{-}6}$$

The fractional helicities expected for $AcHel_1$-peptide conjugates of varying length with $s_1 = s_2 = \cdots = s_n = 1$ are plotted in figure 10 with the fractional helicities of untemplated peptides of the same length for comparison. As shown in the plot, the helicities of the $AcHel_1$-peptide conjugates ought to be much more easily detectable, and therefore easier to study, than the helicities of the untemplated peptides.

Figure 10. Comparison of the fractional helicities of $AcHel_1$-peptide conjugates and untemplated peptides of the same length, assuming s = 1 for all residues.



2.5 Methods of Identifying and Quantifying Helicity

The study of peptide helicity has been blessed with an abundance of methods to detect the presence of helices and, in some cases, to quantify their extent. In particular, IR and NMR spectroscopy, circular dichroism (CD), amide hydrogen exchange, and, recently, reporting conformational templates have been used.

Although they are not used in this thesis, several IR based methods for studying peptide structure have been developed.[44] These methods all take advantage of the sensitivity of the various amide modes to conformation.

Many features of NMR spectra respond to peptide helicity. The coupling constant between the α and NH protons is a function of the ϕ torsion, and is unusually small for an

[44] Bandekar, J. *Biochim. Biophys. Acta* **1992**, *1120*, 123.

amino acid involved in a helix ($^3J_{\alpha NH} < 6$ Hz).[45,46] The chemical shifts of α and NH protons are shifted upfield for amino acids in helices.[47,48,49,50] The strongest NMR evidence for the presence of helices, though, comes from NOE experiments. Helices are very compact structures and several of the protons of consecutive and non-consecutive residues are close to each other.[51] All of these NMR based techniques will be discussed further in chapter 4. The information gained from NMR spectroscopy, and especially NOEs, is extremely powerful for identifying helices, but difficult to use quantitatively to determine the extent of helicity.

In circular dichroism (CD) spectroscopy, the perturbation of a chromophore's UV absorbance by its chiral environment is measured. Helical peptides (in the α form) have very distinctive CD spectra, with minima at 208 and 222 nm.[52] The minimum at 208 nm shifts to lower wavelength and tends to become more intense as peptides become less helical, while the minimum at 222 nm simply becomes less intense. In the limit of an absolute lack of structure, the CD spectrum has a minimum at 195 nm and is almost featureless above 220 nm.[52] The intensity of the 222 nm minimum (measured in units of per residue molar ellipticity: deg cm^2 / dmol res) is believed to be largely independent of the particular amino acids in the helix and directly proportional to a peptide's fractional helicity according to[53]

---

[45] Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, **1986**, pp 167-168.
[46] Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.
[47] Williamson, M. P. *Biopolymers* **1990**, *29*, 1423.
[48] Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.
[49] Ösapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436.
[50] Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.
[51] Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715.
[52] Greenfield, N.; Fasman, G. D. *Biochemistry*, **1969**, *8*, 4108.
[53] Chen, Y.-H., Yang, J. T.; Chau, K. H. *Biochemistry*, **1974**, *13*, 3350.

$$\frac{[\theta]_{222,observed} - [\theta]_{222,random\ coil}}{[\theta]_{222,100\%helix} - [\theta]_{222,random\ coil}} = \text{fractional helicity} \qquad \textbf{2 - 7}$$

Where $[\theta]_{222,observed}$ is the peptide's experimentally determined per residue molar ellipticity, $[\theta]_{222,random\ coil}$ is the per residue molar ellipticity of the random coil state and $[\theta]_{222,100\%\ helix}$ is the per residue molar ellipticity expected for a maximally helical peptide. This method provides a convenient measure of a peptide's global helicity (although it does not specify which regions in a peptide are helical), provided that accurate values for $[\theta]_{222,100\%\ helix}$ and $[\theta]_{222,random\ coil}$ are available. Unfortunately, a range of values have been reported for both of these quantities, especially $[\theta]_{222,100\%helix}$. In this thesis, therefore, $[\theta]_{222,randomcoil}$ will be set to 0, and −32,000 deg cm^2 / dmol res and −42,000 deg cm^2 / dmol res will be set as the upper[54] and lower[55] bounds for $[\theta]_{222,100\%helix}$. This range should subsume any other corrections (for example, for aromatic residues or length) that might have to be made. For short peptides bound to an N-terminal template, the values for $[\theta]_{222,100\%helix}$ and $[\theta]_{222,randomcoil}$ obtained from $AcHel_1AAAAAA\text{-}NH_2$ will be used: −28,520 deg cm^2 / dmol res and −1,890 deg cm^2 / dmol res.[56]

The amide hydrogen exchange method relies on peptide amide hydrogens being unable to exchange with the solvent while they are hydrogen bonded to non-solvent acceptors,[57] as they are in helices. In cases where helix formation is the only factor that can retard hydrogen exchange, the observed exchange rate constant of residue i's amide in a peptide with some helical character (k_{obs}) is related to intrinsic exchange rate constant expected for the amide in an unstructured state (k_{int}) as follows[57]:

---

[54] Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5286.
[55] Andersen, N. H.; Tong, H. *Protein Sci.* **1997**, *6*, 1920.

$$k_{obs} = f_{nhb,i} \times k_{int} = (1 - f_{hb,i}) \times k_{int} \qquad \textbf{2 - 8}$$

where $f_{nhb,i}$ is the fraction of the time in which the amide is not hydrogen bonded in a helix, and $f_{hb,i}$ is the complementary fraction of the time in which amide i is hydrogen bonded in a helix. The exchange rate constant, k_{obs}, can be measured from the exchange reaction an amide hydrogen on a peptide that has been dissolved in deuterated water. This typically involves quickly dissolving a peptide in D_2O and following the disappearance of the peptide's NH peaks in an NMR spectrum or the change in the peptide's mass by mass spectroscopy. The value for k_{int} can either be calculated from literature data[58] or determined experimentally from unstructured model peptides.[59] For a more complete discussion of the amide hydrogen exchange method and a critical evaluation of the methods for determining k_{int}, see chapter 5.

The last method for quantifying peptide helicity is specific to $AcHel_1$-peptide conjugates, and it takes advantage of the reporting feature of the c ⇄ t equilibrium of the acetamide group that was used in chapter 1 for the study of hydrogen bonding. Since helix formation can only take place when the acetamide group of the $AcHel_1$ unit is in the trans state, peptides that are more prone to be helical must cause $AcHel_1$'s c ⇄ t equilibrium to shift more towards the trans state.[36,37,38] This should increase the c ⇄ t equilibrium constant, or t/c ratio, which is measureable by NMR as discussed in chapter 1. The relationship between the experimentally measured t/c ratio and the sum of helix-coil equilibrium constant can be made explicit by noting that it is the sum of trans state equilibrium constants divided by the cis state equilibrium constant (which is 1):

---

[56] Oslick, S. L. Thesis, Massachusetts Institute of Technology, **1996**.
[57] Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1984**, *16*, 521.

$$\left(\frac{t}{c}\right) = A + B + B(s_1 + s_1 s_2 + \cdots + s_1 s_2 \cdots s_n) \qquad \boxed{2-9}$$

The t/c ratio is perhaps the most intrinsically accurate measure of peptide helicity since it does not require that a quantity for some idealized state be known, such as the per residue molar ellipticity at 222 nm for a 100% helix or the intrinsic exchange rate constant for a peptide amide in a random coil state. Relating the t/c ratio to the helix coil equilibrium requires only the known template constants A and B.

2.6 Motivation for Studies of Peptide Helicity

According to the Zimm-Bragg model as presented so far, the helicity of any peptide can be completely described if the universal value for σ and the s values for all of the different types of residues are known. The opportunity to have this much information about the conformational states of a peptide has inspired several efforts to determine these parameters. Varying values for σ have been obtained, but values around 1.5×10^{-3} to 2.5×10^{-3} are most commonly used.[60,61] The midpoint of this range, $\sigma = 2.0 \times 10^{-3}$, will be used in this thesis where required. The s values from four studies of peptide helicity are recorded in table 2 along with, for comparison, the relative frequency with which each residue occurs in the middle of α helices in proteins of known structure.

---

[58] Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75.
[59] Rohl, C. A.; Baldwin, R. W. *Biochemistry* **1994**, *33*, 7760.
[60] Rohl, C. A.; Scoltz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1992**, *31*, 1263.
[61] Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623.

Table 2. Second through fifth columns: helix propensities (s values) experimentally determined according to the Zimm-Bragg (or Lifson-Roig) model. Last column (P_{mid}): relative frequency with which a residue occurs in the middle of α helices in known protein structures. The asterisk indicates that these numbers, unlike those in the rest of the table, are not s values.

Residue	Scheraga[62]	Baldwin[61]	Stellwagen[63,64]	Kallenbach[65]	P_{mid}*[66]
Ala	1.07	1.64	1.81	1.92	1.41
Cys	0.99	0.31	0.43	0.23	0.66
Asp^0	0.78	0.39	0.24		
Asp^-	0.68	0.37	0.24	0.47	0.99
Glu^0	1.35	0.68	1.00		
Glu^-	0.97	0.52	0.88	1.31	1.18
Phe	1.09	0.26	0.79	0.26	1.16
Gly	0.59	0.05	0.05	0.29	0.43
His^0	0.85	0.35	0.21	0.73	1.05
His^+	0.69	0.21	0.21		
Ile	1.14	0.44	0.43	0.96	1.09
Lys^+	0.94	0.97	1.25	0.67	1.23
Leu	1.14	0.84	1.03	1.00	1.34
Met	1.20	0.63	0.79	0.74	1.30
Asn	0.78	0.28	0.23	0.41	0.76
Gln	0.98	0.60	0.58	0.45	1.27

[62] Wojcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers*, **1990**, *30*, 121.
[63] Park, S.-H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 7048.
[64] Park, S.-H.; Shalongo, W.; Stellwagen E. *Biochemistry* **1993**, *32*, 12901.
[65] Yang, J.; Spek, E.; Gong, Y.; Zhou, H.; Kallenbach, N. R. *Protein Sci.* **1997**, *6*, 1264.
[66] Williams, R. W.; Chang, A.; Juretic, D.; Loughran, S. *Biochem. Biophys. Acta* **1987**, *916*, 200.

Arg^+	1.03	1.10	1.94	0.91	1.21
Ser	0.76	0.39	0.29	0.48	0.57
Thr	0.82	0.17	0.18	0.39	0.76
Val	0.95	0.24	0.18	0.51	0.98
Trp	1.11	0.28	0.58	0.26	1.02
Tyr	1.02	0.46	0.43	0.06	0.74

That helix propensities are intrinsic properties of amino acid residues to at least some extent is evident from inspection of the table in that, despite the differences in the systems in which these helix propensities were measured, the various sets of data share some common features. Serine, threonine, asparagine, aspartate, and glycine rank near the bottom in all of the sets of s values and in the P_{mid} scale. Similarly, alanine, leucine, and arginine rank near the top. However, a first sign that the s values in table 2 may be inadequate for describing peptide helicity comes from comparing the various sets. Despite the general agreement about what the most and least helix stabilizing residues are, and although each set of s values succeeds in describing the helix formation in the systems from which they were derived, there are significant differences in the s values across the sets of data. For example, the s values reported for Ile (1.14 (Scheraga), 0.44 (Baldwin), 0.43 (Stellwagen) and 0.96 (Kallenbach)) differ by up to 260%. The s values reported for Phe (1.09 (Scheraga), 0.26 (Baldwin), 0.79 (Stellwagen), and 0.26 (Kallenbach)) differ by up to 420%. The correlation coefficients between the data sets are also not generally very good (especially between the Scheraga s values and the other three sets). They range from 0.32 to 0.91, with an average of 0.62.

Explicit demonstrations of the shortcomings in the assumption that peptide helicity can be described using s values that are independent of a residue's environment (context independent) come from studies in which it is shown that interactions between residues affect helicity. Because helices are such compact structures, there are a number of potential causes of context dependence for helix propensities. For example, in a helix, the side chain of residue i (if it is more than a couple of atoms long) can interact with the side chains of residues i+3 or i+4. This can occur by hydrogen bonding if the side chains have hydrogen bond donors and acceptors,[67,68,69,70] electrostatic attraction (or repulsion) if the side chains are oppositely (or identically) charged,[18,67,68] ion-dipole interactions,[71] or hydrophobic clustering if both side chains are non-polar.[72,73,74] Such interactions would make the helix propensity of residue i a function not only of residue i's type, but also of the types of residues at positions i+3, and i+4. It has also been postulated that the helix propensities of residues with charged side chains can be affected by the putative helix dipole.[75,76,77,78] Having a positively charged residue at the negative end of the helix dipole (the C-terminus) or a negatively charged residue at the positive end of the helix dipole (the N-terminus) would increase a helix propensity while the opposite situations would decrease a helix propensity. Either way, such charge–dipole interactions would

---

[67] Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biophys. Chem.* **1990**, *37*, 107.
[68] Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668.
[69] Huyghues-Despointes, B. M. P.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *41*, 13267.
[70] Stapley, B. J.; Doig, A. J. *J. Mol. Biol.* **1997**, *272*, 465.
[71] Shoemaker, K. R.; Fairman, R.; Schultz, D. A.; Robertson, A. D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1990**, *29*, 1.
[72] Padmanabhan, S.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *241*, 706.
[73] Padmanabhan, S.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 1992.
[74] Zerkowski, J. A.; Powers, E. T.; Kemp, D. S. *J. Am. Chem. Soc.* **1997**, *119*, 1153.
[75] Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563.
[76] Richardson, J. S.; Richardson, D. C.; *Science* **1988**, *240*, 1648.
[77] Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Proteins* **1989**, *5*, 1.
[78] Nicholson, H.; Anderson, D. E.; Dao-pin, S.; Matthews, B. W. *Biochemistry* **1991**, *30*, 9816.

lead to a position dependence of helix propensity. Helix initiation could also be context dependent. The side chain of the first residue with random coil ϕ and ψ angles at the N-terminus of a helix could stabilize the helix by hydrogen bonding to the two NH groups of the residues between it and the first hydrogen bonded residue of the helix. Such interactions are called N-capping interactions (similar C-capping interactions are also known).[79,80,81]

The interactions listed in the preceding paragraph are only those that have been most widely studied. Several other types of interactions are conceivable. For example, side chains could interact with the helix barrel,[39,41,82] resulting in a dependence of a residue's helix propensity on its position in a helix. Interactions could occur between adjacent residues (i, i+1 interactions) provided their side chains were more than a few atoms long. Furthermore, interactions could occur between nearby residues in the random coil state. If such interactions stabilized (or destabilized) the random coil, they would destabilize (or stabilize) the helix.

How can such a complicated problem be profitably studied? Two approaches have been used in the past. The first is the global approach, exemplified by the Baldwin group's study of the ribonuclease C-peptide (see chapter 3 for a review), in which small changes are made in complicated systems and the observed effects are rationalized. The second is the building-up approach, exemplified by the Scheraga group's host-guest studies,[62,83] in which simple systems are characterized at first, then complicating factors are added and the new systems are studied until they are understood, then more

---

[79] Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632.
[80] Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21.
[81] Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325.
[82] Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5585.

complicating factors are added, etc. These approaches have provided the data necessary to construct peptide helicity models with new parameters to account for the effects of interactions. Several such models have been proposed,[84,85,86,87,88,89,90,91] some of them with dozens of parameters in addition to helix propensities and a helix initiation constant, to explain the more well studied of the interactions described above (the capping, charge-helix dipole, and i, i+3 and i+4 side-chain to side-chain interactions). However, both the global and the building-up approaches are limited. The complexity of the systems studied in the global approach ultimately limit one's ability to attribute observed effects to specific causes, while the number of iterations needed in the building-up approach to finally arrive at an understanding of very complicated systems ultimately limits one's ability to achieve results in a reasonable time scale. Therefore, a new approach that lies conceptually somewhere between the global and the building-up approaches is used in the remainder of this thesis.

The next chapter introduces a new, reductionist approach, in which the helicity of short, highly heterogeneous peptides are studied via their N-terminally templated forms. The heterogeneity of the peptides ensures that the networks of interactions that are present in protein fragments can be observed, while their length makes their study more tractable. The N-terminal template serves to further simplify the interpretation of observed effects (by reducing the complexity of the helix-coil equilibrium) and to

---

[83] Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein. Chem.* **1975**, *29*, 205.
[84] Scheraga, H. A.; Vásquez, M. *Biopolymers* **1988**, *27*, 41.
[85] Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605.
[86] Muñoz, V.; Serrano, L. *Nature: Struct. Biol.* **1994**, *1*, 399.
[87] Muñoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 275.
[88] Muñoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 297.
[89] Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 3396.
[90] Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623.
[91] Andersen, N. H.; Tong, H. *Protein Sci.* **1997**, *6*, 1920.

introduce helicity into the peptides despite their length. The reductionist approach is used in the following chapter to study the helicity of the ribonuclease C-peptide via the $AcHel_1$ conjugates of its fragments.

Appendix 1. An Algorithm for Computing Sums of Helix-Coil equilibrium Constants

The sum of equilibrium constants for an n residue peptide, denoted Z, can be expressed as the matrix product[23]

$$Z = \alpha\left(\prod_{ji=3}^{n} \mathbf{M}_j\right)\omega \qquad \textbf{2 - 10}$$

where ω, α and $\mathbf{M_j}$ are

$$\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix};\ \omega = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix};\ \mathbf{M}_j = \begin{bmatrix} 1 & \sigma s_j & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & s_j & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & s_j \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & s_j & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & s_j \end{bmatrix}$$

The index j in the matrix product counts along the residues in the peptide. It starts at the third residue because the closing of the first and second residues' ϕ and ψ torsions into the helical range never results in the formation of an i, i+4 hydrogen bond, so these residues

can only ever contribute a residue-independent factor of σ into the equilibrium constant for a state.

An n residue peptide's fractional helicity can be obtained convenintly from Z as follows

$$\text{fractional helicity} = \frac{\left[\sum_{\substack{\text{each type} \\ \text{of residue, X}}} \frac{s_X}{Z}\frac{\partial Z}{\partial s_X}\right] + 2\frac{\sigma}{Z}\frac{\partial Z}{\partial \sigma}}{n} \qquad \textbf{2 - 11}$$

where the index of the sum in the numerator counts along the types of residues in the peptide (there should be one term for every type of residue). The bracketed sum corresponds to the average number of hydrogen bonded residues in the peptide. The extra σ-containing term adds the two N-terminal residues in every helix that are not hydrogen bonded but have ϕ and ψ torsions in the helical range. The numerator by itself thus is the average number of residues with ϕ and ψ torsions in the helical range. Dividing by the length of the peptide gives the fractional helicity.

The fractional hydrogen bonding of the ith residue in an n residue peptide is given by the following expression

$$f_{hb,i} = \frac{\alpha\left(\prod_{j=3}^{i-2}\mathbf{M}_j\right)\times\mathbf{DM}_{i-1}\times\left(\prod_{j=i}^{n}\mathbf{M}_j\right)\omega}{Z} \qquad \textbf{2 - 12}$$

Where α and ω and $\mathbf{M}_j$ are defined as above and $\mathbf{DM}_{i-1}$ (the dummy matrix) is

$$\mathbf{DM}_{i-1} = \begin{bmatrix} 0 & \sigma s_{i-1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_{i-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{i-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_{i-1} \end{bmatrix}$$

The effect of the dummy matrix is to eliminate all of the terms from the matrix product in the numerator that correspond to states in which residue i is not hydrogen bonded (i.e. those in which residue i-1 does not contribute an s to the states equilibrium constant). The quotient 2-12 is the sum of the equilibrium constants for the states in which i is hydrogen bonded divided by the sum of all equilibrium constants, as required by equation 2-4.

Chapter 3. The Reductionist Approach Applied to the Study of the Ribonuclease C-Peptide's Helicity

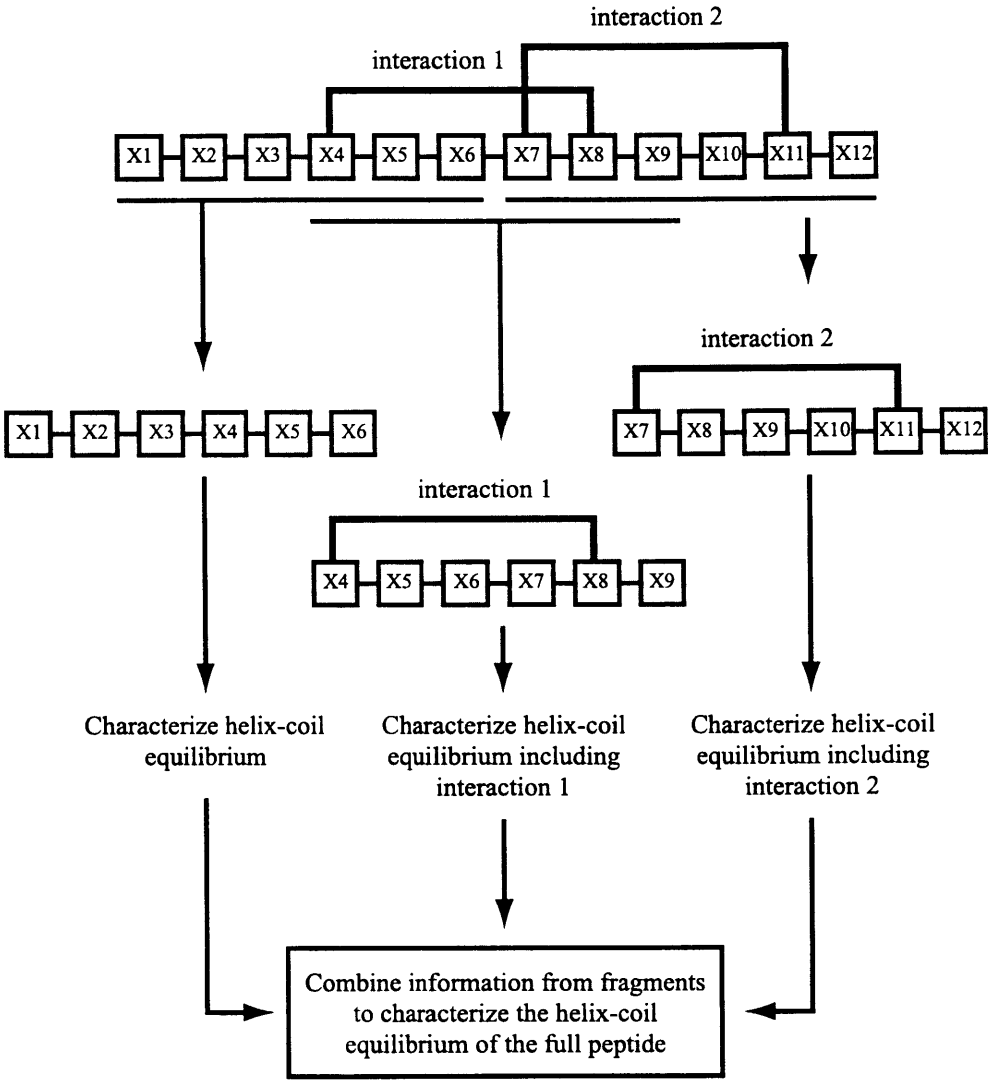
3.1 Introduction

In this chapter, the use of reductionist approaches to the study of peptide helicity is gauged by a preliminary study of a specific peptide, the ribonuclease C-peptide, by a novel method, the divided peptide method. The C-peptide comprises the 13 N-terminal residues of bovine pancreatic ribonuclease (RNase-A) and it was the first short peptide shown to adopt a helical conformation in aqueous solution.[1,2] In the divided peptide method, illustrated schematically in figure 1, peptides are broken into fragments, the fragments are studied in isolation, and the information gained is integrated in an attempt to recover a description of the full peptide's helicity. The application of the divided peptide approach will be described in full detail in section 3.3 of this chapter, while sections 3.4 and 3.5 present and discuss the results from the approach to the C-peptide system. First, however, the following section relates relevant background information on the C-peptide.

---

[1] Brown, J. E.; Klee, W. A. *Biochemistry* **1969**, *8*, 2876.
[2] Brown, J. E.; Klee, W. A. *Biochemistry* **1971**, *10*, 470.

Figure 1. The divided peptide method applied to a twelve residue peptide. Each fragment contains the information necessary to explain the helicity, including interactions, of the corresponding region of the full peptide.



3.2 Background Information on the C-peptide

3.2.1 History

Before recounting the history of the C-peptide it is necessary to describe some experiments that preceded the C-peptide's discovery. In the late 1950's F. M. Richards found that subtilisin cleaved RNase-A in only one place to produce two fragments.[3,4] The first, called the RNase S-peptide, consisted of the N-terminal 20 residues of ribonuclease (sequence: KETAAAKFERQHMDSSTSAA), and the second, called the RNase S-protein, consisted of the remaining 104 residues. The two fragments could be separated, but neither the S-peptide nor the S-protein by themselves showed any ribonuclease activity. However, when mixed together in a 1:1 ratio these two components associated to produce RNase S, a complex that had nearly the full activity of the unmodified protein.[5]

Somewhat after these studies, the ribonuclease C-peptide was first prepared and identified by Gross and Witkop as a product of the cyanogen bromide cleavage of bovine pancreatic ribonuclease in a demonstration of the reagent's utility.[6] They showed it to consist of the first 13 residues of ribonuclease, with the last residue being converted to a homoserine (lactone or free acid, depending on the conditions) as a result of the cyanogen bromide treatment. Its sequence was therefore KETAAAKFERQH-Hse (where Hse is the three-letter code for homoserine). The year after the preparation of the C-peptide Hofmann *et al* showed that a synthetic peptide whose sequence matched the 13 N-

---

[3] Kalman, S. M.; Lindstrøm-Lang, K.; Ottesen, M.; Richards, F. M. *Biochim. Bipophys Acta* **1955**, *16*, 297.
[4] Richards, F. M. *Compt. rend. Lab. Carlsberg, ser. Chim.* **1955**, *29*, 329.
[5] Richards, F. M. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 162.

terminal residues of RNase A (KETAAAKFERQHM), and differing from the C-peptide only at the C-terminal residue, could also induce activity in the RNAse S-protein although a 10:1 peptide: protein ratio was needed.[7] Following this, Parks *et al* demonstrated that the C-peptide itself was similarly capable of inducing activity in the RNase S-protein.[8]

It seemed likely that the C-peptide and the S-peptide reconstituted activity in the S-protein by binding to it in the location and conformation in which they reside in unmodified RNase A. Strong evidence for this hypothesis came from the X-ray structure of the RNase S complex.[9] Of particular interest was that residues 3 through 12 of the S-peptide, which make up almost the entire C-peptide sequence, adopted a helical conformation just as in RNase-A.

The next logical question was, how much of the structure that the C-peptide obtained on binding to the S-protein was already present in its uncomplexed solution form? This was answered when Brown and Klee showed that the free C-peptide in aqueous solution was up to 25% helical[1,2] under optimal helix-forming conditions. The C-peptide together with the S-peptide (which showed similar helicity, but only in the region of its sequence that corresponded to the C-peptide[1]) were the first, and for a time the only, short peptides (< 20 residues) known to display any observable helicity in aqueous solution. In spite of this very interesting result, the study of the C-peptide lay dormant for around 10 years after this work.

---

[6] Gross, E.; Witkop, B. *J. Biol. Chem.* **1962**, *237*, 1856.
[7] Hofmann, K.; Finn, F.; Haas, W.; Smithers, M. J.; Wolman, Y.; Yanaihara, N. *J. Am. Chem. Soc.* **1963**, *85*, 833.
[8] Parks, J. M.; Barancik, M. B.; Wold, F. *J. Am. Chem. Soc.* **1963**, *85*, 3519.
[9] Wyckoff, H. W.; Tsernoglou, D.; Hanson, A. W.; Knox, J. R.; Lee, B.; Richards, F. M. *J. Biol. Chem.* **1970**, *245*, 305.

Interest in the helicity of the C-peptide was revived by the Baldwin group, who made a concerted effort to determine why the C-peptide was helical when, according to the Zimm-Bragg model for peptide helicity and the σ and s values determined from random copolymer systems,[10] it should not have been. Their studies used the pH dependence of the C-peptide's helicity and systematic alteration of the C-peptide's sequence to identify residues that were involved in helix-stabilizing interactions. Their results are described in the next section in which is discussed the forces at the root of the C-peptide's helicity.

3.2.2 Proposed Origins of the C-Peptide's Helicity

The C-peptide's helicity is strongly temperature dependent,[2] increasing monotonically from 26 °C to 1 °C, and is favored by high ionic strengths,[2] increasing monotonically from μ = 0 to 10 M. It also shows a strong dependence on pH, with the maximum helicity near pH 5, and much reduced helicities at low and high pH.[11] According to CD spectroscopy, the fractional helicity at pH 5 is around twice what it is at pH < 3 and pH > 8 as shown in figure 2; recall from section 2.5 that helicity is larger as the CD signal at 222 nm, $[\theta]_{222}$, becomes more negative. (These data pertain to the C-peptide lactone only; the helicity of the C-peptide with a free C-terminal carboxylate is drastically smaller.[12]) The C-peptide's per residue ellipticities at pH 2, 5, and 9 can be converted into fractional helicities as detailed in section 2.5. As described there, a range of $[\theta]_{222,100\%helix}$ values will be used to provide a range of fractional helicities. The ranges of

[10] Wojcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121.
[11] Bierzynski, A.; Kim, P. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2470.
[12] Kim, P. S.; Bierzynski, A.; Baldwin, R. L. *J. Mol. Biol.* **1982**, *162*, 187.

s values that would yield these fractional helicities in a homopeptide the same length as the C-peptide (in this case, an 11 residue homopeptide since neither the first lysine nor the Hse lactone can have their ϕ and ψ angles fixed by helix formation) can then be determined using the plot in figure 3, which shows the dependence of fractional helicity on s value for an 11 residue homopeptide. The per residue ellipticities, the calculated fractional helicities, and the corresponding homopeptide equivalent s values are listed in table 1.

Figure 2. Dependence of the C-peptide's per residue molar ellipticity at 222 nm ($[\theta]_{222}$), and therefore its helicity (since helicity is proportional to $-[\theta]_{222}$) on pH at 3.1 °C (data are taken from ref. 11).

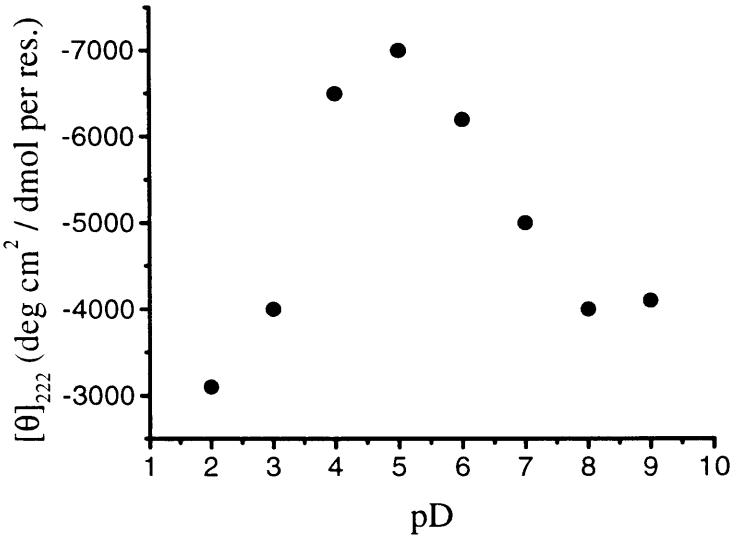
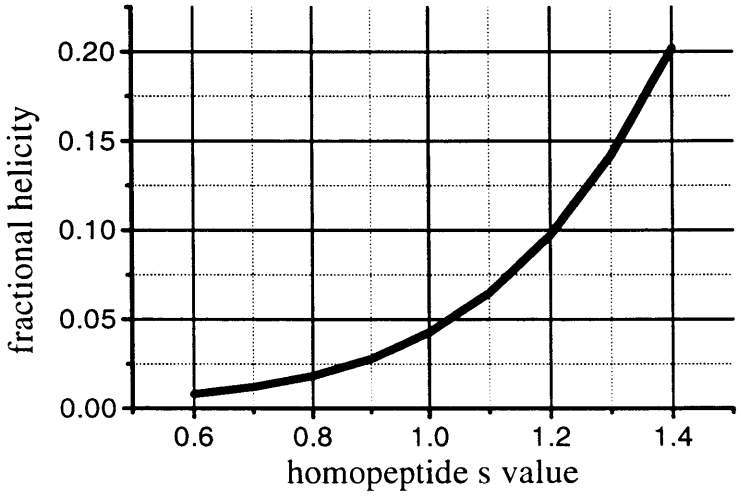


Table 1. The C-peptide's per residue molar ellipticities, $[\theta]_{222}$, at 3.1 °C at pHs 2, 5, and 9 and the corresponding ranges of fractional helicities. In the last two columns are the s values that would yield the same range of fractional helicities for a homopeptide of the same length. Data are taken from ref. 11.

pH	$[\theta]_{222,obs}$ (deg cm^2/dmol res)	fractional helicity calculated with $[\theta]_{222,100\%helix}$ =		homopeptide s equivalent	
		−42,000	−32,000	min	max
2	-3,100	0.074	0.097	1.1	1.2
5	-7,000	0.167	0.219	1.3	1.4
9	-4,000	0.095	0.125	1.2	1.25

Figure 3. The dependence of fractional helicity on s value for an 11 residue homopeptide.



The differences in the homopeptide s equivalents at different pHs are not overwhelmingly large, but it should be kept in mind that this is a per residue difference. The difference in the midpoints of the ranges of the homopeptide s equivalents at pHs 2 and 5, for example, corresponds to −RTln (1.35/1.15) or about −0.1 kcal/mol per residue.

For the 11 residue C-peptide this corresponds to −1.1 kcal/mol in stabilization over the whole peptide.

The dependence of the C-peptide's helicity on pH has probably been the most interpreted datum in attempts to identify helix stabilizing interactions as it indicates that the C-peptide's helicity is dependent on the ionization states of residue side chains. The low pH arm of the helicity vs. pH curve shows that deprotonation of either or both of the glutamic acid side chains to yield a negative charge at either or both of positions 2 and 9 favors helicity. Similarly, the high pH arm indicates that deprotonation of the histidine to yield a neutral side chain at position 12 disfavors helicity. This behavior could be explained in terms of changes in the intrinsic helix propensities of the ionizing residues. However, the differences in the s values of Glu^0 and Glu^- and His^+ and His^0 (see table 2 of chapter 2) are too small to explain the large changes in fractional helicity. In addition, the helicity should decrease upon deprotonation of Glu, whereas the helicity is seen to increase. A preferable explanation is the existence of interactions between residue side chains that result in strongly context dependent helix propensities. It has been suggested that this is the likely basis for all instances in which short peptides are substantially helical.[13,14,15]

The first attempt[11] to interpret the pH dependence of the C-peptide's helicity postulated a helix stabilizing charge-charge interaction between Glu 9^- and His 12^+. This was later proven to be incorrect, as the signature pH dependence of the C-peptide's helicity was shown to persist when Glu 9 was replaced by a leucine (in fact, the Glu 9 → Leu derivative was much more helical than the original C-peptide; this will be discussed

[13] Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein. Chem.* **1975**, *29*, 205.
[14] Vasquez, M.; Pincus, M. R.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 351.

further below).[16] This forced the conclusion that the pH dependence had to be due to Glu 2 and His 12, residues whose presence in the C-peptide was already known to be critical for the activation of RNase S-protein,[17] but these residues were much too far apart to be interacting with each other. Each one had to be interacting with some other residue.

The next proposal to explain the Glu 2^- and His 12^+ helicity effects invoked charge-dipole interactions between Glu 2^- and the N-terminal, positive pole of the helix dipole and His 12^+ and the C-terminal, negative pole of the helix dipole.[18] The existence of charge-dipole effects was supported with experiments in which the N-terminal charge of a C-peptide analog was varied from +2 (a lysine at the N-terminus) to -1 (a succinyl alanine at the N-terminus). The helicity of the analog increased continually as the charge was changed from positive to negative. Although this seemed to established a charge-dipole interaction as a major effect, the presence of other interactions in the C-peptide could not be ruled out.

It is currently believed that the contribution of Glu 2^- to the C-peptide's helicity is due to an interaction with Arg 10^+ that overrides its interaction with the helix dipole. This interaction at first seems improbable because of the distance between Glu 2^- and Arg 10^+. Most of the interactions known to affect helix propensities are i to i+3 or i+4 interactions (see chapter 2, section 2.6), whereas this would be an i to i+8 interaction. However, a great deal of evidence from a variety of sources supports the presence of this interaction. It is directly observed in the crystal structures of both RNase A[19,20] and RNase S.[9]

---

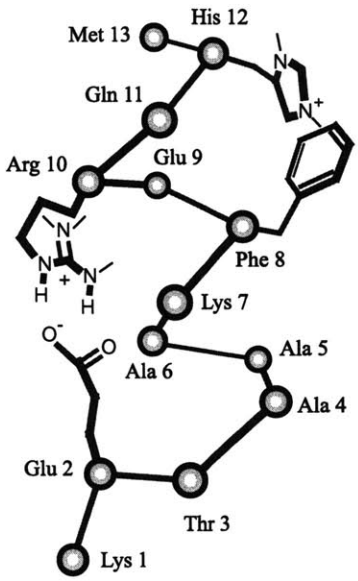
[15] Vasquez, M.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 41.
[16] Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York. E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 2349.
[17] Finn, F. M.; Hofmann, K. *J. Am. Chem. Soc.* **1965**, *87*, 645.
[18] Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563.
[19] Kartha, G.; Bello, J.; Harker, D. *Nature* **1967**, *213*, 862.
[20] Wlodawer, A.; Sjölin, L. *Biochemistry* **1983**, *22*, 2720.

Studies of the activation of the S-protein by synthetic C- and S-peptide analogs[21,22] show that the presence or absence of Glu 2^- only affects peptide-protein binding when Arg 10^+ is intact. NMR studies have shown that the δ methylene protons[23] and the ε NH proton[24] of Arg 10^+ shift significantly (more than 0.1 ppm) in the pH region where as Glu 2^- ionizes, and an NOE between the γ methylene of Glu 2^- and the main chain NH of Lys 7 confirms that the Glu 2^- side chain is oriented toward Arg 10^+.[25] In addition, CD studies show that replacing Arg 10^+ by Ala can eliminate the variation of the helicity of C-peptide analogs at low pH.[24]

---

[21] Hofmann, K.; Visser, J. P.; Finn, F. M. *J. Am. Chem. Soc.* **1970**, *92*, 2900.
[22] Marchiori, F.; Borin, G.; Moroder, L.; Rocchi, R.; Scoffone, E. *Biochim. Biophys. Acta* **1972**, *257*, 210.
[23] Rico, M.; Gallego, E.; Santoro, J.; Bermejo, F. J.; Nieto, J. L.; Herranz, J. *Biochem. Biophys. Res. Comm.* **1984**, *123*, 757.
[24] Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biophys. Chem.* **1990**, *37*, 107.
[25] Osterhout Jr., J. J.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1989**, *28*, 7059.

Figure 4. A trace of the C-peptide as it appears in the crystal structure of RNase A (taken from ref 20). Alpha carbons are shown as shaded circles. To emphasize their interactions, the only side-chains shown are those of Glu 2^-, Arg 10^+, Phe 8 and His 12^+.



It should be noted, though, that the crystal structures and one of the NMR studies[25] show that the i to i+8 interaction is not accommodated within the helix. Although the Glu 2^- / Arg 10^+ interaction favors helicity, the glutamic acid residue is held outside the helix by its involvement with the arginine, and helical ϕ and ψ torsions actually begin at Ala 4. The Glu 2^- / Arg 10^+ interaction is illustrated in the cartoon of the C-peptide helix in figure 4 taken from one of the crystal structure of ribonuclease A.[20]

The role of His 12^+ in stabilizing the C-peptide helix is currently believed to stem from an interaction with Phe 8. The evidence for this interaction is similar to that for the Glu 2^- / Arg 10^+ interaction. The Phe and His rings are close to each other in the crystal structure of RNase A, as shown in figure 4.[20] The aromatic protons of Phe 8 shift over the range in which the His 12 imidazole becomes deprotonated,[23] and the imidazole protons appear at an unusual chemical shift, consistent with their being influenced by Phe

8's ring currents.[26,27] Finally, CD studies show that the variation of the helicity of C-peptide analogs at high pH can be reduced (but not eliminated) by replacing Phe 8 with Ala.[28]

3.2.3 Oddities about the Helicity of the C-peptide and Its Analogues

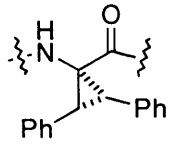
Despite the plentiful data on the helicity of the C-peptide and related peptides, there remain a number of remarkable observations for which there are no explanations. Some examples include the following. Changing Gln 11 and the homoserine lactone 13 of the C-peptide (KETAAAKFERQHHse lactone) both to alanine to yield the peptide KETAAAKFERAHA-NH_2 results in a large decrease in fractional helicity ($[\theta]_{222}$ = −3,000 deg cm^2 / dmol res at pH 5, 3 °C compared to $[\theta]_{222}$ = −7,100 deg cm^2 / dmol res for the C-peptide).[16] This is probably due solely to the Gln 11 → Ala change since replacing the C-terminal residue by Gly has very little effect on helicity.[29] Changing Glu 9 in KETAAAKFERAHA-NH_2 to Leu to yield KETAAAKFLRAHA-NH_2 increases the helicity by 200% ($[\theta]_{222}$ = −6,000 deg cm^2 /dmol res for the latter peptide).[16] This effect is similar in magnitude to that of the Glu 2^- / Arg 10^+ interaction (where $[\theta]_{222}$ changes from −3,100 to −7,000 deg cm^2 / dmol res). Although the Gln 11 → Ala substitution above resulted in a decrease in helicity, it is not always the case that replacing residues

---

[26] Rico, M.; Bermejo, F. J.; Santoro, J.; Nieto, J. L.; Gallego, E.; Herranz, J.; Voskuyl-Holtkamp, I.; Schattenkerk, C. *Int. J. Pept. Prot. Res.* **1987**, *29*, 193.
[27] Dadlez, M.; Bierzynski, A.; Godzik, A.; Sobocinska, M.; Kupryszewski, G. *Biophys. Chem.* **1988**, *31*, 175.
[28] Shoemaker, K. R.; Fairman, R.; Schultz, D. A.; Robertson, A. D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1990**, *29*, 1
[29] Strehlow, K. G.; Baldwin, R. L. *Biochemistry* **1989**, *28*, 2130.

with alanine decreases helicity. Changing Arg 10^+ in the C-peptide analog AcAETAAAKFLRAHA-NH_2 to Ala to yield AcAETAAAKFLAAHA-NH_2 results in an increase in helicity at low pH ($[\theta]_{222} = -10{,}000$ deg cm^2 / dmol res for the former peptide and $-13{,}000$ deg cm^2 / dmol res for the latter) and the same helicity at medium pH, where the former peptide is stabilized by the Glu 2^- / Arg 10^+ interactions but the latter is not ($[\theta]_{222} = -13{,}000$ deg cm^2 / dmol res for both peptides).[24] Thus, while a Gln 11 → Ala substitution decreased helicity, an Arg 10^+ → Ala substitution increased helicity, and to the same extent as the Glu 2^- / Arg 10^+ interaction. Finally, replacing Phe 8 in the sequence succinyl-AETAAAKFLRAHA-NH_2 with the conformationally constrained cyclopropyl amino acid pictured in figure 5 to yield succinyl-AETAAAKXLRAHA-NH2, where X is the constrained residue, almost completely abolishes the temperature and pH dependence of the peptide's helicity.[30] Apparently neither the Glu 2^- / Arg 10^+ interaction nor the X 8 / His 12^+ interaction (if it occurs) contribute to the helicity of this peptide.

Figure 5. Cyclopropyl amino acid used to replace Phe 8.



[30] Moye-Sherman, D.; Jin, S.; Ham, I.; Lim, D.; Scholtz, J. M.; Burgess, K. *J. Am. Chem. Soc.* **1998**, *120*, 9435.

3.3 The Divided Peptide Method Applied to the Ribonuclease C-peptide: a Feasibility Study

The C-peptide is an ideal system in which to test the reductionist approach to studies of peptide helicity in general, and the divided peptide method specifically. Enough of the forces that influence helicity is understood to guide experimentation, but there are still many effects that further studies could contribute to explaining. First, however, a feasibility study is needed to show the possibilities and limitations of the method.

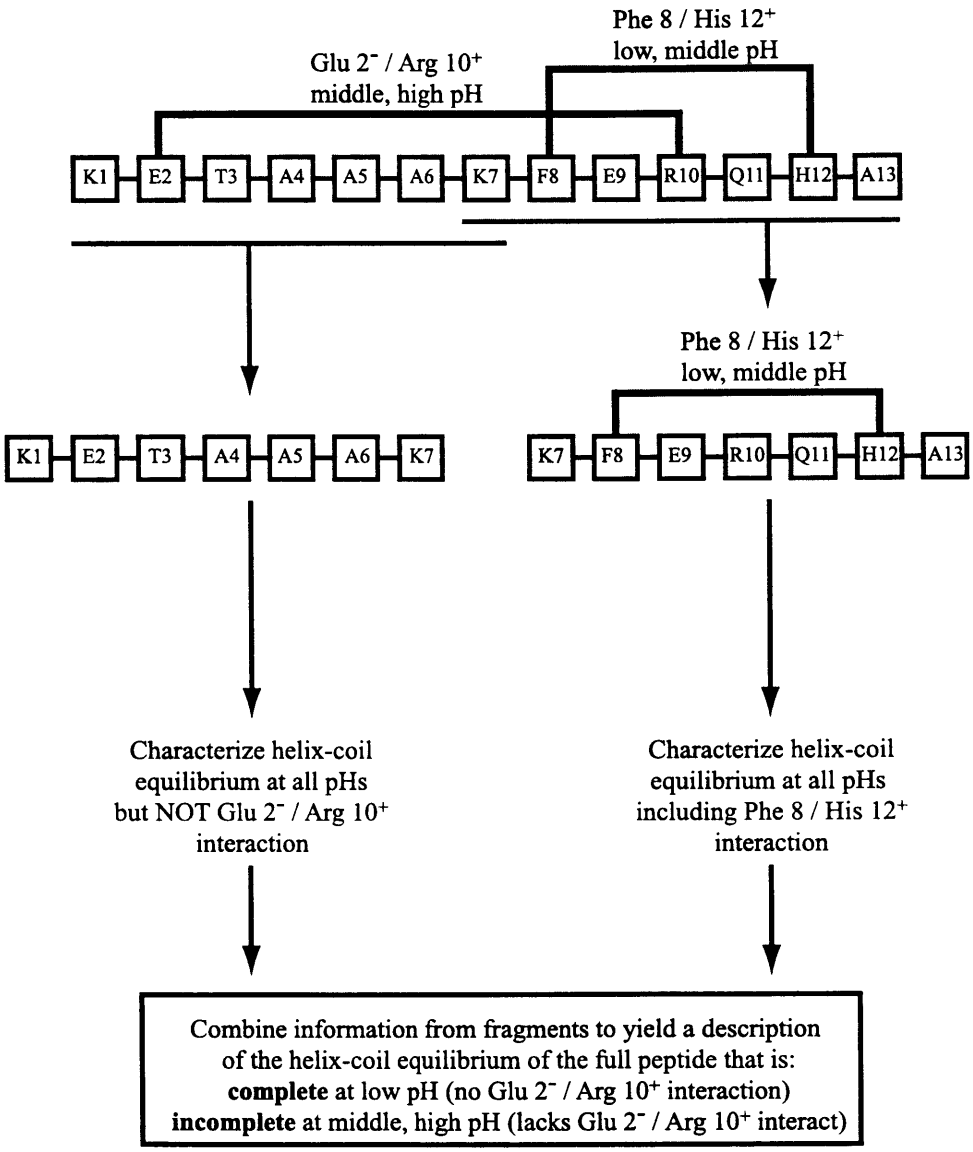
The divided peptide method requires that the C-peptide sequence be split into two or more fragments. In this preliminary study, the C-peptide sequence is disjoined near the middle to yield two fragments, KETAAAK and KFERQHA (henceforth the left- and right-hand fragments, respectively). Note that the second lysine (Lys 7, where the peptide is numbered from left to right, starting at one) overlaps between the fragments so that its effects, if different, on the helicities of both fragments can be assessed. In addition, the homoserine at the C-terminus of the C-peptide has been replaced by alanine in the right-hand fragment KFERQH<u>A</u>. This simple replacement should have a minimal impact on the peptide's helicity, as replacement of the C-terminal residue by glycine is known not to affect the helicity of C-peptide analogs.[29] The C-peptide analogs in which Hse 13 is replaced by Ala will be denoted ^{A13}C-peptide.

One of the most striking features of the C-peptide's helicity is the pH dependence believed to be caused by the Glu 2^- / Arg 10^+ and Phe 8 / His 12^+ interactions. Any attempt to describe the C-peptide's helicity must certainly include this phenomenon in order to be complete. The conformational tendencies of the fragments must therefore be

studied at several pH values; actually, in practice the helicities have to be compared at several pD values since all experiments will be done by NMR in D_2O (see below). In this work, measurements are made at pD 2.3, 5.1, 7.5, and 9.8 so that at the first pD, all of the side chains are protonated; at the second pD, the Glu residues are at about the mid-point of their titration; at the third pD, the His residue is at its mid-point; and at the last pD, all of the ionizable side chains except for Lys and Arg are fully deprotonated.

The application of the divided peptide method to the ^{A13}C-peptide as described thus far is shown schematically in figure 6.

Figure 6. The divided peptide method applied to the Hse 13 → Ala analog of the C-peptide, the ^{A13}C-peptide. The peptide is divided into left- and right-hand fragments. Although the right-hand fragment contains the Phe 8 / His 12^+ interaction, neither fragment contains the Glu 2^- / Arg 10^+ interaction. Consequently, the method should only provide a complete description of the ^{A13}C-peptide's helicity at low pH, where the interaction does not occur.



As shown in figure 6, it is expected that the effects of the Phe 8 / His 12^+ interaction will be observed in the KFERQHA fragment since this fragment contains both of the

relevant residues. However, the Glu 2^- / Arg 10^+ interaction has been broken up between the two fragments, and consequently it cannot be accounted for by this method. Because of this, the approach can yield a complete description of the ^{A13}C-peptide's helicity at low pD when this interaction is not active. At pDs 5.1, 7.5, and 9.8 the combined data from the two fragments should underestimate helicity in a way directly relatable to the strength of the Glu 2^- / Arg 10^+ interaction. It should be noted that a method for characterizing the helix-coil equilibria of the left- and right-hand fragments has not yet been introduced. Given the Zimm-Bragg model for the helix-coil equilibrium introduced in chapter 2, such a characterization requires that the helix propensities (s values) for each residue in both peptides and the helix initiation constant (σ) be determined. If the value given for σ in chapter 2 (2×10^{-3}) is used, then only the s values are left to be measured.

Determining the helix propensities in the peptide fragments would normally be impossible because they are too short to exhibit significant helicity on their own. As discussed in chapter 2 section 2.4, in order to study their helical states they must be conjugated to a helix-inducing template. The template $AcHel_1$ serves in this case as both a template and a means to quantify the helicity of the attached peptides, as detailed below.

3.3.1 The Residue Exclusion Method for Measuring Helix Propensities Using $AcHel_1$-peptide Conjugates

The formula for the t/c ratio of an n residue $AcHel_1$-peptide conjugate determined in chapter 2 (equation 2-9) is reintroduced here:

$$\left(\frac{t}{c}\right) = A + B + B(s_1 + s_1 s_2 + \cdots + s_1 s_2 \cdots s_n) \qquad \boxed{3\text{-}1}$$

Numerous ways to use the t/c ratio to obtain s values suggest themselves upon inspecting the expression for the t/c ratio given above. Thus far, s values for alanine and lysine have been determined by finding the values of s_A and s_K that best fit blocks of t/c ratio data from many $AcHel_1$-peptide derivatives where the peptide portion has only alanine and lysine residues.[31,32,33] The s values for all 20 naturally occurring and a few non-naturally occurring amino acids have also been determined by a host-guest procedure, in which the residue whose s value is of interest is inserted as a guest into an alanine-lysine host peptide conjugated to $AcHel_1$.[34] Unfortunately, both of these methods required that the residues whose s values were of interest be studied in very homogeneous, alanine rich contexts. Neither of these methods can be used to obtain the information needed for this study, that is, the pD dependent s values of residues within the heterogeneous environment of the ^{A13}C-peptide fragments. A different means, described below, for determining s values from t/c ratios has been used for this purpose.

Take two C-terminally amidated peptides conjugated to the template $AcHel_1$, both with an alanine spacer to eliminate the possibility of unusual side-chain-template interactions:

Peptide 1: $AcHel_1$-A-X1-X2-X3-X4-X5-X6-NH_2

Peptide 2: $AcHel_1$-A-X2-X3-X4-X5-X6-NH_2

---

[31] Kemp, D. S.; Oslick, S. L.; Allen, T. J. *J. Am. Chem. Soc.* **1996**, *118*, 4249.
[32] Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4025.
[33] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 12234.

The two peptides differ only in that the residue X1 has been excluded in peptide 2 from its position in peptide 1 between the alanine and X2. According to equation 3-1, the t/c ratios of these peptides are

$$\left(\frac{t}{c}\right)_1 = A + B + Bs_A + Bs_A(s_{X1} + s_{X1}s_{X2} + s_{X1}s_{X2}s_{X3} + \cdots + s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6})$$

$$\left(\frac{t}{c}\right)_2 = A + B + Bs_A + Bs_A(s_{X2} + s_{X2}s_{X3} + \cdots + s_{X2}s_{X3}s_{X4}s_{X5}s_{X6})$$

The formulas above for $(t/c)_1$ and $(t/c)_2$ can be rearranged to give

$$\left[\frac{\left(\frac{t}{c}\right)_1 - A - B - Bs_A}{Bs_A}\right] = Q_1 = (s_{X1} + s_{X1}s_{X2} + s_{X1}s_{X2}s_{X3} \cdots + s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6})$$

$$\left[\frac{\left(\frac{t}{c}\right)_2 - A - B - Bs_A}{Bs_A}\right] = Q_2 = (s_{X2} + s_{X2}s_{X3} + \cdots + s_{X2}s_{X3}s_{X4}s_{X5}s_{X6})$$

In general, the $Q_{peptide}$ values for any $AcHel_1$-peptide conjugate that has an alanine at the template-peptide junction and n residues can be expressed in terms of the peptide's t/c ratio as follows

$$\frac{\left(\frac{t}{c}\right)_{peptide} - A - B - Bs_A}{Bs_A} = Q_{peptide} = s_1 + s_1s_2 + s_1s_2s_3 + \cdots + s_1s_2s_3 \cdots s_n$$

---

[34] Allen, T. J. Thesis, Massachusetts Institute of Technology, **1993**.

Using the known values given in chapter 2 for the template constants A = 0.832 and B = 0.156 and using[31] $s_A = 1.02$, $Q_{peptide}$ is

$$\frac{\left(\frac{t}{c}\right)_{peptide} - 1.15}{0.159} = Q_{peptide} \qquad \boxed{3-2}$$

Returning to the case of the two peptides $AcHel_1A$-X1X2X3X4X5X6-NH_2 and $AcHel_1A$-X2X3X4X5X6-NH_2, one notices that their Q_1 and Q_2 values are very similar except that s_{X1} does not appear in Q_2. In fact,

$$Q_1 = s_{X1}(1+Q_2)$$

Given this relationship between Q_1 and Q_2, s_{X1} is simply

$$s_{X1} = \frac{Q_1}{1+Q_2} \qquad \boxed{3-3}$$

This result can be generalized. All that is necessary to find the s value of the first residue in any peptide is to know the t/c ratio for the peptide attached to $AcHel_1$ (with an alanine spacer between the template and the peptide) and the t/c ratio for the analogous peptide in which the first residue has been excluded. The method can even be extended in an obvious way to obtain the s values of all of the residues in a peptide. For the example given above in which the peptide of interest was X1-X2-X3-X4-X5-X6, s_{X1} was obtained from the t/c ratios of $AcHel_1A$-X1-X2-X3-X4-X5-X6-NH_2 and $AcHel_1A$-X2-X3-X4-X5-X6-NH_2. The s value of X2 could be obtained similarly from the t/c ratios of $AcHel_1A$-X2-X3-X4-X5-X6-NH_2 and $AcHel_1A$-X3-X4-X5-X6-NH_2, that of X3 could be obtained from the t/c ratios of $AcHel_1A$-X3-X4-X5-X6-NH_2 and $AcHel_1A$-X4-X5-X6-NH_2, and so

on until all of the s values were known. Thus, the residue exclusion method can be used to obtain the s values for all of the residues in any peptide fragment, just as required by the divided peptide method for studying peptide helicity. It can provide a complete accounting for the helix-coil equilibrium within the fragment.

The peptides that are needed in order to determine the unknown s values (those other than s_A) of the residues in the left-hand fragment of the ^{A13}C-peptide are shown in table 2.

Table 2. $AcHel_1$-peptides from the left-hand fragment.

Peptides from first fragment	s value determined between peptides
$AcHel_1A$-KETAAAKA-NH_2	
	s_{K1}
$AcHel_1A$-ETAAAKA-NH_2	
	s_{E2}
$AcHel_1A$-TAAAKA-NH_2	
	s_T
$AcHel_1A$-AAAKA-NH_2	

The extra alanine is added at the C-terminus of the peptides to avoid any possible effects that are not present in the ^{A13}C-peptide itself due to having a lysine positioned directly at the C-terminus.

The peptides needed in order to determine the unknown s values of the residues in the right-hand fragment are shown below in table 3.

Table 3. $AcHel_1$-peptides from the right hand fragment.

Peptides from second fragment	s value determined between peptides
$AcHel_1A$-KFERQHA-NH_2	
	s_{K7}
$AcHel_1A$-FERQHA-NH_2	
	s_F
$AcHel_1A$-ERQHA-NH_2	
	s_{E9}
$AcHel_1A$-RQHA-NH_2	
	s_R
$AcHel_1A$-QHA-NH_2	
	s_Q
$AcHel_1A$-HA-NH_2	
	s_H
$AcHel_1A$-A-NH_2	

Extra alanines are not added at the ends of these peptides because their C-termini represent the actual C-terminus of the ^{A13}C-peptide, so there is no worry of introducing any end effects that are not already present ^{A13}C-peptide itself. Note that the t/c ratio of $AcHel_1A$-A-NH_2 need not be physically measured. Given that A = 0.832, B = 0.156, and s_A = 1.02, it can be computed as follows

$$\left(\frac{t}{c}\right)_{AA} = A + B + Bs_A + Bs_A^2 = 0.832 + 0.156 + 0.159 + 0.162 = 1.31$$

Because the template constants and s_A should all be invariant with respect to pD, this t/c ratio can be used at all pDs.

Two s values for Lys 7, which is shared between the two fragments, can be determined using the peptides listed above. The first s_{K7} can be found by the residue exclusion method from the peptides in table 3. The second can be determined directly from the t/c ratio of $AcHel_1A$-AAAKA-NH_2, which is:

$$\left(\frac{t}{c}\right)_{AAAAKA} = A + B + Bs_A + Bs_A(s_A + s_A^2 + s_A^3 + s_A^3 s_{K7} + s_A^4 s_{K7})$$

Since A, B and s_A are all known and the t/c ratio can be measured, s_{K7} is the only unknown in this equation and can therefore be solved for in a straightforward way. If the s value for lysine were not dependent on the context in which it appeared, then the s values found by the two different methods would be the same. However, it is known that the s value of lysine depends on its position relative to the template so one should expect the two determinations of s_{K7} to yield different results.

3.3.2 Properties of s Values Determined by the Residue Exclusion Method

In the absence of any interactions that would cause helix propensities to be context dependent, the s values that are determined by the residue exclusion method are exactly those that would be needed for the divided peptide approach. However, it is prudent to model what happens to these s values when they are context dependent. The effects of two types of context dependence will be addressed below: helix-stabilizing interactions between residues and position dependence of s values.

3.3.2a The Residue Exclusion Method Applied to a Peptide in which a Helix Stabilizing Interaction Between Residues Occurs

For a case that incorporates the first type of context dependence, take the peptide X1-AAA-X2-A where a helix-stabilizing interaction occurs between X1 and X2. To determine the s values of the residue X1 by the residue exclusion method, the t/c ratios of $AcHel_1A$-X1AAAX2A-NH_2 and $AcHel_1A$-AAAX2A-NH_2 are needed. (The s value of X2 could, if desired, be determined directly from the t/c ratio of the second peptide in the same way that s_{K7} can be determined from $AcHel_1A$-AAAKA-NH_2.) Say that the intrinsic value of s_{X1} is 1.1 and that the intrinsic value of s_{X2} is 1.2, but when X1 and X2 are in a helix together they interact in such a way as to stabilize the helix by 0.5 kcal/mol. Then in the peptide $AcHel_1A$-X1AAAX2A-NH_2, the states in which the helix extends from the template-peptide junction to the X2 residue are stabilized by 0.5 kcal/mol. This results in X2 having an effective s value of $s_{X2} \times e^{-(-0.5\text{kcal/mol})/RT} = 2.8$ at 25 °C. Note that the stabilization energy cannot equivalently be assigned to the s value of X1. Since all helices in peptides attached to $AcHel_1$ initiate at the template, there are many states in which the helix extends through X1 but not all the way to X2. In these states the stabilizing interaction between X1 and X2 does not exist and it would be wrong to use a value of s_{X1} that included a −0.5 kcal/mol contribution from the interaction. In order to calculate the t/c ratio of $AcHel_1A$-X1AAAX2A-NH_2 two values of s_{X1} would be needed- one for the states in which just X1 is helical and one for the states in which both X1 and X2 are helical. In contrast, in all of the states in which the helix extends through X2, X1 is also helical and the stabilizing interaction between X1 and X2 is present. Because X2

cannot be helical without X1 also being helical, a single value for s_{X2} that includes the contribution from the X1-X2 interaction can be used in the calculation of the t/c ratio of $AcHel_1A$-X1AAAX2A-NH_2. So, using A = 0.832, B = 0.156, s_A = 1.02, s_{X1} = 1.1 and s_{X2} = 2.8, the t/c ratio of $AcHel_1A$-X1AAAX2A-NH_2 is:

$$\left(\frac{t}{c}\right)_{X1AAAX2A} = A + B + Bs_A + Bs_A(s_{X1} + s_{X1}s_A + s_{X1}s_A^2 + s_{X1}s_A^3 + s_{X1}s_A^3 s_{X2} + s_{X1}s_A^4 s_{X2})$$

$$= 2.91$$

The residue X1 has been excluded from the peptide $AcHel_1A$-AAAX2A-NH_2 so there can be no X1-X2 interaction, and the intrinsic value of s_{X2} = 1.2 should be used for calculating its t/c ratio:

$$\left(\frac{t}{c}\right)_{AAAX2A} = A + B + Bs_A + Bs_A(s_A + s_A^2 + s_A^3 + s_A^3 s_{X2} + s_A^4 s_{X2})$$

$$= 2.05$$

Of course, if one were studying these peptides experimentally, one would not have any of the information detailed above about the intrinsic values of s_{X1} and s_{X2} and the energetics of the X1-X2 interaction. One would only have experimentally determined t/c ratios for $AcHel_1A$-X1AAAX2A-NH_2 and $AcHel_1A$-AAAX2A-NH_2, which should be close to the values computed above (2.91 and 2.05 respectively). These would be used to find the s value of X1 by the residue exclusion method using the Q values for these two peptides. Recall that $Q_{peptide}$ values for any $AcHel_1$-peptide conjugate are given by

$$\left[\frac{\left(\frac{t}{c}\right)-1.15}{0.159}\right]=Q_{peptide}$$

The $Q_{peptide}$ values for $AcHel_1A\text{-}X1AAAX2A\text{-}NH_2$ and $AcHel_1A\text{-}AAAX2A\text{-}NH_2$, $Q_{X1AAAX2A}$ and Q_{AAAX2A} respectively, can be used to find s_{X1} as previously described:

$$s_{X1}=\frac{Q_{X1AAAX2A}}{1+Q_{AAAX2A}}=\frac{11.14}{1+5.69}=1.66$$

Note that the value of s_{X1} determined by the residue exclusion method is substantially larger than the intrinsic value of s_{X1} used to compute the t/c ratio. Furthermore, if one were to make the necessary peptides to determine s_{X2} by the residue exclusion method, or if one were to determine s_{X2} directly from the t/c ratio of $AcHel_1A\text{-}AAAX2A\text{-}NH_2$, one would find that s_{X2} = 1.2, the intrinsic s value for X2.

The residue exclusion method assigns the energy of the X1-X2 interaction to s_{X1} instead of s_{X2}, the reverse of the way in which it was assigned above. Also, the difference between the intrinsic value of s_{X1} (1.1) and the measured value of s_{X1} (1.66) corresponds to only −0.24 kcal/mol (at 25 °C), somewhat less than the true energy of the interaction. It happens, however, that one obtains the same t/c ratio for the peptide $AcHel_1A\text{-}X1AAAX2A\text{-}NH_2$ whether one uses the s values s_{X1} = 1.1 and s_{X2} = 2.8 or the s values s_{X1} = 1.66 and s_{X2} = 1.2. While the former set of s values is the more correct because the energy that corresponds to the difference between s_{X2} = 2.8 and s_{X2} = 1.2 is the true energy of the X1-X2 interaction, the latter set still yields the correct t/c ratio. One can conclude that the residue exclusion method in fact can yield serviceable, if not

rigorously correct, helix propensities even in the face of fairly strong context dependencies.

It should be noted that, unless it is independently known that the intrinsic s_{X1} was actually 1.1, one would not be able to tell that the value s_{X1} = 1.66 calculated above by the residue exclusion method is not itself the intrinsic value of s_{X1}. One would need more information in order to identify the large measured value of s_{X1} as being due to an interaction. One could obtain such information, for example, by measuring the s value of X1 by the residue exclusion method in the absence of the residue X2 using the peptides $AcHel_1A$-X1AAAAA-NH_2 and $AcHel_1A$-AAAAA-NH_2. Alternatively, one could discern whether there was an interaction and avoid synthesizing more peptides using the pD dependence of s_{X1}. If X2 had an ionizable side chain and X1 did not (or had a side chain that ionized over a completely different pD range), and the X1-X2 interaction depended on the protonation state of X2, then s_{X1} would change as X2 changes its protonation state. For example, say the above value for s_{X1} was determined at low pD, where X2 is protonated, and say that at high pD X2 is deprotonated and the interaction no longer occurs. Also, let the intrinsic s_{X1} retain its low pD value of 1.1 at high pD, which is very likely since none of the properties of X1 change between low and high pD. The intrinsic s_{X2} would probably change as X2 changed its ionization state, but for the sake of simplicity let the intrinsic s_{X2} also retain its low pD value of 1.2 high pD. The t/c ratio of $AcHel_1A$-X1AAAX2A-NH_2 would then be 2.32 and $AcHel_1A$-AAAX2A-NH_2 would be 2.05. From these two values one finds that

$$s_{X1}^{\text{high pD}} = \frac{Q_{X1AAAX2A}^{\text{high pD}}}{1 + Q_{AAAX2A}^{\text{high pD}}} = \frac{7.36}{1 + 5.69} = 1.1$$

At high pD one obtains s_{X1} = 1.1 (its intrinsic value), compared to 1.66 at low pD. Since it was stated that only the protonation state of X2 changes between high and low pD, it can be concluded that the difference between the high and low pD values of s_{X1} can be attributed to an interaction between X1 and X2.

3.3.2b The Residue Exclusion Method Applied to a Peptide in Which One of the s Values is Position Dependent

For a case that incorporates the second type of context dependence, compare the peptides $AcHel_1A$-AAAKA-NH_2, which has an experimentally determined t/c ratio of 2.63 (in D_2O at pD 2.3 and 5 °C) and $AcHel_1A$-AAKA-NH_2, which has a t/c ratio of 1.99 (also in D_2O at pD 2.3 and 5 °C). (Note that these t/c ratios are experimentally measured, unlike the t/c ratios in the previous section which were calculated based on arbitrarily assigned s values for X1 and X2.) Assuming the usual values of A, B, and s_A, the s_K values in both peptides can be determined directly from the t/c ratios as discussed in section 3.3.1. For the first peptide,

$$\left(\frac{t}{c}\right)_{AAAAKA} = A + B + Bs_A + Bs_A(s_A + s_A^2 + s_A^3 + s_A^3 s_K + s_A^4 s_K)$$

$$2.63 = 1.64 + 0.34 \times s_K$$

$$s_K = 2.91$$

while for the second peptide

$$\left(\frac{t}{c}\right)_{AAAKA} = A + B + Bs_A + Bs_A(s_A + s_A^2 + s_A^2 s_K + s_A^3 s_K)$$

$$1.99 = 1.47 + 0.33 \times s_K$$

$$s_K = 1.58$$

The strong dependence of s_K on the number of residues between the lysine and the helix initiation site at the template is evident from the computations above. This position dependence, in which the helix propensity of lysine increases until it is five residues away from the template and thereafter remains constant, has also been demonstrated by analysis of a large database of t/c ratio data of $AcHel_1$-peptide conjugates in which the peptide is composed only of alanine and lysine residues.[32]

The t/c ratios for $AcHel_1A$-AAAKA-NH_2 and $AcHel_1A$-AAKA-NH_2 can also be used to determine the s value of alanine by the residue exclusion method, since the first peptide differs from the second only by the exclusion of an alanine residue:

$$s_A = \frac{Q_{AAAAKA}}{1 + Q_{AAAKA}} = \frac{9.36}{1 + 5.38} = 1.47$$

This rather large value for s_A shows how the length dependence of the lysine s value influences s values determined by the residue exclusion method. When a peptide contains a residue whose s value is larger when the residue is further from the helix initiation site, the s values determined by the residue exclusion method for all the residues that precede it will be artificially high. The reverse is true when a peptide contains a residue whose s value is smaller when the residue is further from the helix initiation site; the s values determined by the residue exclusion method for all the residues that precede it will be artificially low. This is again a situation in which context

dependence causes the residue exclusion method to yield s values that are not exactly correct, but that faithfully reflect the underlying effects.

One can conclude that the s value found here for alanine by the residue exclusion method contains a contribution due to the length dependence of lysine's s value because alanine's intrinsic s value is already known and is available for comparison. If this were not the case though, it would again be impossible to know just from this determination of s_A by the residue exclusion method that 1.47 was not the intrinsic value for s_A. As before, more information would be needed from, for example, other $AcHel_1$-peptide conjugates or from pD studies in order to know whether or not the measured s value was due to context dependence.

3.3.2c Distinguishing Effects Due to Residue-Residue Interactions from Those Due to Length Dependence in s Value Determined by the Residue Exclusion Method

Residue-residue interactions and position dependencies have the same effect on s values determined by the residue exclusion method. Both types of context dependence result in residues' measured s values deviating from their intrinsic s values according to whether the context dependence results in a stabilized or a destabilized helical state. Given just one residue whose s value indicates the existence of context dependence it would be impossible to tell what kind of context dependence was operating. However, in the case of a residue-residue interaction only the s value of the residue directly involved in the interaction would show the context dependence. In contrast, in the case of position dependence the s values of all of the residues that precede the residue with the position

dependent helix propensity would show the context dependence. As long as the s values of more than one residue in the peptide are measured it should be possible to distinguish the types of context dependencies.

3.3.3 Testing the Residue Exclusion Method for Determining s Values and the Divided Peptide Method for Determining Helix-Coil Equilibria

In the residue exclusion method, N-terminal residues are excluded one by one from $AcHel_1A$-peptide conjugates and their s values are determined between pairs of t/c ratios. If one excludes a C-terminal residue, one obtains a different kind of information. Consider two peptides

Peptide 1: $AcHel_1$-A-X1-X2-X3-X4-X5-X6-NH_2

Peptide 2: $AcHel_1$-A-X1-X2-X3-X4-X5-NH_2

According to equation 3-2, Q_1 and Q_2 for these two peptides are

$$\left[\frac{\left(\frac{t}{c}\right)_1 - 1.15}{0.159}\right] = Q_1 = (s_{X1} + s_{X1}s_{X2} + \cdots + s_{X1}s_{X2}s_{X3}s_{X4}s_{X5} + s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6})$$

$$\left[\frac{\left(\frac{t}{c}\right)_2 - 1.15}{0.159}\right] = Q_2 = (s_{X1} + s_{X1}s_{X2} + \cdots + s_{X1}s_{X2}s_{X3}s_{X4}s_{X5})$$

Q_1 and Q_2 are the same except for the term $s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6}$ that is present in Q_1 but not Q_2. The difference $Q_1 - Q_2$ is therefore equal to $s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6}$. To test a set of s values s_{X1}, s_{X2}, s_{X3}, s_{X4}, s_{X5}, and s_{X6} determined by the residue exclusion method, the products $s_{X1}s_{X2}s_{X3}s_{X4}s_{X5}s_{X6}$ calculated from these s values and from $Q_1 - Q_2$ can be compared. This procedure for testing the s values produced by the residue exclusion method will not be used in this preliminary work. Instead, a test is used that enables simultaneous evaluation of the s values produced by the residue exclusion method and the utility of the divided peptide method for studying peptide helicity.

The ultimate test of the divided peptide method is whether an accurate description of the helix-coil equilibrium of the peptide of interest can be recovered from the characterization of the helix-coil equilibria of the separate fragments. For this study in particular, this means that the s values obtained from the fragments KETAAAK and KFERQHA have to be used to predict the helicity of the full ^{A13}C-peptide. How this can be done is described in detail in the following sections.

3.3.3a Comparing the Observed Helicity of the Full ^{A13}C-peptide to the Helicity Calculated Using the s Values Determined by the Residue Exclusion Method in the ^{A13}C-peptide Fragments

The calculated and observed helicities of the full ^{A13}C-peptide (KETAAAKFERQHA) can be compared by two methods. First, the fractional helicity of the free ^{A13}C-peptide can be calculated from the Zimm-Bragg model using the s values measured by the residue exclusion method (and assuming $\sigma = 2\times10^{-3}$) and compared to its fractional helicity as ascertained by CD spectroscopy. Second, the t/c ratio that would be expected from having the full ^{A13}C-peptide attached to $AcHel_1$ can be calculated using the measured s values and compared to its experimentally determined t/c ratio. The two peptides used for testing the divided peptide method are H-AKETAAAKFERQHA-NH_2 (for the CD spectroscopy method) and $AcHel_1$A-KETAAAKFERQHA-NH_2 (for the t/c ratio method). These are slightly different from the ^{A13}C-peptide. There is an extra alanine at the N-terminus to separate the peptide from the template-peptide junction in the $AcHel_1$-peptide conjugate and from the positive charge of the N-terminus in the free peptide (an alanine is used instead of an acetyl group because the latter is a strong N-cap). Both of the above methods for comparison are used in this work, but before examining any data derived from them it must be decided whether they should be expected to provide comparable results.

3.3.3b The Correspondence Between the t/c ratios of $AcHel_1A$-KETAAAKFERQHA-NH_2 and the fractional helicity of H-AKETAAAKFERQHA-NH_2

Two scenarios can be envisioned for the correspondence between the pD dependent t/c ratio of $AcHel_1A$-KETAAAKFERQHA-NH_2 and the pD dependent fractional helicity of H-AKETAAAKFERQHA-NH_2. In the first scenario the Glu 2^- / Arg 10^+ interaction in the ^{A13}C-peptide allows propagation of helices to the N-terminus. In this case, the pD dependence of $AcHel_1A$-KETAAAKFERQHA-NH_2's t/c ratio would mirror the pD dependence H-AKETAAAKFERQHA-NH_2's CD derived fractional helicity. As for the comparison of calculated to observed helicities, since the s values derived from the fragments of the ^{A13}C-peptide exclude the effect of the Glu 2^- / Arg 10^+ interaction, the calculated t/c ratios and fractional helicities would only match the observed data at low pD where the interaction does not occur. In the second scenario, the Glu 2^- / Arg 10^+ interaction requires a severe enough distortion to block the helices in which it occurs from extending to the N-terminus. In this case, the pD dependence of $AcHel_1A$-KETAAAKFERQHA-NH_2's t/c ratio would be different from the pD dependence of H-AKETAAAKFERQHA-NH_2's CD derived fractional helicity. The calculated and observed t/c ratios would match each other over the whole pD range since neither the t/c ratios nor the s values determined from the ^{A13}C-peptide fragments would be influenced by the Glu 2^- / Arg 10^+ interaction. However, the calculated and observed fractional helicities of H-AKETAAAKFERQHA-NH_2 would, as before, only match each other at low pD. It should be added that, although the t/c ratios of $AcHel_1A$-KETAAAKFERQHA-NH_2 would not show the influence of the Glu 2^- / Arg 10^+

interaction, the CD spectrum of this template-peptide conjugate should, because $[\theta]_{222}$ is proportional to total helicity. Although the template would not sense the Glu 2^- / Arg 10^+ interaction, it should not prevent it from occurring, and it could still substantially affect the total helicity of the template-peptide conjugate.

Is it possible to tell which of these scenarios is the more likely? Based on the crystal structure of RNase A and solution NMR studies of a C-peptide analog, the peptide must obtain a conformation in which the N-terminus is not helical in order to engage in the Glu 2^- / Arg 10^+ interaction (see figure 4). Given these results, the scenario in which the t/c ratio of $AcHel_1A$-KETAAAKFERQHA-NH_2 is blind to the Glu 2^- / Arg 10^+ is much more likely.

3.4 Results

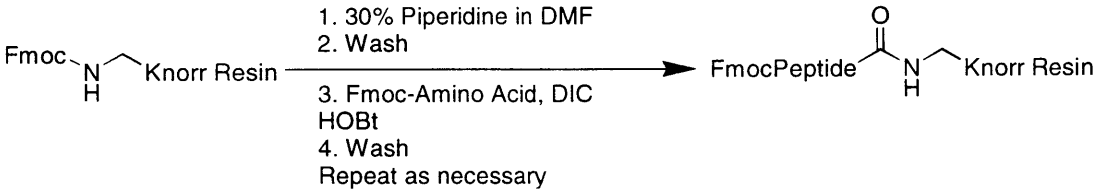
3.4.1 Synthesis

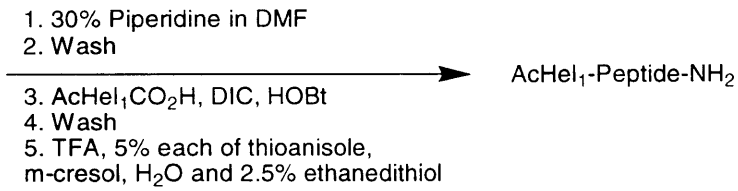
The $AcHel_1$-peptide conjugates were synthesized as C-terminal amides using Fmoc based solid-phase peptide synthesis protocols that have been previously described.[35] A typical procedure starting from Knorr resin[36] (Fmoc-2,4-dimethoxy-4'-(carboxymethyloxy)-benzhydrylamine linked to cross-linked polystyrene) is shown in scheme 1 and detailed in the experimental section. All peptides were characterized by ^{1}H-NMR and mass spectroscopy. Several of the necessary peptides ($AcHel_1A$-

[35] Jones, J. *The Chemical Synthesis of Peptides*; Oxford University Press: Oxford, **1991**.
[36] Bernatowicz, M. S.; Daniels, S. B.; Koster, H. *Tet. Lett.* **1989**, *30*, 4645.

KETAAAKA-NH_2, AcHel_1A-ETAAAKA-NH_2, and AcHel_1A-TAAAKA-NH_2) had been prepared prior to the work in this thesis.[37]

Scheme 1. (DIC = 1,3-Diisopropylcarbodiimide, HOBt = 1-Hydroxybenzotriazole)





3.4.2 General Notes

The C-peptide's helicity is maximal at low temperature and drops off dramatically as the temperature increases. All of the measurements in this chapter were therefore made at 5 °C. In addition, because the vast majority of experiments involved ^{1}H-NMR, any buffers used had to be deuterated. Experiments were performed at the required pDs (2.3, 5.1, 7.5, and 9.4) using a universal buffer, a single buffer with buffering capacity over a large range of pDs, rather than a series of different buffers so that the environment of the peptides would be as constant as possible. The buffer consisted of 50 mM each of phosphate (which is a good buffer at pD 2.3 and 7.5), acetate (which is a good buffer at

[37] Tsang, K. Y. Unpublished results.

pD 5.1), and borate (which is a good buffer at pD 9.8) in D_2O. Finally, the buffer's ionic strength had to be held roughly constant at all pDs because the C-peptide's helicity is dependent on ionic strength (to the extent that increasing the ionic strength from 100 mM to 1 M causes about a two-fold increase in helicity[2]). Hence, the buffer was adjusted to the various pDs in such a way as to maintain an ionic strength around 200 mM (see the experimental section for details on the buffer's preparation).

It has been shown by analytical ultracentrifugation that as long as the ionic strength is greater than 0.1 M, the C-peptide has no tendency to aggregate[2] at 1 °C at 2-3 mM, and concentrations as high as 6 mM have been used in studies of C-peptide analogs with no signs of aggregation.[25] The peptides in this study are accordingly extremely unlikely to aggregate.

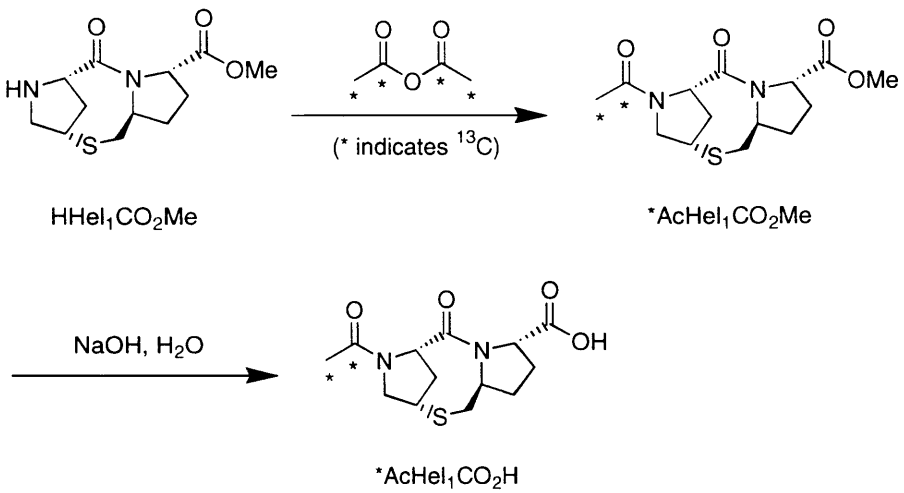
3.4.3 Determination of the t/c Ratios of Highly Heterogeneous $AcHel_1$-Peptide Conjugates

As was discussed in the experimental section of chapter 1, determining accurate t/c ratios hinges on being able to assign and integrate many different peaks and groups of peaks in an $AcHel_1$ derivative's ^{1}H-NMR spectrum. It was anticipated that peak overlaps could interfere with obtaining accurate t/c ratios for the more heterogeneous $AcHel_1$-peptide conjugates used in this work, especially $AcHel_1$A-KETAAAKFERQHA-NH_2. Carbon-13 labels were therefore incorporated into some of the template-peptide conjugates, either in the carbonyl of the first alanine residue or in the acetyl group of the template, so that t/c ratios could be obtained from ^{13}C-NMR spectra. The former was

accomplished simply by substituting the (commercially available) Fmoc-alanine derivative in which the alanine carbonyl carbon is enriched to > 98% ^{13}C for the usual unlabeled Fmoc-alanine derivative (^{13}C labeled alanines will be denoted *A). The latter was accomplished by preparing the $AcHel_1CO_2H$ derivative in which both the acetyl methyl and carbonyl carbons were enriched to > 98% ^{13}C, as shown in scheme 2 (the doubly labeled acetyl group will be denoted *Ac). The free amino template methyl ester ($HHel_1CO_2Me$, whose preparation was developed prior to the work in this thesis[38]) was acylated with the commercially available anhydride ($(^{13}CH_3{}^{13}CO)_2O$, the labeled template ester ($*AcHel_1CO_2Me$) was saponified, and the free acid ($*AcHel_1CO_2H$) was used in solid phase peptide synthesis by the procedures used routinely for their unlabeled counterparts.[39] The peptides synthesized with a ^{13}C enriched alanine were $AcHel_1$*A-KFERQHA-NH_2, $AcHel_1$*A-FERQHA-NH_2, $AcHel_1$*A-ERQHA-NH_2, and $AcHel_1$*A-RQHA-NH_2. The only peptide synthesized with the ^{13}C enriched template was $*AcHel_1$A-KETAAAKFERQHA-NH_2.

[38] Renold, P.; Kemp, D. S. Unpublished results.
[39] Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd, J. G.; Muendel, C. *J. Org. Chem.* **1991**, *56*, 6672.

Scheme 2.



Note that when proton decoupling is used during the acquisition of a spectrum, ^{13}C-NMR cannot provide integrals as accurate as those provided by ^{1}H-NMR. If decoupling is used, as is often necessary in order to obtain an acceptable signal to noise ratio, then inter-nuclear NOEs from the hydrogen to the carbon nuclei result in peak intensity changes such that the integral of a given peak may not correspond only to the number of carbon nuclei contributing to the peak. Accordingly, one must either not use decoupling in the acquisition of the necessary carbon spectra, or if it is necessary to decouple, one must be careful only to compare peaks that are expected to have the same NOE effects. For example, in the decoupled ^{13}C-NMR spectrum of a peptide conjugated to doubly labeled *AcHel$_1$ one would only compare the acetyl carbonyl t state peak to the acetyl carbonyl c state peak and the acetyl methyl t state peak to the acetyl methyl c state peak in order to calculate t/c ratios. Fortunately, only one of the AcHel$_1$-peptide conjugates, *AcHel$_1$A-KETAAAKFERQHA-NH$_2$, had an ^{1}H-NMR spectrum so crowded that it was necessary to estimate t/c ratios from the decoupled ^{13}C-NMR spectra to corroborate the estimates from the ^{1}H-NMR spectra.

3.4.4 Measured t/c Ratios of $AcHel_1$-Peptide Conjugates from the Right- and Left-Hand ^{A13}C-peptide Fragments

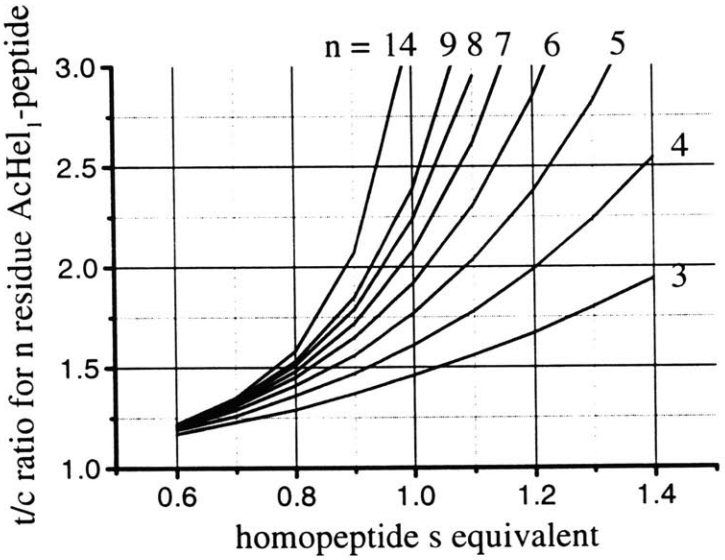
The t/c ratios for all of the $AcHel_1$-peptide derivatives from the ^{A13}C-peptide fragments are listed below in table 4. To give a general sense for how strong the tendencies toward helicity are for the individual fragments, the s values that would produce the same t/c ratios in a template-homopeptide conjugate of the same length are included in the table for the two parent fragments, $AcHel_1$A-KETAAAKA-NH_2 and $AcHel_1$A-FERQHA-NH_2. These homopeptide s values are obtained from the plot in figure 7 that shows the dependence of t/c ratio on peptide length and s value for homopeptides. NMR spectra from which t/c ratios were calculated were obtained at 5 °C and within ±0.2 pD units of pDs 2.3, 5.1, 7.5 and 9.8.

Table 4. t/c Ratios at pDs 2.3, 5.1, 7.5, and 9.8 and 5 °C for the peptides from tables 2 and 3. Standard errors are reported in parenthesis. The s values that would yield the same t/c ratio in a template-homopeptide conjugate of the same length are reported below the t/c ratios for the two parent fragment peptides, $AcHel_1$A-KETAAAKA-NH_2 and $AcHel_1$*A-KFERQHA-NH_2.

Peptides	t/c ratios			
Left-Hand Fragment	pD = 2.3	pD = 5.1	pD = 7.5	pD = 9.8
$AcHel_1$A-KETAAAKA-NH_2	2.33 (±0.12) s_{equiv} = 0.98	2.02 (±0.07) s_{equiv} = 0.93	1.86 (±0.07) s_{equiv} = 0.90	1.73 (±0.05) s_{equiv} = 0.87
$AcHel_1$A-ETAAAKA-NH_2	2.17 (±0.14)	1.92 (±0.11)	2.03 (±0.14)	1.90 (±0.10)
$AcHel_1$A-TAAAKA-NH_2	1.87 (±0.09)	1.70 (±0.04)	1.78 (±0.09)	1.69 (±0.06)
$AcHel_1$A-AAAKA-NH_2	2.63 (±0.08)	2.62 (±0.10)	2.56 (±0.09)	2.47 (±0.08)

Right-Hand Fragment				
$AcHel_1$*A-KFERQHA-NH_2	2.22 (±0.20) s_{equiv} = 1.00	1.85 (±0.21) s_{equiv} = 0.92	1.95 (±0.10) s_{equiv} = 0.93	1.75 (±0.07) s_{equiv} = 0.88
$AcHel_1$*A-FERQHA-NH_2	2.48 (±0.16)	2.27 (±0.15)	2.21 (±0.18)	2.04 (±0.12)
$AcHel_1$*A-ERQHA-NH_2	2.49 (±0.06)	2.08 (±0.07)	1.74 (±0.06)	1.62 (±0.05)
$AcHel_1$*A-RQHA-NH_2	2.35 (±0.06)	2.30 (±0.06)	2.07 (±0.06)	1.92 (±0.08)
$AcHel_1$A-QHA-NH_2	1.94 (±0.04)	1.86 (±0.03)	1.74 (±0.03)	1.56 (±0.05)
$AcHel_1$A-HA-NH_2	1.77 (±0.07)	1.67 (±0.03)	1.50 (±0.03)	1.26 (±0.03)
$AcHel_1$A-A-NH_2	1.31 (±0.04)	1.31 (±0.04)	1.31 (±0.04)	1.31 (±0.04)

Figure 7. The dependence of the t/c ratio of an n-residue $AcHel_1$-homopeptide conjugate on s value.



The t/c ratios in table 4 decrease almost without exception as the pD increases, indicating that, for the most part, where the s values of residues change with pH they

must decrease. This is mirrored in the general decline in the homopeptide s equivalent values for the two parent fragments, $AcHel_1A$-KETAAAKA-NH_2 and $AcHel_1$*A-KFERQHA-NH_2. The t/c ratios that are least affected by increasing pD are those of $AcHel_1A$-TAAAKA-NH_2 and $AcHel_1A$-AAAKA-NH_2, since they do not have any residues whose ionization state changes very much over the pD range from 2.3 to 9.8. Their t/c ratios begin declining only at the highest pD. This is probably because the lysine's protonation state could begin to change around pD 9.8. Based on these t/c ratios, one can anticipate that the helicities predicted for the ^{A13}C-peptide using the s values determined from these data will decrease continually from pD 2.3 to 9.8 rather than having a maximum at medium pD.

Two additional features of the t/c ratios in table 4 can be translated into specific expectations about s values. The first is the abrupt increase in t/c ratio that occurs between $AcHel_1A$-TAAAKA-NH_2 and $AcHel_1A$-AAAKA-NH_2. This indicates that threonine must have a very low helix propensity consistent with the s values recorded for it in table 2 of section 2.6. The second is the abrupt decline in t/c ratio that occurs between pDs 5.1 and 9.8 in $AcHel_1A$-HA-NH_2. This indicates that the helix propensity of histidine must drop significantly over this pD range.

The changes in the t/c ratios as a function of pD, such as those observed in $AcHel_1A$-HA-NH_2, present an opportunity to check a condition necessary for asserting that the peptide's helicities depend on the protonation state of particular side chains. If such were the case, then the t/c ratio, which is a measure of helicity, would have to change with an apparent pK_a that matched the pK_a of the side-chain of the residue in question. The apparent pK_a for the titration of the t/c ratio can be determined as follows for the $AcHel_1$-

peptide conjugates that have only one side-chain that ionizes over the pH range considered (since Lys could start titrating at pD 9.8, this includes only $AcHel_1$*A-RQHA-NH_2, AcHel1A-QHA-NH_2 and $AcHel_1$A-HA-NH_2). Assuming that, for these peptides, the t/c ratio at pD 2.3 represents the t/c ratio in the fully protonated state and the t/c ratio at pD 9.8 represents the t/c ratio for the fully deprotonated state, then the t/c ratios at 5.1 and 7.5 are related to the fraction of deprotonated species ($f_{deprotonated,pD=X}$ where X = 5.1 or 7.5) at that pD by

$$\frac{(t/c)_{pD=2.3} - (t/c)_{pD=X}}{(t/c)_{pD=2.3} - (t/c)_{pD=9.8}} = f_{deprotonated,pD=X}$$

The quantity $f_{deprotonated,pD=X}$ is related to the pK_a of the group being titrated via the Henderson-Hasselbach equation:

$$pD - \log_{10} \frac{f_{deprotonated,pD=X}}{1 - f_{deprotonated,pD=X}} = pK_a$$

For all three template-peptide conjugates, $AcHel_1$*A-RQHA-NH_2, $AcHel_1$A-QHA-NH_2, and $AcHel_1$A-HA-NH_2, the t/c ratios at pD 5.1 are very close to those at pD 2.3. These t/c ratios will not be used to calculate pK_as. Using their t/c ratios at pD = 7.5, which are in all cases about half-way between their t/c ratios at pDs 2.3 and 9.4, yields apparent pK_as for the change in t/c ratio of 7.24, 7.45, and 7.45 respectively. All of these are reasonable pK_as for the side-chain of a potentially helical histidine in D_2O and this correspondence is consistent with the t/c ratios, and thus the helicities of the peptides, varying with histidine's protonation state. The pK_a determined in this way, where measurements have only been made at four pDs, is of necessity very rough, but sufficient for this exploratory work.

3.4.5 Helix Propensities of Residues in the ^{A13}C-peptide Fragments Determined by the Residue Exclusion Method

As described in section 3.3.1, the t/c ratios for each peptide must be converted to $Q_{peptide}$ values in order to obtain s values by the residue exclusion method. These are obtained from the t/c ratios via equation 3-2, and presented for each $AcHel_1$-peptide conjugate in table 5.

Table 5. $Q_{peptide}$ values at pDs 2.3, 5.1, 7.5, and 9.8 and 5 °C for the peptides from tables 2 and 3 calculated from the t/c ratios of table 4 and equation 3-2.

Peptides	$Q_{peptide}$			
Left-Hand Fragment	pD = 2.3	pD = 5.1	pD = 7.5	pD = 9.8
$AcHel_1A$-KETAAAKA-NH_2	7.45	5.51	4.47	3.64
$AcHel_1A$-ETAAAKA-NH_2	6.41	4.84	5.55	4.73
$AcHel_1A$-TAAAKA-NH_2	4.52	3.50	3.98	3.41
$AcHel_1A$-AAAKA-NH_2	9.30	9.25	8.87	8.32
Right-Hand Fragment				
$AcHel_1$*A-KFERQHA-NH_2	6.71	4.39	5.01	3.76
$AcHel_1$*A-FERQHA-NH_2	8.36	7.06	6.70	5.59
$AcHel_1$*A-ERQHA-NH_2	8.43	5.86	3.71	2.97
$AcHel_1$*A-RQHA-NH_2	7.55	7.26	5.77	4.86
$AcHel_1A$-QHA-NH_2	5.00	4.50	3.70	2.57
$AcHel_1A$-HA-NH_2	3.91	3.31	2.19	0.70
$AcHel_1A$-A-NH_2	1.02	1.02	1.02	1.02

The values of $Q_{peptide}$ in this table can now be used directly in the calculation of s values by equation 3-3.

The s values calculated from the $Q_{peptide}$ values in table 5 for the residues in the left-hand fragment KETAAAK will be presented first. Before they are presented, however, the expectations for these s values based on the discussion in section 3.2.2 of the interactions in the C-peptide will be reviewed. The value of s_T should not vary with pD because threonine is neutral and is not known to engage in any interactions that affect helicity. The values of s_{K1} and s_{K7} also should not vary with pD because although these lysines are charged, their ionization states should change very little over the pD range studied. If the helix propensity of lysine were not context dependent, then s_{K1} and s_{K7} would be expected to be the same. However, as Lys s values are known to depend on position when Lys is fewer than five residues away from the template-peptide junction, s_{K1} and s_{K7} are likely to be substantially different. The value of s_{E2} could vary with pD because the protonation state of glutamic acid changes over the pD range studied. Whether or not a significant effect is observed will depend on the extent to which Glu's helix propensity is influenced by the ionization of its side chain. Note that s_{E2} cannot show the effects of the interaction that Glu 2^- has with Arg 10^+ in the C-peptide since this interaction cannot take place in the fragments.

The s values found at 5 °C for the residues in the left-hand ^{A13}C-peptide fragment are arrayed in table 6. Note that the value for s_{K7} in this table was not determined by the residue exclusion method, but directly from the t/c ratio of $AcHel_1A$-AAAKA-NH_2.

Table 6. s Values for the residues in the right hand fragment KETAAAK at 5 °C. Standard errors are reported in parenthesis (their calculation is described in the experimental section).

pD	2.3	5.1	7.5	9.8
s_{K1}	1.00 (±0.16)	0.94 (±0.13)	0.68 (±0.12)	0.64 (±0.09)
s_{E2}	1.16 (±0.20)	1.09 (±0.17)	1.11 (±0.22)	1.07 (±0.17)
s_T	0.44 (±0.06)	0.34 (±0.04)	0.40 (±0.06)	0.37 (±0.05)
s_{K7} (direct from t/c_{AAAAKA})	2.90	2.88	2.70	2.45

The s values of Thr and Lys 7 behave as expected, changing very little over the pD range studied (s_{K7} decreases somewhat at high pD, possibly because the lysine ammonium ion is entering its titration range). The s value of Glu 2 seems to remain constant over the pD range studied. This is perhaps surprising since, according to the s values for Glu^0 and Glu^- in table 2 of chapter 2, the s value of Glu should decrease upon deprotonation. However, considering the error, a small decrease in s_{E2} cannot be ruled out. The behavior of Lys 1's s value is definitely surprising. Whereas it was expected to remain constant, it in fact decreases between pD 2.3 and 7.5 then levels out between pD 7.5 and 9.8.

Now that the s values for the residues in the left-hand fragment have been examined, those calculated from the $Q_{peptide}$ values in table 5 for the residues in the right-hand fragment KFERQHA will be presented. The expectations for these s values based on the discussion in section 3.2.2 of the interactions in the C-peptide are as follows. The value of s_Q should be constant with respect to changing pD since Gln is neutral and not known to engage in any interactions that affect helicity. The values of s_R and s_{K7} also should not

vary with pD because, although Lys and Arg is charged, their ionization states should change very little over the pD range studied. Although Phe 8 is neutral and its intrinsic helix propensity should therefore not be influenced by pD, it is the more N-terminal residue of the Phe 8 / His 12^+ interaction. According to the discussion in section 3.3.2a, the expected interruption in the Phe 8 / His 12^+ interaction when His 12 is deprotonated should manifest itself as a decrease in s_F over the pD range in which His 12 is deprotonated. The s value of histidine, however, should show no effects of the Phe 8 / His 12^+ interaction. If s_H changes with pD it would only reflect the difference in the helix propensity between protonated and unprotonated histidine. Similarly, if the s value of Glu 9 changes with pD, this would reflect only the difference in the helix propensities of protonated and unprotonated glutamic acid.

Table 7. s Values for the residues in the ^{A13}C-peptide at 5 °C. Standard errors are reported in parenthesis (their calculation is described in the experimental section).

pD	2.3	5.1	7.5	9.8
s_{K7} (residue exclusion method	0.72 (±0.16)	0.54 (±0.18)	0.65 (±0.13)	0.57 (±0.10)
s_F	0.89 (±0.11)	1.03 (±0.15)	1.41 (±0.27)	1.41 (±0.22)
s_{E9}	0.99 (±0.06)	0.71 (±0.06)	0.55 (±0.07)	0.51 (±0.07)
s_R	1.26 (±0.08)	1.32 (±0.08)	1.23 (±0.10)	1.35 (±0.19)
s_Q	1.01 (±0.10)	1.05 (±0.07)	1.16 (±0.09)	1.52 (±0.26)
s_H	1.93 (±0.39)	1.62 (±0.28)	1.10 (±0.20)	0.34 (±0.14)

The s values of Lys 7 and Arg remain roughly constant over the pD range, as expected. The s values of Glu 9 and His both change substantially over the pD range, the

former by a factor of 1/2 and the latter by a factor of 1/6, indicating that both of these residues' helix propensities are strongly dependent on the protonation states of their side chains. In both cases, the bulk of the change is over the pD range in which their respective side-chains titrate. The decrease in s_{E9} occurs almost entirely between pD 2.3 and pD 7.5, while the decrease in s_H occurs almost entirely between pD 5.1 and 9.8. The s values of Gln and Phe do not at all behave as expected. Whereas s_Q should have remained constant, it increases by 50% largely between pD 5.1 and 9.8. This is the same pD region where the drop in s_H is observed. Whereas s_F should have decreased over the pD range where the His side-chain titrated, it increases by 60% between pD 2.3 and 7.5. This is the same pD region where the drop in s_{E9} is observed.

With the sets of s values derived from both fragments in hand, some comparisons between the sets can be made. Two types of residues (other than alanine) are repeated in the sequence of the ^{A13}C-peptide: Lys and Glu. Since none of the Lys or Glu residues are expected to interact with anything in the peptide fragments, two pairs of s values that should have been the same over the entire pD range. The first pair is s_{K1} determined in the left-hand fragment and s_{K7} determined in the right-hand fragment (because of lysine's position dependent helix propensity, s_{K7} determined directly from the t/c ratio of $AcHel_1A$-AAAKA-NH_2 should not be the same as either of these; this issue is addressed below). The second pair is s_{E2} determined in the left-hand fragment and s_{E9} determined in the right-hand fragment. Of these two pairs, neither of them fit the expectation of similarity. The values of s_{K1} and s_{K7} are moderately different at low pD, although the amount of error in these s values precludes making any definite statements. The values of s_{K1} and s_{K7} only coincide at the higher pDs. The values of s_{E2} and s_{E9} are significantly

different at high pD. In particular, at pDs 7.5 and 9.8 the errors are small enough and the differences are large enough that the preceding statement can be made with confidence. The values of s_{E2} and s_{E9} only coincide at low pD.

The last feature of the s values in tables 6 and 7 that begs commentary is the two different values determined for s_{K7}, one in the left-hand fragment directly from the t/c ratio of $AcHel_1A$-AAAKA-NH_2 and one in the right-hand fragment by the residue exclusion method. If context did not affect the helix propensity of lysine, then one would expect the two to be the same. However, when Lys is less than five residues away from the $AcHel_1$ its helix propensity is known to be strongly position dependent. The value of s_{K7} determined in the left-hand fragment is that for a lysine at the fifth position beyond the helix initiation site, while the value of s_{K7} determined in the right-hand fragment is that for a lysine at the second position beyond the helix initiation site. These two determinations of s_{K7} should not be commensurate, and in fact they are not. They differ by a factor of between four and five. Nevertheless, each should authentically represent the contribution of lysine toward the stability of helices in the left- or right-hand fragments. How to use these values of s_{K7} to best represent Lys 7's contribution to the ^{A13}C-peptide's helicity will be addressed in the following section.

The s values presented in this section represent a characterization of the helicities of the left- and right-hand fragments of the ^{A13}C-peptide. A rigorous application of this data to the full ^{A13}C-peptide's helicity requires two more pieces of information: the energetics of the Glu 2^- / Arg 10^+ interaction, and how the two different values of s_{K7} determined in the left- and right-hand fragments should be used in the description of the full peptide's helicity. Neither of these two issues can be addressed directly using the data presented

thus far, but a preliminary comparison can be made between the experimental helicity of the ^{A13}C-peptide and the helicity calculated using the s values determined in the right and left hand fragments.

3.4.6 Comparing the Predicted and Observed Helicities of the ^{A13}C-peptide

With the s values from tables 6 and 7, predicted and observed measures of the ^{A13}C-peptide's helicity can be compared to test how well the s values measured in the ^{A13}C-peptide fragments by the residue exclusion method reflect the helicity of the full peptide. Recall that this can be done in two ways: by comparing the calculated and observed fractional helicities of H-AKETAAAKFERQHA-NH_2, and by comparing the calculated and observed t/c ratios of *$AcHel_1$A-KETAAAKFERQHA-NH_2.

3.4.6a Predicted and Observed t/c Ratios of *$AcHel_1$A-KETAAAKFERQHA-NH_2

The latter comparison will be made first because it will provide information relating to three questions. First, does the Glu 2^- / Arg 10^+ interaction block helices from extending to the template, as suggested in section 3.3.3b? Second, can this comparison be used to justify the residue exclusion and divided peptide methods to the study of helicity? Third, are the s_{K7} values determined by the residue exclusion method closer to those in the ^{A13}C-peptide than those determined directly from t/c_{AAAAKA}, or vice versa, or must they be used in combination? The answer to the first question will come directly from the

experimentally measured t/c ratios. If they were to show the pD dependence expected based on the known pD dependence of the C-peptide, then it would be likely that the Glu 2^- / Arg 10^+ interaction was compatible with helices initiated from the template. If they did not show the expected pD dependence, then the opposite hypothesis would be supported. The answer to the second question will come from comparing the predicted t/c ratios calculated using the two sets of s_{K7} values with the observed t/c ratios. Either one of them matching the observed t/c ratios could be regarded as evidence that the assumptions underlying both the use of the residue exclusion method and the divided peptide method are valid. Note that if the effects of the Glu 2^- / Arg 10^+ are witnessed in the pD dependence of the t/c ratio, the predicted and observed t/c ratios can only be expected to match at low pD. Otherwise, they should match throughout the pD range. The answer to the third question derives naturally from the answer to the second. If one of the sets of predicted t/c ratios were to match the observed t/c ratios, then the set of s_{K7} values used to calculate them would have to be the more sound. If neither of them did, then the possibility of somehow using them in combination would have to be examined.

The predicted and observed t/c ratios at 5 °C are listed in table 8. The ^{1}H-NMR spectra of *AcHel$_1$A-KETAAAKFERQHA-NH_2 at all pDs were very crowded and t/c ratios could only be extracted from them with difficulty. In fact, a t/c ratio could not be obtained from the ^{1}H-NMR spectrum at pD 9.8. Because of the ^{13}C labels in the template acetyl group, however, there was the opportunity to obtain the necessary t/c ratios from ^{13}C-NMR spectra. In practice, the ^{13}C-NMR spectra had to be ^{1}H decoupled in order to get satisfactory signal to noise ratios, and only the ^{13}C signals of the acetyl carbonyl

group were well enough separated in the t and c states to provide t/c ratios, but as shown in table 8 the t/c ratios from the ^{13}C-NMR and ^{1}H-NMR agree very well.

Table 8. Predicted vs. observed t/c ratios for *AcHel$_1$A-KETAAAKFERQHA-NH$_2$ at 5 °C.

	predicted t/c ratio		observed t/c ratio		
pD	using s_{K7} from residue exclusion method	using s_{K7} direct from t/c_{AAAAKA}	^{1}H-NMR	^{13}C-NMR	average
2.3	2.41	4.17	2.60	2.66	2.63
5.1	1.94	3.05	2.22	2.25	2.24
7.5	1.83	2.63	1.56	1.75	1.66
9.8	1.69	2.22	-	1.60	1.60

The data in table 8 provide the answers to the questions posed above. First, the t/c ratios do not show the pD dependence one would have expected based on the pH dependence of the C-peptide's helicity. Instead of having a maximum at pD 5.1 that decreases at low and high pD there is a constant decrease from pD 2.3 to 9.8, suggesting that the Glu 2$^-$ / Arg 10$^+$ interaction does in fact block the helices that contain it from propagating all the way to the N-terminus. Second, the agreement between the observed t/c ratios and those calculated with the s_{K7} values derived from the left-hand fragment by the residue exclusion method is good. The root mean square deviation (RMSD) between the predicted and observed t/c ratios is only 0.21, about 10% of the observed t/c ratios. This degree of accuracy demonstrates the utility of the s values determined by the residue exclusion method from fragments of the full ^{A13}C-peptide and that the divided peptide method can be successful. This is only a partial success, though, because it results from the inability of the t/c ratio to respond to the Glu 2$^-$ / Arg 10$^+$ interaction. The predicted and observed t/c ratios match each other because neither of them shows any contributions

from this interaction. Third, the match between the t/c ratios predicted using the s_{K7} values determined by the residue exclusion method with the observed t/c ratios indicates that these s values are the more relevant to the context of the full peptide. The environment of the right-hand fragment is apparently more like Lys 7's situation in the ^{A13}C-peptide than the environment of the left-hand fragment. An explanation for why this should be could be addressed in the future by characterizing the helicity of central fragments of the ^{A13}C-peptide.

3.4.6b Predicted and Observed Fractional Helicities of H-AKETAAAKFERQHA-NH_2

The t/c ratios of *AcHel$_1$A-KETAAAKFERQHA-NH_2 do not transmit any information about the Glu 2^- / Arg 10^+ interaction, and therefore provide only an incomplete picture of the ^{A13}C-peptide's helicity. The fractional helicities of H-AKETAAAKFERQHA-NH_2, however, do not suffer from this drawback. These fractional helicities can therefore be used to ascertain how much ignoring the Glu 2^- / Arg 10^+ interaction compromises the accuracy of helicity predictions. The discrepancy between predicted and observed helicities at higher pDs should be in direct proportion to the magnitude of the aforementioned interaction. However, at low pD, where the interaction doesn't exist, the predicted and observed fractional helicities should match each other fairly well.

Helicities predicted for H-AKETAAAKFERQHA-NH_2 using the Zimm-Bragg model (taking $\sigma = 2 \times 10^{-3}$) are compared with the helicities observed at 5 °C in table 9, where the observed fractional helicities are computed from the per residue molar ellipticities at

222 nm as described in section 2.5. For the reasons discussed there, the observed fractional helicities are shown as ranges depending on the value used for $[\theta]_{222,100\%\ helix}$. Note that the values of s_{K7} derived from the right-hand fragment by the residue exclusion method are used to compute the predicted helicities in table 9 since these provided the more accurate predicted t/c ratios in table 8.

Table 9. Predicted vs. observed fractional helicities of H-AKETAAAKFERQHA-NH_2 at 5 °C. Measurements were made using the same deuterated buffer as was used in the NMR experiments. All values of $[\theta]_{222}$ are reported in deg cm^2 / dmol res. Peptide concentration was 75 μM.

pD	predicted fractional helicity	$[\theta]_{222}$ (deg cm^2/dmol res)	observed fractional helicity	
			$[\theta]_{222,100\%helix}$ = −42,000	$[\theta]_{222,100\%helix}$ = −32,000
2.3	0.063	-2,480	0.059	0.078
5.1	0.046	-4,690	0.112	0.147
7.5	0.045	-3,450	0.082	0.108
9.8	0.033	-2,650	0.063	0.083

It should first be noted that the helicities for H-AKETAAAKFERQHA-NH_2 reported in this table are somewhat less than those reported for the C-peptide in table 1. It is unlikely that the extra alanine at the N-terminus of this peptide or the Hse → Ala substitution at the C-terminus could be responsible for this. These two replacements should increase helicity, if only because they increase the number of residues in the peptide (H-AKETAAAKFERQHA-NH_2 has 13 residues compared to 11 in the C-peptide). It has been noted that problems of concentration determination may have influenced the values of $[\theta]_{222}$ obtained in some studies of C-peptide analogs.[24] Since

concentrations were determined in the same way in the study from which the data in table 1 were taken,[11] this could account for the discrepancy.

As shown in table 9, the predicted fractional helicity falls within the range of the observed fractional helicity at pD 2.3, as expected. Also as expected, predicted fractional helicities fall well outside the ranges of the observed fractional helicities at all other pDs. The RMSD between the predicted fractional helicities and of the mid-point of observed helicities' ranges at pDs 5.1, 7.5 and 9.8 is 0.061, which corresponds to 50% – 75% of the observed fractional helicities. Given that this discrepancy is due entirely to the Glu 2^- / Arg 10^+ interaction, this interaction is clearly an important effect and leaving it unaccounted for significantly diminishes the accuracy of predictions of the ^{A13}C-peptide's helicity. Still, the success at pD 2.3 indicates both that the residue exclusion method yields reasonable s values for use in the Zimm-Bragg model and that the divided peptide method is valid when all interactions in the full peptide are accounted for in the fragments.

3.4.6c The Glu 2^- / Arg 10^+ Interaction and the Total Helicity of *AcHel$_1$A-KETAAAKFERQHA-NH$_2$

The two indices of the ^{A13}C-peptides helicity that have been used so far, the t/c ratio of *AcHel$_1$A-KETAAAKFERQHA-NH$_2$ and the CD derived fractional helicity of H-AKETAAAKFERQHA-NH$_2$, show sharply differing pD dependencies. It is a legitimate concern that perhaps the cause of the discrepancy is not that a distortion caused by the Glu 2^- / Arg 10^+ interaction prevents helices from extending to the template, but that the

template engenders distorted helices in which the Glu 2^- / Arg 10^+ interaction cannot occur. If it were true that the Glu 2^- / Arg 10^+ interaction did not occur in *AcHel$_1$A-KETAAAKFERQHA-NH_2 (as opposed to it occurring and merely being ignored in the t/c ratio), then one would expect the pD dependence of the CD derived fractional helicity of the template-peptide conjugate to mirror the pD dependence of the t/c ratio. All contributions to *AcHel$_1$A-KETAAAKFERQHA-NH_2's total helicity would originate at the template, so a measure of the total helicity (such as $[\theta]_{222}$) would match a measure of the N-terminally initiated helicity (such as the t/c ratio). If this were not true, if the Glu 2^- / Arg 10^+ interaction was the distorting influence as has been supposed, then one would expect a different behavior. The Glu 2^- / Arg 10^+ interaction would influence *AcHel$_1$A-KETAAAKFERQHA-NH_2's total helicity independently of the template's presence and one would expect to see a superposition of effects in the pD dependence of the fractional helicity. On the one hand, the helicity should have a tendency to decrease with increasing pD as manifested in the pD dependence of *AcHel$_1$A-KETAAAKFERQHA-NH_2's t/c ratio. On the other hand, the helicity should also have a tendency to increase between pD 2.3 and 5.1, and then to decrease again between pD 5.1 and 9.8 as observed in the pD dependence of H-AKETAAAKFERQHA-NH_2's fractional helicity. The sum of these two tendencies should yield a helicity that stays nearly constant between pD 2.3 and 5.1 (since the two tendencies act contrarily to each other) and then declines regularly form pD 5.1 to 9.8.

The measured per residue molar ellipticities of *AcHel$_1$A-KETAAAKFERQHA-NH_2 after correction for the template's known CD spectrum[40] are listed in table 10 (the application of template corrections is described in the experimental section).

Table 10. Measured per residue molar ellipticities of *AcHel$_1$A-KETAAAKFERQHA-NH_2 at 5 °C. Measurements were made using the same deuterated buffer as was used in the NMR experiments. All values of $[\theta]_{222}$ are reported in deg cm^2 / dmol res. Peptide concentration was 56 μM.

pD	$[\theta]_{222}$ (deg cm^2 / dmol res)	observed fractional helicity	
		$[\theta]_{222,100\%helix}$= -42,000	$[\theta]_{222,100\%helix}$= -32,000
2.3	-7,780	0.185	0.243
5.1	-7,860	0.187	0.246
7.5	-5,920	0.141	0.185
9.8	-4,320	0.103	0.135

As shown in table 10, the fractional helicity of *AcHel$_1$A-KETAAAKFERQHA-NH_2 changes very little between pD 2.3 and 5.1 and then drops at higher pDs. These data therefore support the hypothesis that the Glu 2$^-$ / Arg 10$^+$ interaction requires a distortion that prevents helices from reaching the template. The Glu 2$^-$ / Arg 10$^+$ interaction and AcHel$_1$ act independently to promote helicity.

[40] Oslick, S. L. Thesis, Massachusetts Institute of Technology, **1996**.

3.5 Discussion

3.5.1 Evidence for Unexpected Interactions

There is an unexpected feature of the s values reported in tables 6 and 7 that should be emphasized: the dependence of residues' s values on the protonation states of the residues that succeed them. For example, s_{K1} changes over the pD range in which the Glu 2's side chain is deprotonated, s_F changes over the pD range in which the Glu 9's side chain is deprotonated, and s_Q changes over the pD range in which His 12 is deprotonated. This behavior could be an indication that either Lys 1, Phe 8, and Gln 11 all interact with the residues that succeed them, or Glu 2, Glu 9, and His 12 all have position dependent s values. Although much more experimentation will be required to establish which of these is the case, a preliminary judgement can be made using the criteria discussed in section 3.3.2c. If an interaction occurs, only the s value of the more N-terminal of the two interacting residues should be influenced by the protonation state of the more C-terminal residue's protonation state. In contrast, if a residue has a position dependent s value, its protonation state should affect the s values of all of the residues that precede it. According to these criteria, since only s_Q changes in the pD range where histidine's protonation state changes, the effect observed in s_Q is likely to be due to an interaction between Gln 11 and His 12 rather than a position dependence of s_H. The same argument can be made for the effect of Glu 9's protonation state on s_F. As for the effect of Glu 2's protonation state on s_{K1}, Lys 1 is the only residue that precedes Glu 2 so no other s values are available to distinguish the effects of an inter-residue interaction from those of a

position dependent s value. No judgement can be made in this case until more data become available.

3.5.2 The Glu 2^- / Arg 10^+ and Phe 8 / His 12^+ Interactions in the ^{A13}C-peptide

The Glu 2^- / Arg 10^+ and Phe 8 / His 12^+ interactions have been proposed in the literature as the causes of the C-peptide's high helicity. There should accordingly be evidence in the results above indicating the presence of these interactions.

Such evidence for the Glu 2^- / Arg 10^+ interaction is present in the discrepancy between calculated and observed helicities of the ^{A13}C-peptide. Assuming that the discrepancy between calculated and observed helicities is due solely to the Glu 2^- / Arg 10^+ interaction, it is possible to quantify the energetics of this interaction using the Zimm-Bragg model and the s values from tables 6 and 7. The sum of equilibrium constants for H-AKETAAAKFERQHA-NH_2, which is given in appendix 1, can be modified to include terms that represent the presence of helices that contain the Glu 2^- / Arg 10^+ interaction. In these helices, helical ϕ and ψ torsions begin at Ala 4 and extend through at least Arg 10^+ and up to Ala 13. Thus, four helical states can support the Glu 2^- / Arg 10^+ interaction and these have the following equilibrium constants:

1. $\sigma s_A s_{K7} s_F s_{E9} s_R \times e^{-\Delta Gint/RT}$ for the helix from Ala 4 to Arg 10^+
2. $\sigma s_A s_{K7} s_F s_{E9} s_R s_Q \times e^{-\Delta Gint/RT}$ for the helix from Ala 4 to Gln 11
3. $\sigma s_A s_{K7} s_F s_{E9} s_R s_Q s_H \times e^{-\Delta Gint/RT}$ for the helix from Ala 4 to His 12^+
4. $\sigma s_A s_{K7} s_F s_{E9} s_R s_Q s_H s_A \times e^{-\Delta Gint/RT}$ for the helix from Ala 4 to Ala 13

where ΔG_{int} is the energy of the Glu 2^- / Arg 10^+ interaction. (Recall that for helices beginning at Ala 4, the first hydrogen bonded residue is two residues later in the sequence; hence the first s value in these terms is that of Ala 6.) These terms can be added to H-AKETAAAKFERQHA-NH_2's sum of equilibrium constants to represent the interaction's contribution to the helicity, and then the values of ΔG_{int} that produce the most accurate predicted fractional helicities at each pD can be found. These are listed in table 11. Of course, since a range of fractional helicities is reported for H-AKETAAAKFERQHA-NH_2 in table 9, a range of interaction energies is given. All calculations were made assuming $\sigma = 2\times10^{-3}$.

Table 11. Energies of the Glu 2^- / Arg 10^+ interaction at 5 °C (calculated using $\sigma = 2 \times 10^{-3}$).

pD	range of ΔG_{int} (kcal/mol)	range of fractional helicity
2.3	[no interaction]	[no interaction]
5.1	−1.7 to −2.0	0.112 to 0.147
7.5	−1.4 to −1.7	0.082 to 0.108
9.8	−1.5 to −1.8	0.063 to 0.083

Given the above ranges, a reasonable estimate for the interaction energy is −1.7 kcal/mol, an energy that corresponds to increasing an equilibrium constant by a factor of about 20 at 5 °C. This interaction energy is large, especially considering that it takes place in aqueous solution between residues that are at positions i and i+8 relative to each other. For comparison, the total charge-dipole / hydrogen bonding interaction determined in chapter 1 between the ammonium ion and acetamide carbonyl in $AcHel_1CH_2NH_3^+$

amounts to −0.70 kcal/mol (t/c ratio = 5.23 for $AcHel_1CH_2NH_3^+$ compared to t/c ratio = 1.60 for $AcHel_1CH_2OMe$ at 25 °C; see chapter 1). The interaction energy between Lys^+ and Glu^- at an i, i+4 spacing in helical peptides has been reported at only −0.3 kcal/mol.[41] Finally, the interaction energies between hydrogen bonding pairs at protein surfaces where at least one member is charged range from −0.3 to −1.0 kcal/mol.[42] If the interaction energies listed in table 11, and therefore the assumption underlying their calculation, are correct, then the Glu 2^- / Arg 10^+ interaction is substantially more stabilizing than any of these.

Unlike the situation of the Glu 2^- / Arg 10^+ interaction, none of the results presented thus far support the existence of the Phe 8 / His 12^+ interaction. Had this interaction been the cause of the helicity decrease seen at the basic end of the C-peptide's helicity vs. pD curve, s_F should have decreased over the pD range in which His 12 is deprotonated. Such behavior is not observed; rather, s_F increases over the pD range in which Glu 9 is deprotonated. The decrease in the C-peptide's helicity observed at high pD can instead be explained in terms of the drop in histidine's intrinsic s value at high pD, which decreases enough between pD 5.1 and 9.8 to explain the bulk of the C-peptide's helicity decrease at high pD. This is not entirely inconsistent with previous work on the C-peptide. While it has been demonstrated that C-peptide analogs with the Phe 8 / His 12 combination show a dramatic helicity decrease when the histidine is deprotonated, it has also been demonstrated that C-peptide analogs in which Phe 8 has been replaced by an alanine show a similar, albeit smaller, helicity decrease.[28] Nevertheless, it is difficult to argue that there is no interaction between Phe 8 and His 12. Even if the decrease in s_H can

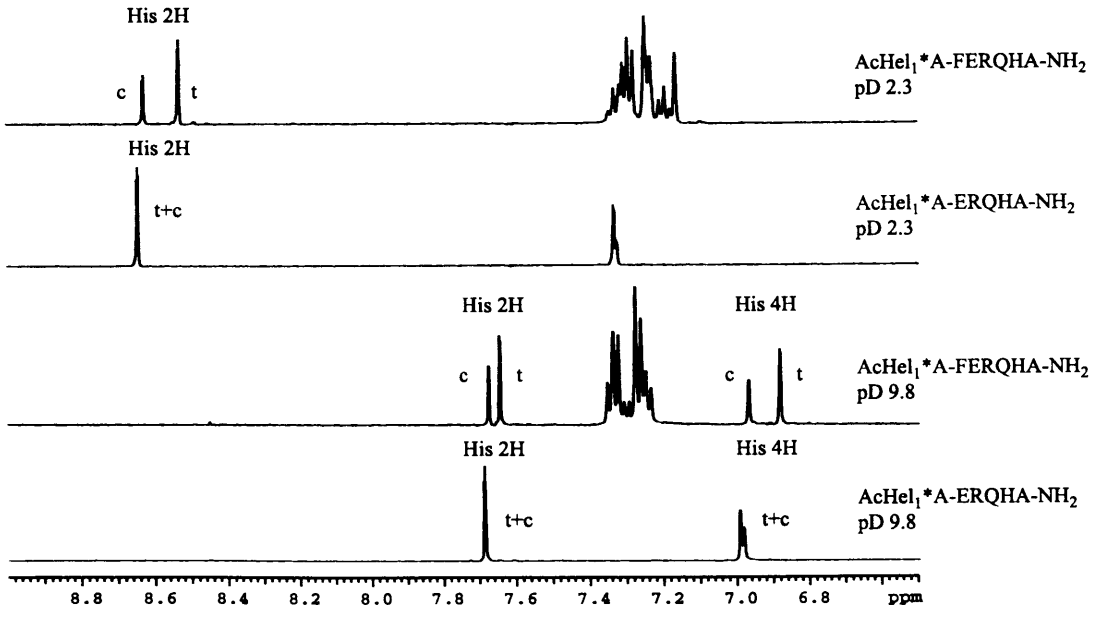
---

[41] Scholtz, J. M.; Qian, H.; Robbins, V.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668.
[42] Serrano, L.; Kellis, Jr. J. T.; Cann, P.; Matouschek, A. Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 783.

partly explain the decrease in the C-peptide's helicity at high pD, there is still evidence for the interaction from analyses of the chemical shifts of their respective aromatic protons.[23,26,27]

A model proposed by Dadlez et al.[27] in which Phe 8 interacts not only with His 12^+ but also with neutral His 12^0 can reconcile the observations from this study and the literature. If this were the case, one would expect s_F to reflect the different contributions of two stabilizing interactions rather than the difference between the presence and absence of an interaction. While s_F should still have decreased over the pD range in which His 12 was deprotonated, the decrease would have been much less and could easily have been covered up in the experimental error in s_F.

Figure 8. Aromatic regions of the NMR spectra of $AcHel_1$*A-FERQHA-NH_2 and $AcHel_1$*A-ERQHA-NH_2 at pDs 2.3 and 9.8. Only the His imidazole peaks of the former show a chemical shift difference between the t and c states



Evidence for this model is present in the ^{1}H-NMR spectra of AcHel$_1$*A-FERQHA-NH$_2$ and AcHel$_1$*A-ERQHA-NH$_2$. The aromatic regions of the ^{1}H-NMR spectra of these two compounds at pD 2.3 and 9.8 are shown in figure 8. The chemical shifts of the histidine imidazole 2H differ by 0.1 ppm between the t state ($\delta_{2H,t}$ 8.53 ppm) and the c state ($\delta_{2H,c}$ 8.63 ppm) at pD 2.3 in AcHel$_1$*A-FERQHA-NH$_2$. In contrast, the difference between t and c states for the corresponding proton in AcHel$_1$*A-ERQHA-NH$_2$ is barely detectable ($\delta_{2H,t}$ and $\delta_{2H,c}$ both occur at 8.65 ppm) at the same pD. Note that $\delta_{2H,c}$ is about the same in both AcHel$_1$-peptide conjugates. This indicates that Phe 8 has an influence on the environment of the His 12^+ side chain, but only in the t state where the two residues can be helical. Similarly, at pD 9.8 the chemical shifts of both the histidine imidazole 2H and 4H differ substantially between the t and c states of AcHel$_1$*A-FERQHA-NH$_2$ ($\delta_{2H,t}$ = 7.64 ppm, $\delta_{2H,c}$ = 7.68 ppm and $\delta_{4H,t}$ = 6.88 ppm, $\delta_{4H,c}$ = 6.97 ppm) but not in AcHel$_1$*A-ERQHA-NH$_2$ ($\delta_{2H,t} \approx \delta_{2H,c}$ = 7.69 ppm and $\delta_{4H,t} \approx \delta_{4H,c}$ = 6.99 ppm). Again, $\delta_{2H,c}$ is the same in both AcHel$_1$-peptide conjugates and, additionally, $\delta_{4H,c}$ is the same in both AcHel$_1$-peptide conjugates. This indicates that Phe 8 retains its influence on His 12 in the t state even when His 12 is uncharged.

3.5.3 Evaluation of the Reductionist Approach to the Study of Peptide Conformation

This is a preliminary work to test the feasibility of the divided peptide and residue exclusion methods and the reductionist approach in general. Thus, the most significant result from this chapter is the evidence that the s values determined from the left- and right-hand ^{A13}C-peptide fragments could be used to accurately predict both the t/c ratio of

*$AcHel_1$A-KETAAAKFERQHA-NH_2 at several pDs and the fractional helicity of H-AKETAAAKFERQHA-NH_2 at low pD. It appears that the helical tendencies of the ^{A13}C-peptide fragments accurately represent the helical tendencies of the full peptide so long as the Glu 2^- / Arg 10^+ interaction, which is not accounted for in the fragments, is not active. This success encourages further work to test the potential of the methods used in this chapter. In the case of the C-peptide and its analogs, many issues regarding the nature of its helicity remain unresolved that could be addressed by the application of the divided peptide and residue exclusion methods. For example, it is not understood why using the value of s_{K7} determined by the residue exclusion method in the left-hand ^{A13}C-peptide fragment in the calculation of the *$AcHel_1$A-KETAAAKFERQHA-NH_2's t/c ratio should yield such better results than using the value of s_{K7} determined directly from the t/c ratio of $AcHel_1$A-AAAKA-NH_2. In addition, as discussed in section 3.2.3, it is not understood why C-peptide analogs with Leu at the 9 position have much larger helicities than otherwise identical C-peptide analogs with Glu in the 9 position, or why C-peptide analogs with Ala at the 11 position have much smaller helicities than otherwise identical C-peptide analogs with Gln in the 11 position. Extensions of the work in this chapter in which the helicity of fragments encompassing the central part of the ^{A13}C-peptide's sequence and fragments with specific substitutions, such as Glu 9 → Leu and Gln 11 → Ala, ought to provide the information necessary to fully explain the helicities of the C-peptide and its analogs.

3.6 Summary

The divided peptide method, a reductionist approach in which a peptide's helicity is studied in terms of the helicities of its fragments, has been applied to a close analog of the ribonuclease C-peptide. The ^{A13}C-peptide (in which alanine replaces the homoserine at the C-terminus of the C-peptide) was divided into two fragments. The helicities of the fragments were characterized by determining the s values of each residue in both fragments at four pDs. The s values were determined by a novel method, the residue exclusion method, from the t/c ratios of a series of $AcHel_1$-peptide conjugates each one related to the previous one by the exclusion of a single residue. A consistent and surprising trend was observed in which helix propensities seemed to depend on the protonation states of the residues that succeeded them. The s values determined by the residue exclusion method were used to predict the t/c ratios of the ^{A13}C-peptide attached to $AcHel_1$ via an alanine spacer (*$AcHel_1$A-KETAAAKFERQHA-NH_2) and the fractional helicities of the free ^{A13}C-peptide with an added N-terminal alanine (H-AKETAAAKFERQHA-NH_2) at four pDs. The Glu 2^- / Arg 10^+ interaction was the determining factor in whether the predictions fit the experimental observations. The s values determined by the residue exclusion method did not account for this interaction, because neither fragment contained both of the interacting residues. The t/c ratios of *$AcHel_1$A-KETAAAKFERQHA-NH_2 did not show the effects of this interaction because the interaction interferes with N-terminal helix propagation. On the other hand, the H-AKETAAAKFERQHA-NH_2's fractional helicity (determined by CD spectroscopy) did show the effects of the interaction. Thus, the predicted and observed t/c ratios matched

well at all four pDs, while the predicted and observed fractional helicities only matched at low pD where Glu 2 was protonated and the interaction did not occur. Whereas the impact of the Glu 2^- / Arg 10^+ interaction was clear from this study, the Phe 8 / His 12^+ interaction did not manifest itself very strongly. This was consistent with earlier proposals[27] that an interaction took place between Phe 8 and His 12^0 as well as Phe 8 and His 12^+, so that deprotonation of the histidine would not necessarily affect helicity by destroying the interaction. The decline in the C-peptide's helicity at high pD was instead attributed to the sharp decrease in histidine's intrinsic helix propensity upon its deprotonation.

The success of the divided peptide and residue exclusion methods in this limited study justifies their future use in more rigorous studies of the C-peptide and other peptides. Furthermore, it justifies the use of reductionist approaches in general.

3.7 Experimental

Equipment. ^{1}H-NMR spectra were measured at 500 MHz and ^{13}C-NMR spectra were measured at 125 MHz on Varian VXR500S and 501S spectrometers and processed using the Varian Instruments VNMR 3.1 software. Chemical shifts were measured relative to the reference signal of (trimethylsilyl) propionic-2,2,3,3-d_4 acid (TMSP). ^{1}H-NMR spectra were obtained using a 60° pulse width with a 4 second acquisition time (80,000 points) and a 3 second delay between pulses. Between 128 and 512 transients were acquired, depending on the concentration of the sample (typically between 0.5 and 2 mM). ^{13}C-NMR spectra were obtained using a 60° pulse width and broadband proton

decoupling with a 1 second acquisition time and a 3 second delay between pulses. Typically, > 2500 transients were required in order to get accebtable signal to noise ratios. CD spectra were measured on an Aviv 62DS circular dichroism spectrometer using 5 mm strain free quartz cells (Hellma). CD spectra were processed with Aviv 62DS version 4.0s software. Analytical high performance liquid chromatography (HPLC) was performed on a Waters system consisting of two 501 pumps, a rheodyne injector, a model 660 automated gradient controller, a model 740 data module, a model 484 detector, and a Vydac 0.46 × 25 cm (218TP54) C_{18} reverse phase column. Flow rates were 1.0 mL/min. Preparative scale HPLC was performed on a Waters system consisting of a model 590 pump fitted with preparative heads, an Autochrome DPG/S pre-pump solvent mixer, a Rheodyne injector, a model 484 variable wavelength detector, and a Waters 2.5 × 10 cm radial compression column housed in a PrepLC 2.5 cm radial compression module (RCM). Flow rates for preparative HPLC were 12 mL/min. Detection in all uses of HPLC was carried out at 214 nm unless otherwise specified. The pDs of solutions were measured using a Cole-Parmer pH meter (model # 5982-00) and a 3.5 × 183 mm glass electrode with a calomel reference (model # 5990-30). The pH meter was referenced to pH 4.00 and 7.00 certified buffers. The H/D isotope effect was accounted for by adding 0.4 to the pD that was read off the meter ($pD = pD_{read} + 0.4$).[43] Mass spectra were obtained courtesy of Dr. P. Wishnok and Prof. S. Tannenbaum on a Hewlett-Packard HP5989 electrospray ionization mass spectrometer from samples dissolved in 1:1 water: methanol with 0.1% acetic acid, detecting either positive or negative ions as necessary.

---

[43] Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188.

Measurement of t/c Ratios. The procedure by which t/c ratios were obtained from ^{1}H-NMR spectra was described in chapter 1. Obtaining t/c ratios from ^{13}C-NMR spectra was much simpler because very few t and c state peaks were available for integration. The proton decoupled ^{13}C-spectrum of *AcHel$_1$A-KETAAAKFERQHA-NH$_2$, the only compound for which t/c ratios had to be obtained from the ^{13}C-NMR spectrum, had two sets of t and c state peaks. Each set of peaks consisted of doublets (due to the ^{13}C-^{13}C coupling, $J \approx 50$ Hz) for the t and c states. Unfortunately, the doublets for the acetyl methyl carbon overlapped, but the doublets for the acetyl carbonyl of *AcHel$_1$A-KETAAAKFERQHA-NH$_2$ were well separated. The t and c state doublets for the acetyl carbonyl were integrated, and the ratio of the integrals taken to yield the t/c ratio.

Standard Errors of s Values.[44] Given a random variable, y, that is a function f(x_1, x_2, ...) of other random variables x_1, x_2, ... that have expectations μ_1, μ_2, ..., then the expectation of y is

$E(y) = f(\mu_1, \mu_2, ...)$

and if the standard deviations of x_1, x_2, etc are small relative to the expectation values, the standard deviation of y is given by:

$$sd_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \times sd_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \times sd_{x_2}^2 + \cdots$$

Values of s determined by the residue exclusion method are functions of Q_1 and Q_2, the Q values used to determine them, which are in turn functions of $(t/c)_1$, $(t/c)_2$, the two corresponding t/c ratios, A, B, and s_A:

---

[44] Devore, J. L. *Probability and Statistics for the Sciences, 4th ed.*; Wadsworth: Belmont, CA, **1995**.

$$s = \frac{Q_1}{1+Q_2} = \frac{\dfrac{(t/c)_1 - A - B - Bs_A}{Bs_A}}{1+\dfrac{(t/c)_2 - A - B - Bs_A}{Bs_A}}$$

The standard deviation of s is therefore given by

$$sd_s^2 = \left(\frac{\partial s}{\partial (t/c)_1}\right)^2 \times sd^2_{(t/c)_1} + \left(\frac{\partial s}{\partial (t/c)_2}\right)^2 \times sd^2_{(t/c)_2} + \left(\frac{\partial s}{\partial A}\right)^2 \times sd_A^2 +$$

$$\left(\frac{\partial s}{\partial B}\right)^2 \times sd_B^2 + \left(\frac{\partial s}{\partial s_A}\right)^2 \times sd^2_{s_A}$$

Actually, since only standard errors are known for the t/c ratios, the above computation yields the standard error of s rather than the standard deviation. The standard error of s_A has been given as 0.035 and for the purposes of the computation, 5% errors in A and B were assumed so that $sd_A = 0.0416$ and $sd_B = 0.00078$.

Buffer. The deuterated forms of phosphoric acid, acetic acid and boric acid were needed for the buffer. The first two were both commercially available from Cambridge Isotope Labs, D_3PO_4 as an 85% w/w solution in D_2O (14.66 M) and CD_3COOD as a neat solution. D_3BO_3 was prepared by repeatedly (3×) dissolving 1.55 g (0.025 mol) of H_3BO_3 in 20 mL D_2O, heating for 15 min and removing the solvent. The resulting D_3BO_3 was dissolved in 50 mL D_2O to yield a 0.5 M solution of D_3BO_3 in D_2O.

To create the buffer, 680 μL of 85% D_3PO_4 in D_2O (0.010 mol), 570 μL of CD_3COOD (0.64 g, 0.010 mol), 20 mL of 0.5 M D_3BO_3, and 0.0412 g (0.00024 mol) of (trimethylsilyl) propionic-2,2,3,3-d_4 acid (TMSP, which was required as an NMR reference and as a concentration standard) were diluted to 200 mL with D_2O. The final

buffer was therefore 1.2 mM in TMSP and 50 mM each in D_3PO_4, CD_3COOD, and D_3BO_3.

Rather than adjusting the pD of the buffer after the sample was dissolved, the buffers were pre-adjusted to the desired pDs. Since the peptides in samples made for NMR spectroscopy were only ~ 1 mM, it was not expected that they would significantly alter the pD of the buffer upon their dissolution; in practice, the pD would vary at most by 0.3 pD units and usually much less. Four different versions of the buffer were needed, with pD 2.3, 5.1, 7.5, and 9.8. These were obtained adding the following to 30 mL aliquots:

pD 9.8: 350 µL of 40% w/w NaOD in D_2O (14.8 M)

pD 7.5: 350 µL of 40% w/w NaOD in D_2O (14.8 M) then 100 µL of 37% w/w DCl in D2O (12.3 M)

pD 5.1: 350 µL of 40% w/w NaOD in D_2O (14.8 M) then 230 µL of 37% w/w DCl in D2O (12.3 M)

pD 2.3: 350 µL of 40% w/w NaOD in D_2O (14.8 M) then 350 µL of 37% w/w DCl in D2O (12.3 M)

Note that the buffer was always adjusted to the highest desired pD first, then acidified to all the other pDs. This was done to maintain a roughly constant ionic strength. If the pD had been adjusted upward by addition of NaOD from the lowest desired pD, then each addition of base would increase the ionic strength as the undissociated buffer species (DX) were converted to their ionic sodium salts (NaX). When the pD is adjusted downward by addition of DCl, each addition of acid neutralizes the base forms of the buffer species thus removing an ionic species from solution, but also introduces NaCl ($NaX + DCl \rightarrow HX + NaCl$) thus maintaining the ionic strength. The only change in

ionic strength came as doubly charged phosphate was converted to singly charged phosphate, since ionic strength is related to the square of the charges of the species. Thus, at pD 9.8, 350 μL of 14.8 M NaOD were added to the base solution, introducing 170 mM sodium which was balanced by 50 mM acetate, 50 mM of doubly charged phosphate and 20 mM borate (the remaining borate being undissociated), yielding an ionic strength of 220 mM. At pD 2.3, 350 μL of 14.8 M NaOD and 350 μL of 12.3 M DCl were added, introducing 170 mM of sodium which was balanced by 144 mM chloride and 26 mM of singly charged phosphate, yielding an ionic strength of 170 mM (or 180 mM including the ~ 10 mM dissociated phosphoric acid).

Determination of Concentration for CD Samples. The buffer described above contained 1.2 mM TMSP not only for a chemical shift reference, but also as a concentration standard. The TMSP singlet, which is well separated from the sample peaks, was integrated and divided by 9 to obtain the peak area that corresponded to 1.2 mM of a single proton. The peak area for a single proton was then determined for the sample, and the sample concentration was calculated as follows:

$$\frac{\text{peak area for a single sample proton}}{\text{peak area for a single TMSP proton}} \times 1.20\ \text{mM} = \text{peptide concentration}$$

A very long delay (12 s) was used in NMR spectra obtained for concentration determination to ensure that integrals were accurate. Once the concentration of the NMR sample was known, it could be diluted (usually ~ 20 fold, from ~ 1 mM to ~ 50 μM) to a known concentration for CD spectroscopy. Concentrations determined in this way are expected to be accurate to within ± 20%.

Correcting $[\theta]_{222}$ in CD spectra of AcHel$_1$-peptide conjugates. For the te state of the template the molar ellipticity has been found to be[40] $[\theta]_{222,te} = -1{,}510$ deg cm^2 / dmol. Likewise, for the cs+ts states of the template[40] $[\theta]_{222,cs+ts} = -23{,}740$ deg cm^2 / dmol. The mole fractions of the AcHel$_1$-peptide conjugate residing in the te state and cs+ts states are[45]:

$$\chi_{te} = \left(\frac{(t/c) - A}{1 + (t/c)} \right); \chi_{cs+ts} = \left(\frac{1 + A}{1 + (t/c)} \right)$$

If both the t/c ratio and CD spectrum of an AcHel$_1$-peptide conjugate have been measured, then its $[\theta]_{222}$ can be corrected as follows:

$$[\theta]_{222,corrected} = [\theta]_{222,measured} + \chi_{te} \times [\theta]_{222,te} + \chi_{cs+ts} \times [\theta]_{222,cs+ts}$$

Peptide Synthesis. Peptides were prepared using an Advanced Chemtech model 90 peptide synthesizer. All amino acids were protected with the Fmoc group at the α amine and with the following protecting groups at the side chains: t-butyl for Glu and Thr, Boc for Lys, trityl for Gln and His, and 2,2,5,7,8 pentamethylchroman-6-sulfonyl for Arg. All peptides were prepared as the C-terminal amides using Knorr resin (2,4-dimethoxy-4'-(carboxymethyloxy)-benzhydrylamine linked to polystyrene) functionalized at 0.85 mmol/g. A typical procedure is as follows.

An Advanced Chemtech model 90 peptide synthesis vessel (a glass vessel equipped with a frit at the bottom) was charged with 200 mg (0.17 μmol of functionality) of Knorr resin, swollen for 1 h in DMF, treated with 2 mL of 30% piperidine twice for 15 min each, and washed 9× with 1 mL of DMF (washing entails filling the vessel with the

[45] Kemp, D. S.; Allen, T. J.; Oslick, S. L. *J. Am. Chem. Soc.* **1995**, *117*, 6641.

desired solvent, agitating for 30 s and then removing the solvent by forcing it through the frit with positive nitrogen pressure). The first amino acid, Fmoc-AA1-OH, was pre-activated by dissolving it (0.64 mmol, ~ 4 eq. relative to the resin) and 86 mg of hydroxybenzotriazole (HOBt, 0.64 mmol) in 2 mL of DMF, and treating the solution with 100 μL of diisopropyl carbodiimide (DIC, 80 mg, 0.64 mmol). After standing for 5 min, the solution was added to the Knorr resin in the peptide synthesis vessel. After 2 h, the activated peptide solution was removed and the resin washed 6× with DMF. The completeness of the reaction was checked by performing a qualitative ninhydrin test on the beads (see below). If the test was negative, the second amino acid was coupled just as the first amino acid was, starting from the piperidine deprotection. If the test was positive, then either the coupling was repeated (if it was clear that the coupling was nowhere near completion) or the peptide was capped by treatment with acetic anhydride (if the coupling was close to completion). If, in the course of synthesizing a given peptide, a positive ninhydrin test was obtained at any point, all subsequent coupling times were doubled to 4h. After the final non-template amino acid had been coupled and deprotected leaving a peptide with a free N-terminus bound to the resin, $AcHel_1CO_2H$ was coupled to the peptide. This proceeded similarly to the coupling of the other amino acids, except that the scale was smaller. Instead of 0.64 mmol, 0.08 mmol of $AcHel_1CO_2H$ (25 mg, ~ 0.5 eq relative to the resin) and 0.010 mmol (13 mg) of HOBt were dissolved in DMF and treated with 100 μL of DIC (80 mg, 0.64 mmol). This solution was then added to the resin as described above. Note that $AcHel_1CO_2H$ is the limiting reagent; at the completion of the coupling at most half of the resin bound peptide

could have $AcHel_1$ at the N-terminus. The remaining half would have a free amine at the N-terminus.

After the $AcHel_1$ coupling, the resin was washed 6× with methylene chloride and dried by first forcing a stream of nitrogen through the synthesis vessel and then placing the resin under vacuum. After 2 h under vacuum, approximately half of the resin was removed and placed in a new vessel, also equipped with a glass frit at the bottom. The resin was treated with a deprotection solution consisting of 2 mL trifluoroacetic acid, 0.1 mL thioanisole, 0.1 mL *m*-cresol, 0.1 mL H_2O, and 0.05 mL 1,2 ethanedithiol for 2 h. During this time the suspension changed from colorless to either a deep purple (when trityl groups were present) or to yellow (when no trityl groups were present). The solution was forced through the frit into 40 mL of ether, causing the peptide to precipitate. The precipitate was centrifuged, the solvent decanted from the pellet, and the pellet resuspended in 30 mL of ether. This was repeated twice, following which the pellet was dried under vacuum for 2 h. The pellet was dissolved in water and purified by preparative HPLC (gradient 5 to 100% CH_3CN over 40 min, remainder 0.1% TFA in water). Note that both the $AcHel_1$-peptide conjugate and the free amino peptide were present in the product from the deprotection/ resin cleavage reaction. These were easily separated by HPLC, however, since the free peptide eluted much earlier than the $AcHel_1$-peptide conjugate.

The calculated vs. observed masses and HPLC retention times for the $AcHel_1$-peptide conjugates and H-AKETAAAKFERQHA-NH_2 are summarized below in table 12.

Table 12. Summary of peptide characterization. Retention times are for the gradient mentioned in the text.

Peptide	observed m/z (calculated m/z)	retention time
*AcHel$_1$A-KETAAAKFERQHA-NH_2	$(M+2H)^{2+}$: 920 (920.0) $(M+3H)^{2+}$: 614 (613.6)	11.8 min
H-AKETAAAKFERQHA-NH_2	$(M+Na+H)^{2+}$: 1558 (1556.8)	9.8 min
AcHel$_1$A-KETAAAKA-NH_2	$(M-H)^-$: 1137 (1137.6)	10.4 min
AcHel$_1$A-ETAAAKA-NH_2	$(M+H)^+$: 1011 (1011.5)	12.0 min
AcHel$_1$A-TAAAKA-NH_2	$(M+H)^+$: 882 (882.4)	12.0 min
AcHel$_1$A-AAAKA-NH_2	$(M+H)^+$: 781 (781.4)	12.4 min
AcHel$_1$A-AAKA-NH_2	$(M+H)^+$: 710 (710.4)	9.5 min
AcHel$_1$*A-KFERQHA-NH_2	$(M+H)^+$: 1266 (1265.5)	14.7 min
AcHel$_1$*A-FERQHA-NH_2	$(M-H)^-$: 1136 (1136.5)	13.6 min
AcHel$_1$*A-ERQHA-NH_2	$(M-H)^-$: 989 (989.5)	10.4 min
AcHel$_1$*A-RQHA-NH_2	$(M+H)^+$: 862 (862.4)	9.7 min
AcHel$_1$A-QHA-NH_2	$(M+H)^+$: 705 (705.3)	9.8 min
AcHel$_1$A-HA-NH_2	$(M+H)^+$: 577 (577.2)	9.7 min

Qualitative Ninhydrin Test.[46] To a small test tube was added 10 μL of a solution of 0.5 g ninhydrin in n-butanol, 10 μL of a solution of 80 g phenol in 20 mL n-butanol, and 20 μL of a solution of 2 mL 10 mM KCN (aq.) in 98 mL pyridine. To this was added about 50 resin beads, the mixture was heated over a steam bath for 2 min, and the solution

[46] Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.

checked for color. An incomplete peptide coupling was indicated by the beads retaining a dark purple color.

Preparation of *AcHel$_1$CO$_2$H. To 0.06 g of HHel$_1$CO$_2$Me (0.22 mmol) in 2 mL of methylene chloride was added 50 μL of *Ac$_2$O (doubly labeled with ^{13}C: (^{13}CH$_3$^{13}CO)$_2$O, 0.51 mmol). After 2 h, the reaction was quenched with 0.5 mL methanol, the solvent was removed, the residue was redissolved in water and purified by preparative HPLC (gradient 5% to 100% CH_3CN over 40 min, remainder 0.1% TFA in water) to yield *AcHel$_1$CO$_2$Me (0.07g, 0.22 mmol, quantitative yield), which was then saponified and purified as described in the literature[39] to yield *AcHel$_1$CO$_2$H. *AcHel$_1$CO$_2$H was identical chromatographically and spectroscopically to AcHel$_1$CO$_2$H, except that the t and c state acetyl peaks in the ^{1}H-NMR spectrum were doublets of doublets each with coupling constants J = 129 Hz for the 1 bond ^{13}C-^{1}H coupln and 1.7 Hz for the 3 bond ^{13}C-^{1}H coupling and the proton decoupled ^{13}C-NMR spectrum had exceptionally strong peaks at 174.1 ppm (d, J = 50 Hz) and 174.2 ppm (d, J = 50 Hz) for the t and c state acetyl carbonyl peaks and 22.2 ppm (d, J = 50 Hz) and 21.7 ppm (d, J = 50 Hz) for the t and c state acetyl methyl groups.

Appendix. The State Sum of H-KETAAAKFERQHA-NH$_2$

The sum of equilibrium constants for H-KETAAAKFERQHA-NH$_2$ in the Zimm-Bragg model can be calculated by the methods detailed in chapter 2:

$$Q = \alpha \times \mathbf{M}_T \times \mathbf{M}_A \times \mathbf{M}_A \times \mathbf{M}_A \times \mathbf{M}_{K7} \times \mathbf{M}_F \times \mathbf{M}_{E9} \times \mathbf{M}_R \times \mathbf{M}_Q \times \mathbf{M}_H \times \mathbf{M}_A \times \omega$$

Recall that Lys 1 and Glu 2 do not contribute to the matrix product since they are the first two residues. This matrix product is shown below.

$$
\begin{aligned}
&1 + 4\,\sigma\,S_A + 2\,\sigma\,S_A^2 + 3\,\sigma^2\,S_A^2 + \sigma\,S_A^3 + 2\,\sigma^2\,S_A^3 + \sigma^2\,S_A^4 + \sigma\,S_{E9} + 3\,\sigma^2\,S_A\,S_{E9} + \sigma^2\,S_A^2\,S_{E9} + 2\,\sigma^3\,S_A^2\,S_{E9} + \\
&\sigma^3\,S_A^3\,S_{E9} + \sigma\,S_F + 2\,\sigma^2\,S_A\,S_F + \sigma^3\,S_A^2\,S_F + \sigma\,S_{E9}\,S_F + 2\,\sigma^2\,S_A\,S_{E9}\,S_F + \sigma^3\,S_A^2\,S_{E9}\,S_F + \sigma\,S_H + \sigma\,S_A\,S_H + \\
&3\,\sigma^2\,S_A\,S_H + 5\,\sigma^2\,S_A^2\,S_H + 3\,\sigma^2\,S_A^3\,S_H + \sigma^2\,S_A^4\,S_H + \sigma^2\,S_F\,S_H + \sigma^2\,S_A\,S_F\,S_H + \sigma^3\,S_A\,S_F\,S_H + \sigma^3\,S_A^2\,S_F\,S_H + \\
&\sigma\,S_{K7} + \sigma\,S_A\,S_{K7} + \sigma^2\,S_A\,S_{K7} + \sigma\,S_A^2\,S_{K7} + \sigma^2\,S_A^2\,S_{K7} + \sigma\,S_A^3\,S_{K7} + \sigma^2\,S_A^3\,S_{K7} + \sigma^2\,S_A^4\,S_{K7} + \sigma\,S_F\,S_{K7} + \\
&\sigma\,S_A\,S_F\,S_{K7} + \sigma^2\,S_A\,S_F\,S_{K7} + \sigma\,S_A^2\,S_F\,S_{K7} + \sigma^2\,S_A^2\,S_F\,S_{K7} + \sigma\,S_A^3\,S_F\,S_{K7} + \sigma^2\,S_A^3\,S_F\,S_{K7} + \sigma^2\,S_A^4\,S_F\,S_{K7} + \\
&\sigma\,S_{E9}\,S_F\,S_{K7} + \sigma\,S_A\,S_{E9}\,S_F\,S_{K7} + \sigma^2\,S_A\,S_{E9}\,S_F\,S_{K7} + \sigma\,S_A^2\,S_{E9}\,S_F\,S_{K7} + \sigma^2\,S_A^2\,S_{E9}\,S_F\,S_{K7} + \\
&\sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7} + \sigma^2\,S_A^3\,S_{E9}\,S_F\,S_{K7} + \sigma^2\,S_A^4\,S_{E9}\,S_F\,S_{K7} + \sigma^2\,S_H\,S_{K7} + 2\,\sigma^2\,S_A\,S_H\,S_{K7} + 2\,\sigma^2\,S_A^2\,S_H\,S_{K7} + \\
&2\,\sigma^2\,S_A^3\,S_H\,S_{K7} + \sigma^2\,S_A^4\,S_H\,S_{K7} + \sigma^2\,S_F\,S_H\,S_{K7} + 2\,\sigma^2\,S_A\,S_F\,S_H\,S_{K7} + 2\,\sigma^2\,S_A^2\,S_F\,S_H\,S_{K7} + \\
&2\,\sigma^2\,S_A^3\,S_F\,S_H\,S_{K7} + \sigma^2\,S_A^4\,S_F\,S_H\,S_{K7} + \sigma\,S_Q + 3\,\sigma^2\,S_A\,S_Q + 2\,\sigma^2\,S_A^2\,S_Q + \sigma^2\,S_A^3\,S_Q + \sigma\,S_H\,S_Q + \sigma\,S_A\,S_H\,S_Q + \\
&3\,\sigma^2\,S_A\,S_H\,S_Q + 5\,\sigma^2\,S_A^2\,S_H\,S_Q + 3\,\sigma^2\,S_A^3\,S_H\,S_Q + \sigma^2\,S_A^4\,S_H\,S_Q + \sigma^2\,S_{K7}\,S_Q + \sigma^2\,S_A\,S_{K7}\,S_Q + \sigma^2\,S_A^2\,S_{K7}\,S_Q + \\
&\sigma^2\,S_A^3\,S_{K7}\,S_Q + \sigma^2\,S_H\,S_{K7}\,S_Q + 2\,\sigma^2\,S_A\,S_H\,S_{K7}\,S_Q + 2\,\sigma^2\,S_A^2\,S_H\,S_{K7}\,S_Q + 2\,\sigma^2\,S_A^3\,S_H\,S_{K7}\,S_Q + \\
&\sigma^2\,S_A^4\,S_H\,S_{K7}\,S_Q + \sigma\,S_R + 3\,\sigma^2\,S_A\,S_R + 2\,\sigma^2\,S_A^2\,S_R + \sigma^2\,S_A^3\,S_R + \sigma\,S_{E9}\,S_R + 2\,\sigma^2\,S_A\,S_{E9}\,S_R + \sigma^2\,S_A^2\,S_{E9}\,S_R + \\
&\sigma\,S_{E9}\,S_F\,S_R + \sigma^2\,S_A\,S_{E9}\,S_F\,S_R + \sigma\,S_{E9}\,S_F\,S_{K7}\,S_R + \sigma\,S_A\,S_{E9}\,S_F\,S_{K7}\,S_R + \sigma\,S_A^2\,S_{E9}\,S_F\,S_{K7}\,S_R + \\
&\sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7}\,S_R + \sigma\,S_Q\,S_R + 3\,\sigma^2\,S_A\,S_Q\,S_R + 2\,\sigma^2\,S_A^2\,S_Q\,S_R + \sigma^2\,S_A^3\,S_Q\,S_R + \sigma\,S_{E9}\,S_Q\,S_R + \\
&2\,\sigma^2\,S_A\,S_{E9}\,S_Q\,S_R + \sigma^2\,S_A^2\,S_{E9}\,S_Q\,S_R + \sigma\,S_{E9}\,S_F\,S_Q\,S_R + \sigma^2\,S_A\,S_{E9}\,S_F\,S_Q\,S_R + \sigma\,S_H\,S_Q\,S_R + \\
&\sigma\,S_A\,S_H\,S_Q\,S_R + 3\,\sigma^2\,S_A\,S_H\,S_Q\,S_R + 5\,\sigma^2\,S_A^2\,S_H\,S_Q\,S_R + 3\,\sigma^2\,S_A^3\,S_H\,S_Q\,S_R + \sigma^2\,S_A^4\,S_H\,S_Q\,S_R + \\
&\sigma\,S_{E9}\,S_H\,S_Q\,S_R + \sigma\,S_A\,S_{E9}\,S_H\,S_Q\,S_R + 2\,\sigma^2\,S_A\,S_{E9}\,S_H\,S_Q\,S_R + 3\,\sigma^2\,S_A^2\,S_{E9}\,S_H\,S_Q\,S_R + \sigma^2\,S_A^3\,S_{E9}\,S_H\,S_Q\,S_R + \\
&\sigma\,S_{E9}\,S_F\,S_H\,S_Q\,S_R + \sigma\,S_A\,S_{E9}\,S_F\,S_H\,S_Q\,S_R + \sigma^2\,S_A\,S_{E9}\,S_F\,S_H\,S_Q\,S_R + \sigma^2\,S_A^2\,S_{E9}\,S_F\,S_H\,S_Q\,S_R + \\
&\sigma\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R + \sigma\,S_A\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R + \sigma\,S_A^2\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R + \sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R + \\
&\sigma\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R + 2\,\sigma\,S_A\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R + 2\,\sigma\,S_A^2\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R + \\
&2\,\sigma\,S_A^3\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R + \sigma\,S_A^4\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R + \sigma\,S_T + \sigma\,S_A\,S_T + \sigma^2\,S_A\,S_T + \sigma\,S_A^2\,S_T + \\
&\sigma^2\,S_A^2\,S_T + \sigma\,S_A^3\,S_T + \sigma^2\,S_A^3\,S_T + \sigma^2\,S_A^4\,S_T + \sigma^2\,S_{E9}\,S_T + \sigma^2\,S_A\,S_{E9}\,S_T + \sigma^3\,S_A\,S_{E9}\,S_T + \sigma^2\,S_A^2\,S_{E9}\,S_T + \\
&\sigma^3\,S_A^2\,S_{E9}\,S_T + \sigma^3\,S_A^3\,S_{E9}\,S_T + \sigma^2\,S_F\,S_T + \sigma^2\,S_A\,S_F\,S_T + \sigma^3\,S_A\,S_F\,S_T + \sigma^3\,S_A^2\,S_F\,S_T + \sigma^2\,S_{E9}\,S_F\,S_T + \\
&\sigma^2\,S_A\,S_{E9}\,S_F\,S_T + \sigma^3\,S_A\,S_{E9}\,S_F\,S_T + \sigma^3\,S_A^2\,S_{E9}\,S_F\,S_T + \sigma^2\,S_H\,S_T + 2\,\sigma^2\,S_A\,S_H\,S_T + 2\,\sigma^2\,S_A^2\,S_H\,S_T + \\
&2\,\sigma^2\,S_A^3\,S_H\,S_T + \sigma^2\,S_A^4\,S_H\,S_T + \sigma^3\,S_F\,S_H\,S_T + 2\,\sigma^3\,S_A\,S_F\,S_H\,S_T + \sigma^3\,S_A^2\,S_F\,S_H\,S_T + \sigma^2\,S_{K7}\,S_T + \\
&\sigma^3\,S_A\,S_{K7}\,S_T + \sigma\,S_A^3\,S_{K7}\,S_T + \sigma^2\,S_A^4\,S_{K7}\,S_T + \sigma^2\,S_F\,S_{K7}\,S_T + \sigma^3\,S_A\,S_F\,S_{K7}\,S_T + \sigma\,S_A^3\,S_F\,S_{K7}\,S_T + \\
&\sigma^2\,S_A^4\,S_F\,S_{K7}\,S_T + \sigma^2\,S_{E9}\,S_F\,S_{K7}\,S_T + \sigma^3\,S_A\,S_{E9}\,S_F\,S_{K7}\,S_T + \sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7}\,S_T + \sigma^2\,S_A^4\,S_{E9}\,S_F\,S_{K7}\,S_T + \\
&\sigma^3\,S_H\,S_{K7}\,S_T + \sigma^3\,S_A\,S_H\,S_{K7}\,S_T + \sigma^2\,S_A^3\,S_H\,S_{K7}\,S_T + \sigma^2\,S_A^4\,S_H\,S_{K7}\,S_T + \sigma^3\,S_F\,S_H\,S_{K7}\,S_T + \\
&\sigma^3\,S_A\,S_F\,S_H\,S_{K7}\,S_T + \sigma^2\,S_A^3\,S_F\,S_H\,S_{K7}\,S_T + \sigma^2\,S_A^4\,S_F\,S_H\,S_{K7}\,S_T + \sigma^2\,S_Q\,S_T + \sigma^2\,S_A\,S_Q\,S_T + \sigma^2\,S_A^2\,S_Q\,S_T + \\
&\sigma^2\,S_A^3\,S_Q\,S_T + \sigma^2\,S_H\,S_Q\,S_T + 2\,\sigma^2\,S_A\,S_H\,S_Q\,S_T + 2\,\sigma^2\,S_A^2\,S_H\,S_Q\,S_T + 2\,\sigma^2\,S_A^3\,S_H\,S_Q\,S_T + \sigma^2\,S_A^4\,S_H\,S_Q\,S_T + \\
&\sigma^3\,S_{K7}\,S_Q\,S_T + \sigma^2\,S_A^3\,S_{K7}\,S_Q\,S_T + \sigma^3\,S_H\,S_{K7}\,S_Q\,S_T + \sigma^3\,S_A\,S_H\,S_{K7}\,S_Q\,S_T + \sigma^2\,S_A^3\,S_H\,S_{K7}\,S_Q\,S_T + \\
&\sigma^2\,S_A^4\,S_H\,S_{K7}\,S_Q\,S_T + \sigma^2\,S_R\,S_T + \sigma^2\,S_A\,S_R\,S_T + \sigma^2\,S_A^2\,S_R\,S_T + \sigma^2\,S_A^3\,S_R\,S_T + \sigma^2\,S_{E9}\,S_R\,S_T + \\
&\sigma^2\,S_A\,S_{E9}\,S_R\,S_T + \sigma^2\,S_A^2\,S_{E9}\,S_R\,S_T + \sigma^2\,S_{E9}\,S_F\,S_R\,S_T + \sigma^2\,S_A\,S_{E9}\,S_F\,S_R\,S_T + \sigma^2\,S_{E9}\,S_F\,S_{K7}\,S_R\,S_T + \\
&\sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7}\,S_R\,S_T + \sigma^2\,S_Q\,S_R\,S_T + \sigma^2\,S_A\,S_Q\,S_R\,S_T + \sigma^2\,S_A^2\,S_Q\,S_R\,S_T + \sigma^2\,S_A^3\,S_Q\,S_R\,S_T + \\
&\sigma^2\,S_{E9}\,S_Q\,S_R\,S_T + \sigma^2\,S_A\,S_{E9}\,S_Q\,S_R\,S_T + \sigma^2\,S_A^2\,S_{E9}\,S_Q\,S_R\,S_T + \sigma^2\,S_{E9}\,S_F\,S_Q\,S_R\,S_T + \\
&\sigma^2\,S_A\,S_{E9}\,S_F\,S_Q\,S_R\,S_T + \sigma^2\,S_H\,S_Q\,S_R\,S_T + 2\,\sigma^2\,S_A\,S_H\,S_Q\,S_R\,S_T + 2\,\sigma^2\,S_A^2\,S_H\,S_Q\,S_R\,S_T + \\
&2\,\sigma^2\,S_A^3\,S_H\,S_Q\,S_R\,S_T + \sigma^2\,S_A^4\,S_H\,S_Q\,S_R\,S_T + \sigma^2\,S_{E9}\,S_H\,S_Q\,S_R\,S_T + 2\,\sigma^2\,S_A\,S_{E9}\,S_H\,S_Q\,S_R\,S_T + \\
&2\,\sigma^2\,S_A^2\,S_{E9}\,S_H\,S_Q\,S_R\,S_T + \sigma^2\,S_A^3\,S_{E9}\,S_H\,S_Q\,S_R\,S_T + \sigma^2\,S_{E9}\,S_F\,S_H\,S_Q\,S_R\,S_T + 2\,\sigma^2\,S_A\,S_{E9}\,S_F\,S_H\,S_Q\,S_R\,S_T + \\
&\sigma^2\,S_A^2\,S_{E9}\,S_F\,S_H\,S_Q\,S_R\,S_T + \sigma^2\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R\,S_T + \sigma\,S_A^3\,S_{E9}\,S_F\,S_{K7}\,S_Q\,S_R\,S_T + \\
&\sigma^2\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R\,S_T + \sigma^2\,S_A\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R\,S_T + \sigma\,S_A^3\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R\,S_T + \\
&\sigma\,S_A^4\,S_{E9}\,S_F\,S_H\,S_{K7}\,S_Q\,S_R\,S_T
\end{aligned}
$$

Chapter 4. A New Template for the Initiation of Peptide Helices

4.1 Introduction

The methods for inducing helicity in peptides include constraining a peptide via its side chains,[1,2,3,4,5,6,7,8,9,10] via its main chain,[11] or using N-terminal templates[12,13] such as $AcHel_1$.[14,15,16,17] $AcHel_1$, in addition to being a helix initiator, is an effective tool[18,19,20,21,22] for the study of peptide helicity by virtue of the reporting function of its trans ⇄ cis equilibrium (see section 2.5). This feature was used in the previous chapter to study the helicity of the ribonuclease C-peptide, and was used in chapter 1 to study

---

[1] Ösapay, G.; Taylor, J. W. *J. Am. Chem. Soc.* **1990**, *112*, 6046.
[2] Ösapay, G.; Taylor, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 6966.
[3] Bracken, C.; Gulyás, J.; Taylor, J. W.; Baum, J. *J. Am. Chem. Soc.* **1994**, *116*, 6431.
[4] Jackson, D. Y.; King, D. S.; Chmielewski, J.; Singh, S.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 9391.
[5] Chorev, M.; Roubini, E.; McKee, R. L.; Gibbons, S. W.; Goldman, M. E.; Caulfield, M. P.; Rosenblatt, M. *Biochemistry* **1991**, *30*, 5968.
[6] Zhou, H. X.; Hull, L. A.; Kallenbach, N. R.; Mayne, L.; Bai, Y.; Englander, S. W. *J. Am. Chem. Soc.* **1994**, *116*, 6482.
[7] Phelan, J. C.; Skelton, N. J.; Braisted, A. C.; McDowell, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 455.
[8] Ghadiri, M. R.; Fernholz, A. K. *J. Am. Chem. Soc.* **1990**, *112*, 9633.
[9] Ghadiri, M. R.; Choi, C. *J. Am. Chem. Soc.* **1990**, *112*, 1630.
[10] Ruan, F.; Chen, Y.; Hopkins, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 9403.
[11] Satterthwiat, A. C.; Arrhenius, T.; Hagopian, R. A.; Zavala, F.; Nussenzwieg, V.; Lerner, R. A. *Phil. Trans. R. Soc. Lond. B* **1989**, *323*, 565.
[12] Müller, K.; Obercht, D.; Knierzinger, A.; Stankovic, C.; Spiegler, C.; Bannwarth, W.; Trzeciak, A.; Englert, G.; Labhardt, A. M.; Scoenholzer, P. *Perspect. Med. Chem.* **1993**, 513.
[13] Austin, R. E.; Maplestone, R. A.; Sefler, A. M.; Liu, K.; Hruzewicz, W. N.; Liu, C. W.; Cho, H. S.; Wemmer, D. E.; Bartlett, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6461.
[14] Kemp, D. S.; Curran, T. P.; Davis, W. M.; Boyd, J. G.; Muendel, C. C. *J. Org. Chem.* **1991**, *56*, 6672.
[15] Kemp, D. S.; Curran, T. P.; Boyd, J. G.; Allen, T. *J. Org. Chem.* **1991**, *56*, 6683.
[16] Kemp, D. S.; Boyd, J. G.; Muendel, C. C. *Nature* **1991**, *352*, 451.
[17] Kemp, D. S.; Allen, T. J.; Oslick, S. O. *J. Am. Chem. Soc.* **1995**, *117*, 6641.
[18] Kemp, D. S.; Allen, T. J.; Oslick, S. O.; Boyd, J. G. *J. Am. Chem. Soc.* **1996**, *118*, 4240.
[19] Kemp, D. S.; Oslick, S. O.; Allen, T. J. *J. Am. Chem. Soc.* **1996**, *118*, 4249.
[20] Zerkowski, J. A.; Powers, E. T.; Kemp, D. S. *J. Am. Chem. Soc.* **1997**, *119*, 1153.
[21] Gorebke, K.; Penold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2025.
[22] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 12234.

hydrogen bonding in variants of $AcHel_l$. However, for two reasons the very equilibrium that makes $AcHel_l$ such a useful tool in some applications thwarts its use in others. First, the substantial population of the unproductive cis state diminishes the efficiency of $AcHel_l$ as a helix initiator. Second, the most important method for quantifying the helicity at individual sites along a peptide chain, the amide hydrogen exchange technique (introduced in section 2.5), cannot be used with $AcHel_l$-peptide conjugates because the trans $\rightleftarrows$ cis interconversion occurs on the same time scale as amide hydrogen exchange reactions. The interconversion influences the exchange kinetics, setting a lower limit on the magnitude of the exchange rate decreases that can be observed, and complicating the simple relationship between exchange rate and local fractional helicity given in section 2.5.[18] This chapter describes the design, synthesis, and characterization of an $AcHel_l$ analog intended both to be a stronger helix initiator and to permit hydrogen exchange studies.

4.2 Design

Developing a template based on the Hel_l skeleton that can be used with the amide hydrogen exchange technique requires only that the acetyl group of $AcHel_l$ be replaced by another good hydrogen bond acceptor that does not undergo a trans $\rightleftarrows$ cis interconversion. This aspect of this chapter's goals does not pose a significant design challenge, but what features of $AcHel_l$ can be changed to optimize helix initiation? Since helix formation only proceeds in $AcHel_l$ after hydrogen bond formation between the amide at the template-peptide junction and the acetamide carbonyl (see section 2.4), it is

natural to replace the acetyl group with other, potentially more potent, hydrogen bond acceptors. Indeed, in previous work, the acetyl group has been replaced by four neutral acceptors[23]: pivaloyl ($(CH_3)_3CO$-), methanesulfonamide (CH_3SO_2-), methoxycarbonyl (CH_3OCO-), and dimethyl urea ($(CH_3)_2NCO$-). Unfortunately, none of these were significantly more adept at promoting helicity in X-Hel_1-peptide conjugates than the acetyl group. How else could the N-terminal group be tuned?

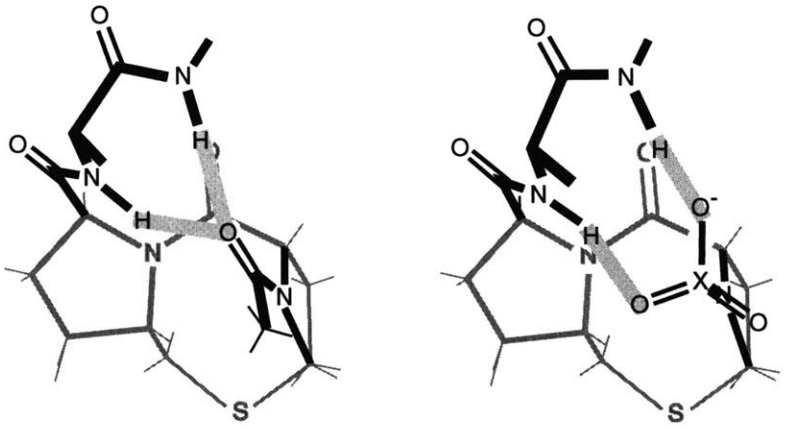
Two design criteria can be added to replacing AcHel_1's acetamide with a hydrogen bond acceptor that does not undergo a trans $\rightleftarrows$ cis equilibrium. First, the new acceptor should be negatively charged. In the best case, the charged acceptor would form stronger hydrogen bonds than its neutral counterparts, as has been suggested it should on the basis of protein mutagenesis experiments (see section 1.2.4b). Alternatively, in light of what was observed for charged donors in chapter 1, the charged acceptor might not form stronger hydrogen bonds, but might interact strongly with the N-terminal amide NHs via charge-dipole interactions. This could be enough to cause the amide in question to assume the geometry required for helix initiation. In addition, a negative charge at the N1 position of the template might interact with other sources of positive charge. It has been shown that tethering negatively charged functional groups such as sulfonates[24] or carboxylates[25] to a peptide's N-terminus via alkane linkers favors helix formation. Whether this effect is attributed to an interaction between the charge and the positive end of the helix macrodipole,[25,26] or to interactions between the charge and the microdipoles

---

[23] Renold, P.; Kemp, D. S. Unpublished work.
[24] Forood, B.; Reddy, H. K.; Nambiar, K. P. *J. Am. Chem. Soc.* **1994**, *116*, 6935.
[25] Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563.
[26] Hol, W. G. J. *Prog. Biophys. Molec. Biol.* **1985**, *45*, 149.

of nearby polar groups,[27] placing a negative charge exactly at the peptide N-terminus, so that the interaction is less solvent shielded, ought to substantially magnify it. Second, the new acceptor should have multiple hydrogen bond acceptor sites. As shown on the left-hand side of figure 1, in the hydrogen bond between the first amide NH and the acetamide, the NH is directed at the carbonyl from outside of the plane in which the carbonyl's lone pairs lie (a π-type hydrogen bond). Furthermore, the first two amide NH groups must share the acetamide carbonyl as a hydrogen bonding partner in $AcHel_1$-peptide conjugates. The hydrogen bonding pattern might be improved if the acetyl were replaced by a group with three tetrahedrally disposed oxygen atoms (XO_3^-), as shown on the right-hand side of figure 1. The first amide could then approach its hydrogen bonding partner from a better angle, and the first two amides could each hydrogen bond to a different atom.

Figure 1. The hydrogen bonding of the first two amides in $AcHel_1$ vs. XO_3Hel_1, a derivative of $AcHel_1$ in which the acetyl is replaced with multiple hydrogen bond acceptors (hydrogen bonds are indicated with thick gray lines).

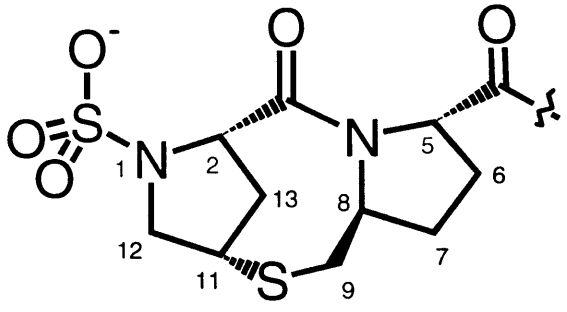


$AcHel_1$-peptide XO_3Hel_1-peptide

[27] Tidor, B. *Proteins: Struct. Funct. Genet.* **1994**, *19*, 310.

Only two functional groups can fulfill all of the above requirements: the phosphoramidic acid (H_2O_3P-N) and the sulfamic acid (HO_3S-N). Neither of these would have any unproductive conformations due to rotation about the P-N or S-N bonds, both would be negatively charged over much of the aqueous pH range (first pK_a = 2.74 for $H_2PO_3NH_2$[28] and 1.1 for HO_3SNH_2[29]), and both have three tetrahedrally disposed oxygens. However, phosphoramidic acids rapidly decompose at low pH while sulfamic acids, especially when they are N-substituted, do not. The half-life of $Et_2NSO_3^-Na^+$ at pH 0 and 90 °C is about 17 hours,[30] and it ought to be much longer at higher pHs and lower temperatures. The remainder of this chapter therefore focuses on the preparation and properties of SO_3Hel_1, the sulfamic acid analog of $AcHel_1$, shown in figure 2. (Note that SO_3Hel_1 is shown as the anion in figure 2 since it should be charged under most of the conditions in which it is used.)

Figure 2. SO_3Hel_1, in which a negatively charged SO_3^- replaces the acetyl of $AcHel_1$.



[28] Peakcock, C. J.; Nickless, G. Z. *Naturforch. A* **1969**, *24*, 245.
[29] Benson. G. A.; Spillane, W. J. *Chem. Rev.* **1980**, *80*, 151.

4.3 Results

4.3.1 Preparation and Stability of SO_3Hel_1-peptide Conjugates

Two strategies for preparing SO_3Hel_1-peptide conjugates were envisioned. Either the SO_3Hel_1 group could be incorporated as a whole during solid phase peptide synthesis, or the Hel_1 unit could be sulfamated at N1 (see figure 2 for numbering scheme) after the template-peptide conjugate had been deprotected and cleaved from the resin. The first strategy requires $SO_3Hel_1CO_2H$, which could be prepared either by sulfamation of $HHel_1CO_2Me$ followed by hydrolysis or by direct sulfamation of $HHel_1CO_2H$ (both $HHel_1CO_2Me$ and $HHel_1CO_2H$ were prepared prior to the work in this thesis[23]). The sulfamation of amines is an old reaction,[31] and can be effected under a number of solvent conditions, including aqueous solution.[32] Sulfamation probably occurs by direct displacement of pyridine from the SO_3 sulfur by the nucleophilic amine.[33] The sulfamation of $HHel_1CO_2Me$ in SO_3-pyridine complex in DMF with 3 eq. diisopropylethylamine, and of $HHel_1CO_2H$ in 10% aqueous Na_2CO_3 both yielded inseparable mixtures as judged by NMR. These probably consisted of the free amine starting materials together with the sulfamated products. It is possible that as the sulfamation progressed under these conditions, acid was generated that protonated the remaining amine making it unavailable for further reaction. Because the preparation of the necessary $SO_3Hel_1CO_2H$ appeared problematic, efforts in this direction were stopped.

---

[30] Spillane, W. J.; Scott, F. L.; Goggin, C. B. *Int. J. Sulfur Chem. A* **1971**, *1*, 223.
[31] Beilstein, F.; Wiegand, E. *Ber.* **1883**, *16*, 1264.
[32] Gilbert, E. E. *Chem. Rev.* **1962**, *62*, 549.

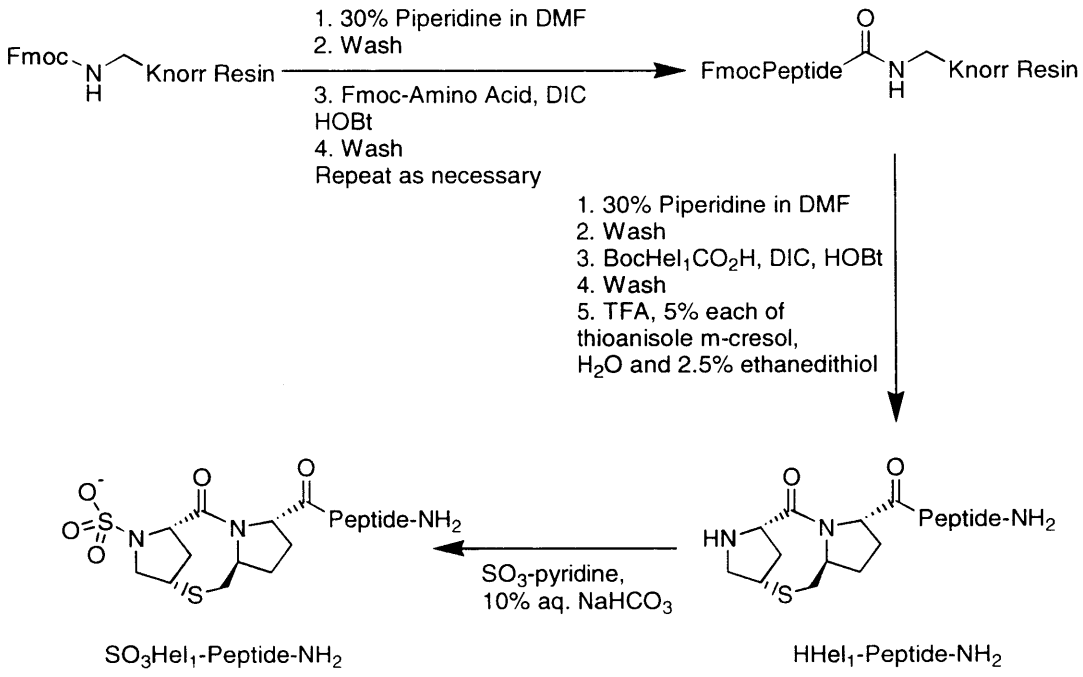
The second strategy listed above requires the selective sulfamation of an α amine in the presence of potentially reactive amino acid side chain nucleophiles. The sulfamation of amino acids was studied over 60 years ago,[34] and it was found that most amino acids could be monosulfamated at the α amine with SO_3-pyridine complex in mild base (aqueous carbonate). The only residues whose side chains had nucleophiles reactive enough to be sulfated under these conditions were Cys, His, Lys, and Tyr, which were respectively sulfated at their thiol, imidazole, amine, and phenol, as well as their α amines. Given these observations it was reasoned that an $HHel_1$-peptide conjugate, in which the template is unsubstituted at N1, could be selectively monosulfamated at the template amine provided that the peptide did not contain Cys, His, Lys, or Tyr. The preparation of SO_3Hel_1-peptide conjugates as C-terminal amides proceeded as shown in scheme 1. Peptides were elongated from Knorr resin by the Fmoc-based solid-phase peptide synthesis techniques described in chapter 3. $BocHel_1$, in which the removable Boc group replaces the acetyl group of $AcHel_1$, was attached at the end of the peptide instead of $AcHel_1$ (the synthesis of $BocHel_1$, achieved in prior work in these labs,[23] is a simple variation on the synthesis of $AcHel_1$). Upon deprotection and cleavage of the peptide from the resin, the Boc group was removed from the template to yield the fully unprotected template-peptide conjugate. This was sulfamated by dissolving the peptide at between 2 and 5 mg / mL in 10% aqueous sodium bicarbonate and treating the solution with a large excess (~ 100 equivalents) of SO_3-pyridine complex for 12 h at 25 °C. This reaction never went to completion. Even after 12h, unreacted starting material was apparent when the progress of the reaction was checked by analytical HPLC. However,

---

[33] Bourne, N.; Hopkins, A.; Williams, A. *J. Am. Chem. Soc.* **1985**, *107*, 4327.

unlike the cases in the preceding paragraph, the unreacted $HHel_1$-peptide conjugates were easily separable from the desired SO_3Hel_1-peptide conjugates, which were typically obtained in 50 – 60% isolated yields. This method has been used to prepare over a dozen SO_3Hel_1-peptide conjugates containing Ala, Glu, Phe, Leu, Gln, Arg, and Ser.

Scheme 1.



The stability of SO_3Hel_1-peptide conjugates in aqueous solution is excellent. Spectroscopic measurements and hydrogen exchange experiments (see chapter 5) could be carried out without difficulty, and no hydrolysis of SO_3Hel_1-peptide conjugates has been detected by NMR for periods up to a month in moderately acidic solutions (pH 3).

---

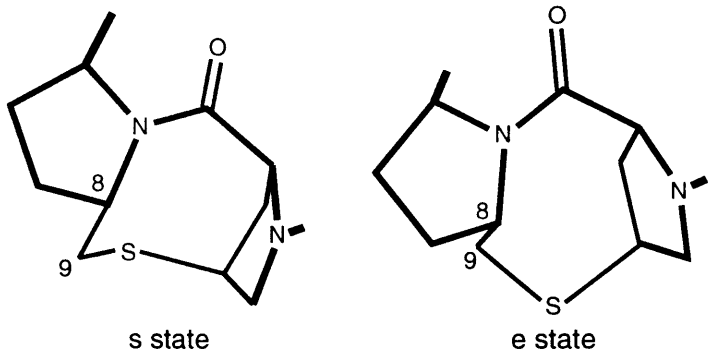
[34] Baumgarten, P.; Marggraff, I.; Dammann, E. *Hoppe-Seyler's Z. Physiol. Chem.* **1932**, *209*, 145.

4.3.2 The Conformation of SO_3Hel_1

4.3.2a The staggered (s) ⇄ eclipsed (e) equilibrium about the C8-C9 bond

Two conformational equilibria in the $AcHel_1$ framework influence helix initiation (see chapter 2 for a complete discussion). The first is the trans ⇄ cis equilibrium of the acetamide that has already been mentioned in this chapter. The second is the staggered ⇄ eclipsed equilibrium about the C8-C9 bond within the trans state (ts ⇄ te), illustrated in figure 3, whose importance results from the following.[17]

Figure 3. The s and e states in the Hel_1 skeleton.



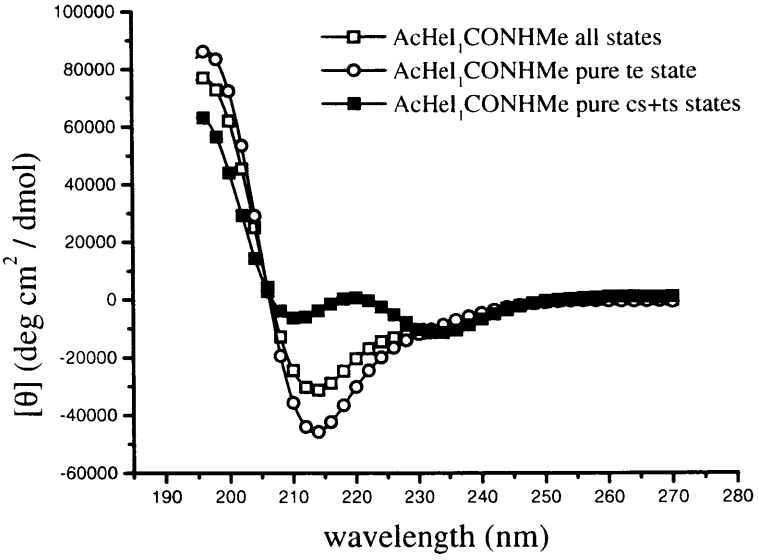
Helix induction by $AcHel_1$ must be preceded by hydrogen bonding between the amide NH at the template-peptide junction and the acetamide carbonyl oxygen. This hydrogen bond is stronger, probably because it is shorter, in the te state than in the ts state (see chapter 1 for a detailed discussion of the effect of this equilibrium on hydrogen bonding).

Therefore, helix initiation by $AcHel_1$ can only occur through a concomitant increase in the population of the te state. There should be no equivalent of the trans ⇄ cis equilibrium in SO_3Hel_1, but since the cyclic core of SO_3Hel_1 is the same as in $AcHel_1$, it should have an staggered (s) ⇄ eclipsed (e) equilibrium. This aspect of the conformation of SO_3Hel_1 is addressed in the present section.

Far UV circular dichroism (CD) is a technique for measuring how local chiral environments perturb ultraviolet absorbances. CD chromophores are therefore UV chromophores, and in the case of $X\text{-}Hel_1$ derivatives, these include the amides and the thioether group. Since the dispositions of these groups differ between the s and the e conformations, the CD spectra of $X\text{-}Hel_1$ derivatives should be different depending on the position of the s ⇄ e equilibrium. The CD spectrum of $AcHel_1CONHMe$[35] is shown in figure 4. Also shown in this figure are the limiting CD spectra expected for the pure te and the pure cs+ts conformations of $AcHel_1CONHMe$,[17,35] extrapolated from the dependencies of this compound's CD spectrum and its te state population on TFE concentration. The cs+ts spectrum has a single deep minimum at around 214 nm, while the te spectrum has two shallow minima at 210 and 233 nm and a small maximum at 222nm.

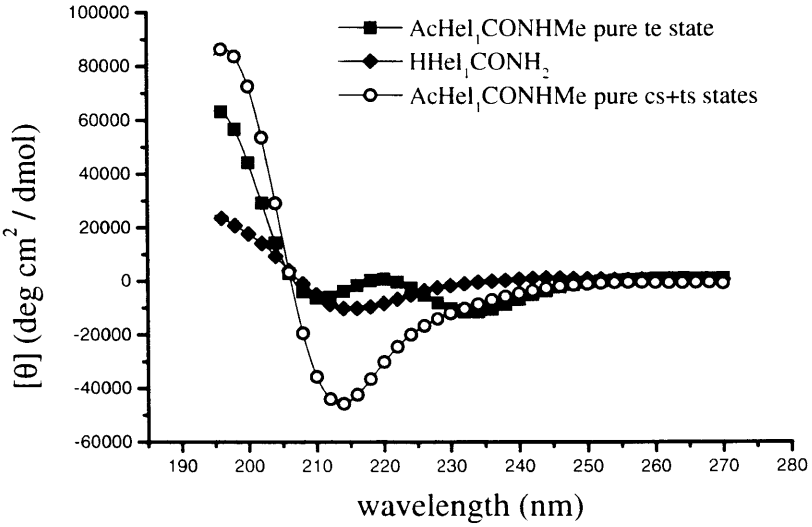
[35] Oslick, S. L. Thesis, Massachusetts Institute of Technology, **1996**.

Figure 4. CD spectrum of $AcHel_1CONHMe$, and of its pure cs+ts and te states.



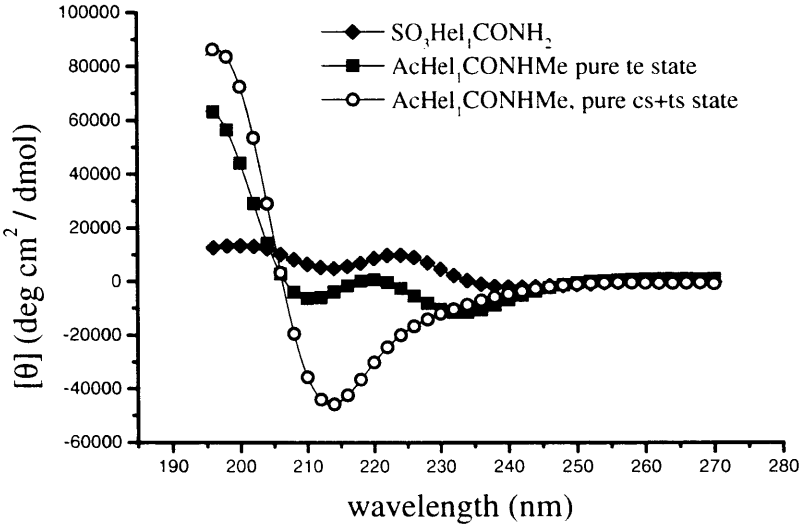
The acetamide group, a chromophore that contributes to the CD spectrum of $AcHel_1$, is missing in SO_3Hel_1. How does its removal alter the CD spectrum of $X\text{-}Hel_1$ derivatives? The CD spectrum of $HHel_1CONH_2$ is shown in figure 5, along with the limiting cs+ts and te CD spectra of $AcHel_1CONHMe$ for comparison. The CD spectrum of $HHel_1CONH_2$ has a minimum at 214 nm, similar to, but less intense than, that in the limiting cs+ts spectrum CD spectrum. Thus, removing the acetyl group from $AcHel_1$ leaves the minimum of its CD spectrum where it was found in its limiting cs+ts CD spectrum, but decreases the intensity of the signal.

Figure 5. CD spectra of $HHel_1CONH_2$ and the pure cs+ts and te states of $AcHel_1CONHMe$. Note that minima occur at the same position for $HHel_1CONH_2$ and the pure cs+ts states of $AcHel_1CONHMe$.



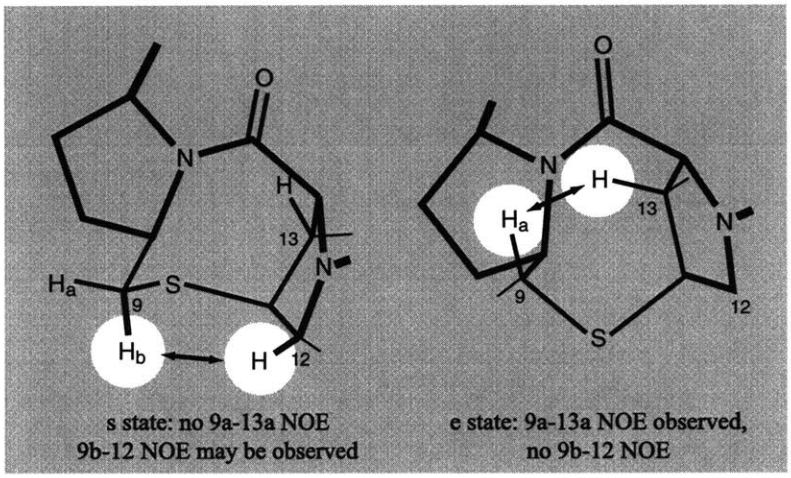
Now that the effect of the acetamide group on the CD spectrum of $AcHel_1$ is known, the effect of replacing it with a sulfamate group can be examined. The CD spectrum of $SO_3Hel_1CONH_2$ is shown in figure 6, along with the limiting cs+ts and te CD spectra of $AcHel_1CONHMe$ for comparison. The CD spectrum of $SO_3Hel_1NH_2$ has two minima at 215 and 238 nm and a maximum at 224 nm. This pattern of extrema, a minimum followed by a maximum followed by another minimum as one proceeds from low to higher UV wavelengths, is similar to that seen in the te state CD spectrum of $AcHel_1CONHMe$, although the extrema are shifted to higher wavelengths and the CD signal is generally more positive.

Figure 6. CD spectra of $SO_3Hel_1CONH_2$ and the pure cs+ts and te states of $AcHel_1CONHMe$. Note that extrema occur near the same positions for $SO_3Hel_1CONH_2$ and the pure te state of $AcHel_1CONHMe$.



The chemical shift of the 9b proton on the Hel_1 framework has been linked to the s ⇄ e equilibrium, as discussed in chapter 1.[17] There it was noted that the chemical shift of this proton changes from δ 3.32 to 2.87 ppm as $AcHel_1$ passes from the pure s state to the pure e state. The chemical shift of the 9b proton in SO_3Hel_1 is 2.89 ppm, very close to the limiting e state chemical shift observed in $AcHel_1$ derivatives. Since the s ⇄ e transition involves a rotation about the C8-C9 bond, the coupling constants between the 8 proton and the 9a and 9b protons ($^3J_{8,9a}$ and $^3J_{8,9b}$) also ought to depend on the position of the equilibrium. In fact, these coupling constants decrease from 9.9 and 5.3 Hz in $AcHel_1$ derivatives with almost no e character in the t state to 5.3 and 2.5 Hz in $AcHel_1$ derivatives with a great deal of e character in the t state.[15] The coupling constants of H9a and H9b with H8 in $SO_3Hel_1CONH_2$ are $^3J_{8,9a}$ = 5.5 Hz and $^3J_{8,9b}$ = 3.5 Hz, close to the limiting e state coupling constants observed in $AcHel_1$ derivatives.

Figure 7. NOEs characteristic of the Hel_1 s (H9b-H12) and e states (H9a-H13a). The cross peak indicating the e state is much stronger than that indicating the s state in the NOESY spectrum of $SO_3Hel_1CONH_2$.



CD spectra, chemical shifts, and coupling constants provide only circumstantial evidence for conformation. More direct indications of a molecule's conformation are available from measurements of nuclear Overhauser enhancements (NOEs). Such enhancements occur between pairs of protons that are consistently closer than 4.5 Å, and the intensity of the effect decreases with the sixth power of the distance between them. [36] In the Hel_1 system, H9a and H13a are only within NOE range in the e conformation; in the s conformation they are too far apart, but an NOE can occur between H9b and the H12 protons (see figure 7).[14] The NOESY spectrum of $SO_3Hel_1CONH_2$ has a medium-strength cross peak corresponding to the H9a, H13a interaction, indicating that this molecule significantly populates the e state. In contrast, only a weak cross peak is observed at the location expected for the H9b, H12 interaction. The cross peak cannot be

[36] Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, **1986**, pp. 93 – 133.

assigned definitely because the H9b signal overlaps with the H13b signal, but if it were due to an H9b, H12 interaction, its weakness would indicate that $SO_3Hel_1CONH_2$ populates the s state less than it populates the e state.

Four lines of evidence point to $SO_3Hel_1CONH_2$ residing largely in the e state. Its CD spectrum has the same pattern of extrema as the limiting te state CD spectrum of $AcHel_1CONHMe$. The chemical shift of the 9b proton and the coupling constants $^3J_{8,9a}$ and $^3J_{8,9b}$ are close to the limiting te state chemical shift reported for $AcHel_1$ derivatives. Finally, a medium-strength NOE characteristic of the e state is detected between the 9b and 13a protons of $SO_3Hel_1CONH_2$, while only a weak NOE characteristic of the s state is detected between the 9a and 12 protons.

4.3.2b The Hybridization of N1

In $AcHel_1$ derivatives, where N1 is acetylated, the N1 nitrogen atom is sp^2 hybridized because of amide resonance. The NMR signal of H8 appears between 4.3 and 4.5 ppm in these compounds.[14] Similarly, in $HHel_1CO_2H$ at low pH, where N1 is protonated, the NMR signal for H8 appears at 4.3 ppm. However, for this same compound at high pH, where N1 is deprotonated and has a free lone pair, the signal for H8 appears at 5.3 ppm, almost a full part per million downfield. H8 is five bonds away from N1, so this deshielding is probably not attributable to a through bond effect. It is more likely that the N1 lone pair influences H8's chemical shift via a through space effect, since N1 is only around 2.2 to 2.3 Å away from H8 in both the s and e states of $HHel_1CO_2H$ (according to molecular mechanics calculations using the CHARMm force field as in chapter 1).

Similar effects have been observed in the differences between axial and equatorial chemical shifts of ring protons in nitrogen heterocycles.[37]

The NMR signal of H8 in $SO_3Hel_1CONH_2$ appears at 5.3 ppm, similar to what is observed in deprotonated $HHel_1CO_2H$. This suggests that N1 has a free lone pair that influences the chemical shift of H8, and therefore that it is sp^3 hybridized, when it is sulfamated. This is not unexpected; the nitrogen atom is tetrahedral in the crystal structure of potassium sulfamate (KNH_2SO_3)[38] and normal coordinate analysis of the sulfamate ion[39] has yielded a fairly low bond order (1.2) for the ^-O_3S-N bond.

4.3.3 The Conformation of Peptides Attached to SO_3Hel_1

4.3.3a Expectations for Helices Induced by SO_3Hel_1: Helix Fraying

What can be expected of helices initiated by SO_3Hel_1? This question was discussed in section 2.4 for N-terminal templates in general. In brief, N-terminally initiated helices should fray unidirectionally from the N- to the C-terminus. Measures of the fractional helicity at individual residues should show helicity decreasing monotonically with distance from the template. Measures of the global fractional helicity should show both the random coil and helical contributions from all of the possible states, from the fully random coil state to the various partly helical and partly random coil states, to the fully helical state. Whether this is observed in the measures of local and global fractional

---

[37] Katritzky, A. R.; Halls, P. J.; Jones, R. A. Y.; Snarey, M.; Trepanier, D. L. *J. Chem. Soc. B* **1971**, 1320.
[38] Cox, G. W.; Sabine, T. M.; Padmanabhan, V. M.; Ban, N. T.; Chung, M. K.; Surjadi, A. J. *Acta Crystallographica, Sect. B* **1967**, *23*, 578.

helicities of SO_3Hel_1-peptide conjugates will be determined using the evidence presented in the following five sections. In the first four of these sections, results from helicity measurements on SO_3Hel_1-peptide conjugates will be compared to those observed in analogous $HHel_1$-peptide conjugates. The latter should not be able to form helices (since the $HHel_1$ lacks a hydrogen bond acceptor at N1), so these are natural models for the random coil states in SO_3Hel_1-peptide conjugates. The comparison ought to provide a better sense for whether the helicity measurements made on SO_3Hel_1-peptide conjugates are genuinely perturbed from their random coil values than comparison to literature data for random coils. Any sequence-specific effects in an SO_3Hel_1-peptide conjugate that might affect helicity measurements without being due to actual helix formation should be accounted for in the $HHel_1$-peptide conjugate. In the fifth section, direct evidence of helicity will be sought from the pattern of NOEs present in an SO_3Hel_1-peptide conjugate.

4.3.3b CD spectra of SO_3Hel_1-peptide Conjugates

CD spectroscopy provides a global measure of peptide helicity, as discussed in section 2.5. Recall that the CD spectra of helical peptides have two minima at 208 and 222 nm, while the CD spectra of random coil peptides have a single minimum at 195 nm and are featureless above 220 nm. The CD spectra of frayed helices should therefore be a combination of the two. They should have two minima, one at low wavelength between 195 and 208 nm, depending on the peptide's helical content, and one at 222 nm. The

---

[39] Sundara Raj, A.; Muthusubramanian, P. *J. Mol. Struct. (THEOCHEM)*, **1982**, *89*, 291.

depth of the minimum at 222 nm is proportional to fractional helicity, as per equation 2-7, which is reproduced below:

$$\frac{[\theta]_{222,observed} - [\theta]_{222,random\ coil}}{[\theta]_{222,100\%helix} - [\theta]_{222,random\ coil}} = \text{fractional helicity} \qquad \textbf{4 - 1}$$

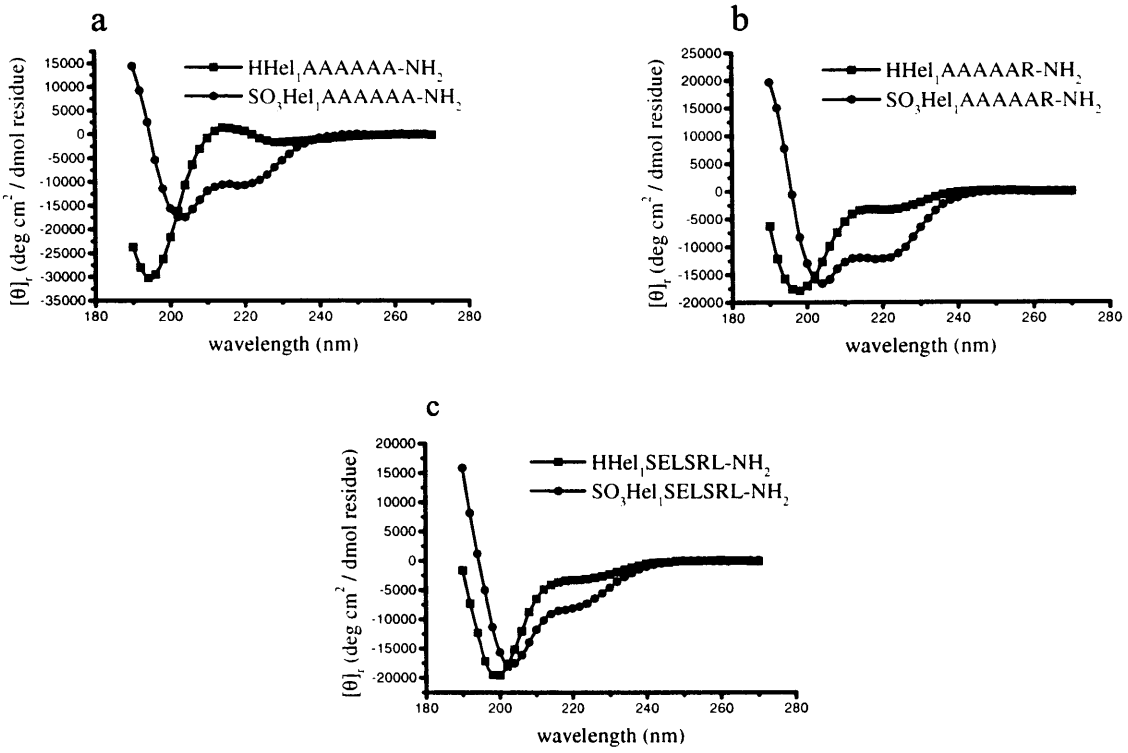
where $[\theta]_{222,observed}$, $[\theta]_{222,randomcoil}$, and $[\theta]_{222,100\%helix}$ are the per residue molar ellipticities (in deg cm^2 / dmol res) at 222 nm for the peptide of interest, the pure random coil state, and the pure helical state respectively.

The per residue CD spectra at 25 °C and pH 6 for three SO_3Hel_1-peptide conjugates, $SO_3Hel_1AAAAAA\text{-}NH_2$, $SO_3Hel_1AAAAAR\text{-}NH_2$, and $SO_3Hel_1SELSRL\text{-}NH_2$ are shown in figures 8a through 8c along with the CD spectra of their un-sulfamated $HHel_1$-peptide analogs. (All of these CD spectra have been corrected for the contribution of the template by subtracting the appropriate template spectrum from the template-peptide CD spectrum. See the experimental section for details.) $SO_3Hel_1AAAAAA\text{-}NH_2$ was chosen for study so that its CD spectrum could be compared to that of $AcHel_1AAAAAA\text{-}NH_2$, to gauge SO_3Hel_1's ability to initiate helices. $SO_3Hel_1AAAAAR\text{-}NH_2$ was chosen for study to see whether replacing the last alanine of the previous sequence with an arginine would increase helicity as judged by CD the same way substituting lysine at the C-terminus increases helicity, as judged by t/c ratios, in analogous $AcHel_1$-peptide conjugates.[40] $SO_3Hel_1SELSRL\text{-}NH_2$ was chosen for study to see whether peptides that were not alanine

[40] The t/c ratio at 25 °C of $AcHel_1AAAAAA\text{-}NH_2$ is 1.9 compared to 2.2 for $AcHel_1AAAAAK\text{-}NH_2$. Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp. D. S. *Proc. Natl. Acad. Sci.* **1996**, *93*, 4025.

rich could be made helical using this template. Its sequence comes from that of the gut hormone, secretin, which is known to be helical.[41]

Figures 8a-c. Per residue CD spectra of three pairs of $HHel_1$ and SO_3Hel_1-peptide conjugates (25 °C, pH ~ 6 in unbuffered water). Helical signatures are only observed for the peptides attached to SO_3Hel_1. CD spectra have been corrected for template contributions.



The CD spectra for these peptides will be discussed further in sections 4.3.3g (for $SO_3Hel_1AAAAAR\text{-}NH_2$ and $SO_3Hel_1SELSRL\text{-}NH_2$) and 4.3.4a (for $SO_3Hel_1AAAAAA\text{-}NH_2$). For now, the following two points are noted. First, the CD spectra of the analogous $HHel_1$-peptide conjugates have the single pronounced minimum below 200 nm

[41] Gronenborn, A. M.; Bovermann, G.; Clore, G. M. *FEBS Lett.* **1987**, *215*, 88.

characteristic of random coil peptide and a weak signal at 222nm. Second, and more importantly, the CD spectra of all three SO_3Hel_1-peptide conjugates have the two minima characteristic of helical peptides, with the lower wavelength minimum shifted down from 208 nm as expected for frayed, N-terminally initiated helices.

The CD spectra for two other peptides attached to $HHel_1$ and SO_3Hel_1 were measured to further test the capability of SO_3Hel_1 to induce helicity in heterogeneous peptides: ALQEAA, a randomly chosen sequence, and SEAQAL, a sequence taken from one of the helices in barnase[42] (residues 28-33). The differences between $[\theta]_{222}$ for the $HHel_1$ and SO_3Hel_1 versions of these two peptides were not as large as those seen for the three peptides whose CD spectra are shown in figure 8, but otherwise the results were similar (for ALQEAA, $HHel_1$: $[\theta]_{222} = -2{,}250$ deg cm^2 / dmol res, SO_3Hel_1: $[\theta]_{222} = -5{,}440$ deg cm^2 / dmol res; for SEAQAL, $HHel_1$: $[\theta]_{222} = -2{,}040$ deg cm^2 / dmol res, SO_3Hel_1: $[\theta]_{222} = -5{,}170$ deg cm^2 / dmol res).

4.3.3c Amide NH and α-H Chemical Shifts

The amide NH and αH chemical shifts in the NMR spectra of helical peptides are both expected to be found upfield relative to their positions in the spectra of random coil peptides.[43,44,45,46] The chemical shifts of NH protons in the random coil model peptide

---

[42] Martin, C.; Richard, V.; Salem, M.; Hartley, R.; Mauguen, Y. *Acta. Cryst. D* **1999**, *55*, 386, part 2.
[43] Williamson, M. P. *Biopolymers* **1990**, *29*, 1423.
[44] Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.
[45] Ösapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436.
[46] Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.

GGXA (where X is variable) fall between 8.1 and 8.8 ppm,[47] while the chemical shifts of the NH protons in partially helical peptides typically fall upfield by an average of 0.2 ppm (based on the chemical shifts of NHs in protein helices).[43,44] Comparable differences are observed for α protons, with the chemical shifts of the α protons of residues in protein helices being, on average, 0.26 ppm upfield of the chemical shifts of the α protons of residues in coil regions of proteins (regions that lack definite secondary structure).[44] Based on this information, and given that SO_3Hel_1 should induce C-terminally frayed helices in attached peptides, one would expect first, that the NH and αH protons of SO_3Hel_1-peptide conjugates should resonate upfield of the corresponding protons in $HHel_1$-peptide conjugates and second, that the difference should be less for the more C-terminal residues. This latter trend may not be monotonic, since the chemical shift change for full helix formation is probably residue dependent, but it should manifest itself to some extent. It should also be noted that the residue at the template-peptide junction may behave differently from the others because of the effect of local charge and the particular geometry it must assume in order for the first two hydrogen bonds to form.[18]

To test these expectations, the chemical shifts of the NH and αH protons in the representative pair of peptides SO_3Hel_1AQSFLR-NH_2 and $HHel_1$AQSFLR-NH_2, whose design and use in amide hydrogen exchange studies will be described in chapter 5, are listed in table 1. Also listed are the differences between the αH chemical shifts ($\Delta\delta_{\alpha H} = \delta_{\alpha H,SO3Hel} - \delta_{\alpha H,HHel}$) and the NH chemical shifts ($\Delta\delta_{NH} = \delta_{NH,SO3Hel} - \delta_{NH,HHel}$) for the

[47] Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, **1986**, p. 17.

SO_3Hel_1 and $HHel_1$ versions of this peptide. The $\Delta\delta$ values should be negative if SO_3Hel_1 induces helices. The NH and αH peaks were assigned by total correlation spectroscopy (TOCSY) for both of these compounds.

Table 1. Chemical shift data for $HHel_1AQSFLR\text{-}NH_2$ and $SO_3Hel_1AQSFLR\text{-}NH_2$ at 25 °C in 9:1 H_2O:D_2O. $\Delta\delta_{NH}$ are the differences between the NH chemical shifts, and $\Delta\delta_{\alpha H}$ are the differences between the αH chemical shifts, of the SO_3Hel_1 and $HHel_1$ versions of the peptide.

Residue	δ_{NH} for X-AQSFLR-NH_2 (ppm)			$\delta_{\alpha H}$ for X-AQSFLR-NH_2 (ppm)		
	X=SO_3Hel_1	X=$HHel_1$	$\Delta\delta_{NH}$	X=SO_3Hel_1	X=$HHel_1$	$\Delta\delta_{\alpha H}$
A	7.70	8.54	-0.84	4.29	4.28	0.01
Q	7.58	8.31	-0.73	4.19	4.29	-0.10
S	8.03	8.22	-0.19	4.24	4.40	-0.16
F	7.94	8.17	-0.24	4.53	4.62	-0.09
L	7.91	8.08	-0.17	4.24	4.28	-0.04
R	8.02	8.17	-0.15	4.24	4.32	-0.08

The δ_{NH} values for $HHel_1AQSFLR\text{-}NH_2$ range from 8.08 to 8.54, consistent with the range quoted above for random coils, while the δ_{NH} values for $SO_3Hel_1AQSFLR\text{-}NH_2$ range between 7.58 and 8.02, consistent with their being partially helical. In almost all of the cases, the chemical shifts of the NH and αH protons are at higher field in $SO_3Hel_1AQSFLR\text{-}NH_2$ than in $HHel_1AQSFLR\text{-}NH_2$, as expected if SO_3Hel_1 induced helicity. The only exception is for the $\delta_{\alpha H}$ of the residue at the template-peptide junction. The magnitudes of $\Delta\delta_{NH}$ and $\Delta\delta_{\alpha H}$ tend to be smaller for the more C-terminal residues

than for the more N-terminal residues, consistent with the helices being frayed. The largest change in chemical shift is observed for the NH protons of the first two residues (Ala and Gln), probably because the NHs of these residues hydrogen bond directly to the SO_3^- in helices initiated by SO_3Hel_1.

4.3.3d NH-αH Coupling Constants

The NH-αH coupling constant ($^3J_{NH,\alpha}$), which is a function of a residue's ϕ torsion, should be smaller for helical residues (3.9 Hz for a residue in an ideal helix,[48] 4.8 Hz considering the actual average ϕ angles observed in protein helices[49]) than random coil residues (range: 6.1 – 7.7 Hz, depending on the residue[49]). $^3J_{NH,\alpha}$ should therefore decrease as residues pass from the random coil to the helical state. The magnitude of the decrease depends on the fractional helicity of the residue in question, since the more helical the residue is, the larger the difference will be. It also depends on the identity of the residue in question, since the random coil coupling constant depends on the conformations that the residue samples in the random coil state. Residues that often have helical ϕ torsions when in the random coil state will have smaller random coil coupling constants, and the difference observed in $^3J_{NH,\alpha}$ upon their obtaining a helical conformation will be less. Random coil coupling constants aside, three consecutive residues with $^3J_{NH,\alpha} < 6.0$ Hz has been offered as a general criterion for local helicity.[48]

---

[48] Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore, **1986**, pp. 167-168.
[49] Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.

Based on this information, and keeping the effect of fraying in mind, the $^3J_{NH,\alpha}$ values for the residues in an SO_3Hel_1-peptide conjugate ought to decrease relative to the corresponding coupling constants in the analogous $HHel_1$-peptide conjugate. The magnitude of the decrease ought to diminish for the more C-terminal residues, but this trend may not be monotonic. Also, as with the chemical shifts, the change in the first residue's value of $^3J_{NH,\alpha}$ may be an exception because of the template-peptide junction.

To test these expectations, the values of $^3J_{NH,\alpha}$ measured from the NMR spectra of $HHel_1$AQSFLR-NH_2 at 25 °C and SO_3Hel_1AQSFLR-NH_2 at 25 and 5 °C are compared in table 2 (note that some amide peaks overlapped and their coupling constants could not be measured).

Table 2. Coupling constants of residues in $HHel_1$AQSFLR-NH_2 at 25 °C and SO_3Hel_1AQSFLR-NH_2 at 25 and 5 °C. The two spectra at 25 °C were both measured in unbuffered 9:1 H_2O:D_2O. The spectrum at 5 °C was measured in a pD 3.0, 200 mM phosphate D_2O buffer (it was possible to determine the $^3J_{NH,\alpha}$ values from this spectrum because the amide NHs had not yet exchanged fully with solvent deuterons).

Residue	$HHel_1$AQSFLR-NH_2 $^3J_{NH,\alpha}$ (Hz), 25 °C	SO_3Hel_1AQSFLR-NH_2 $^3J_{NH,\alpha}$ (Hz), 25 °C	SO_3Hel_1AQSFLR-NH_2 $^3J_{NH,\alpha}$ (Hz), 5 °C
A	5.5	6.7	6.7
Q	6.9	6.1	6.1
S	6.9	-	5.1
F	-	5.9	5.4
L	6.8	6.3	5.8
R	-	-	6.8

The αH-NH coupling constants, as expected if SO_3Hel_1 induced helicity in the attached peptide, are generally smaller in SO_3Hel_1AQSFLR-NH_2 than in $HHel_1$AQSFLR-NH_2 at 25 °C and even more so at 5 °C. The only exception is again due to the alanine residue at the template-peptide junction. Note that at 5 °C, the residues Q, S, F, and L all have coupling constants either very close to or less than 6.0 Hz, meeting the criterion set above for the presence of a helix.

4.3.3e Amide Hydrogen Exchange Kinetics

As discussed in section 2.5, the exchange rate of an amide that is hydrogen bonded in helices is slowed relative to the exchange rate of the same amide in a random coil because the hydrogen bond limits the accessibility of solvent to the amide.[50] If SO_3Hel_1 induced C-terminally frayed helices in attached peptides, then the exchange rates of amides in SO_3Hel_1-peptide conjugates should be slowed relative to those of $HHel_1$-peptide conjugates, with the slowing being less for the amides of residues that are closer to the C-terminus. The slowing of the hydrogen exchange rates in SO_3Hel_1-peptide conjugates is discussed in depth in chapter 5, where the effect is used quantitatively to measure fractional helicities at individual residues. For now, it is only noted that slowing of hydrogen exchange in eight SO_3Hel_1-peptide conjugates has been observed, and that the magnitude of the slowing decreases from the N- to the C-terminus.

4.3.3f NOEs in SO_3Hel_1-peptide Conjugates

In this section, the direct observation of helicity in an SO_3Hel_1-peptide conjugate is attempted by measuring NOEs. Just as CD spectra, chemical shifts, and coupling constants provided only circumstantial evidence for the conformation of SO_3Hel_1, so these three techniques only provide circumstantial evidence for a peptide's conformation. The presence of an appropriate set of NOEs is direct evidence for helicity in a peptide, and under favorable circumstances, NOEs can even be used to distinguish between α and 3_{10} helices. Recall that NOEs can be observed between protons within about 4.5 Å of each other, and the strength of the effect decreases with the sixth power of the distance. α Helices are compact structures with a number of pairs of protons close enough to produce NOEs.[51] These are listed in table 3 (adapted from table 1 of ref. 51).

Table 3. Table of NOE contacts in an α-helix. The subscripts refer to the residue's relative position in a peptide.

Contact	Distance
$\alpha H_i - NH_{i+1}$ ($\alpha N_{i,i+1}$)	3.5 Å
$\alpha H_i - NH_{i+2}$ ($\alpha N_{i,i+2}$)	4.4 Å
$\alpha H_i - NH_{i+3}$ ($\alpha N_{i,i+3}$)	3.4 Å
$\alpha H_i - NH_{i+4}$ ($\alpha N_{i,i+4}$)	4.2 Å

[50] Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1984**, *16*, 521.
[51] Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715.

$NH_i - NH_{i+1}$ ($NN_{i,i+1}$)	2.8 Å
$NH_i - NH_{i+2}$ ($NN_{i,i+2}$)	4.2 Å
$\beta H_i - NH_{i+1}$ ($\beta N_{i,i+1}$)	2.5 – 4.1 Å[52]
αH_i - βH_{i+3} ($\alpha\beta_{i,i+3}$)	2.5 – 4.4 Å[52]

Several of these pairs, αH_i –NH_{i+1} (shorthand: $\alpha N_{i,i+1}$), $NH_i - NH_{i+1}$ ($NN_{i,i+1}$), and $\beta H_i - NH_{i+1}$ ($\beta N_{i,i+1}$), are on adjacent residues. Although NOEs between these protons should be observed in α helices, the protons are never very far from each other in any conformation. In particular, the $\alpha N_{i,i+1}$ is observed in random coils as well as α helices.[53] These NOEs, therefore, cannot by themselves confirm α helicity. The NOEs between pairs of protons on residues that are positioned i, i+2 relative to each other, αH_i - NH_{i+2} ($\alpha N_{i,i+2}$) and N_i - $N_{i,i+2}$ ($NN_{i,i+2}$) also are not very useful for defining α helices, because the distances between both of these pairs of protons should be long (4.2 and 4.4 Å respectively). These NOEs should be weak in α helices, but stronger in other structures, such as turns or 3_{10} helices.[51] The NOEs that most emphatically identify α helices are those that occur between pairs of protons separated by two or more residues, the $\alpha H_i - NH_{i+3}$ ($\alpha N_{i,i+3}$), αH_i - βH_{i+3} ($\alpha\beta_{i,i+3}$), and $\alpha H_i - NH_{i+4}$ ($\alpha N_{i,i+4}$) NOEs. These limit the conformations of the two or three intervening residues to only a few possibilities. The $\alpha N_{i,i+3}$ should be fairly strong for α helices but could also be attributed to 3_{10} helices or some turn conformations.[51] The $\alpha\beta_{i,i+3}$ should be only be observable in α helices and 3_{10} helices, and should be much stronger in the former than in the latter.[51] Note that the

[52] Varies with the χ_1 torsional angle.

strength of this NOE depends on the χ_1 torsional angle of the i+3rd residue's side chain. The $\alpha\beta_{i,i+4}$ NOE should be weak, but also unique, for α helices. If this NOE can be identified in a peptide, it virtually guarantees that the residues between i and i+4 are not only helical, but α helical.[51]

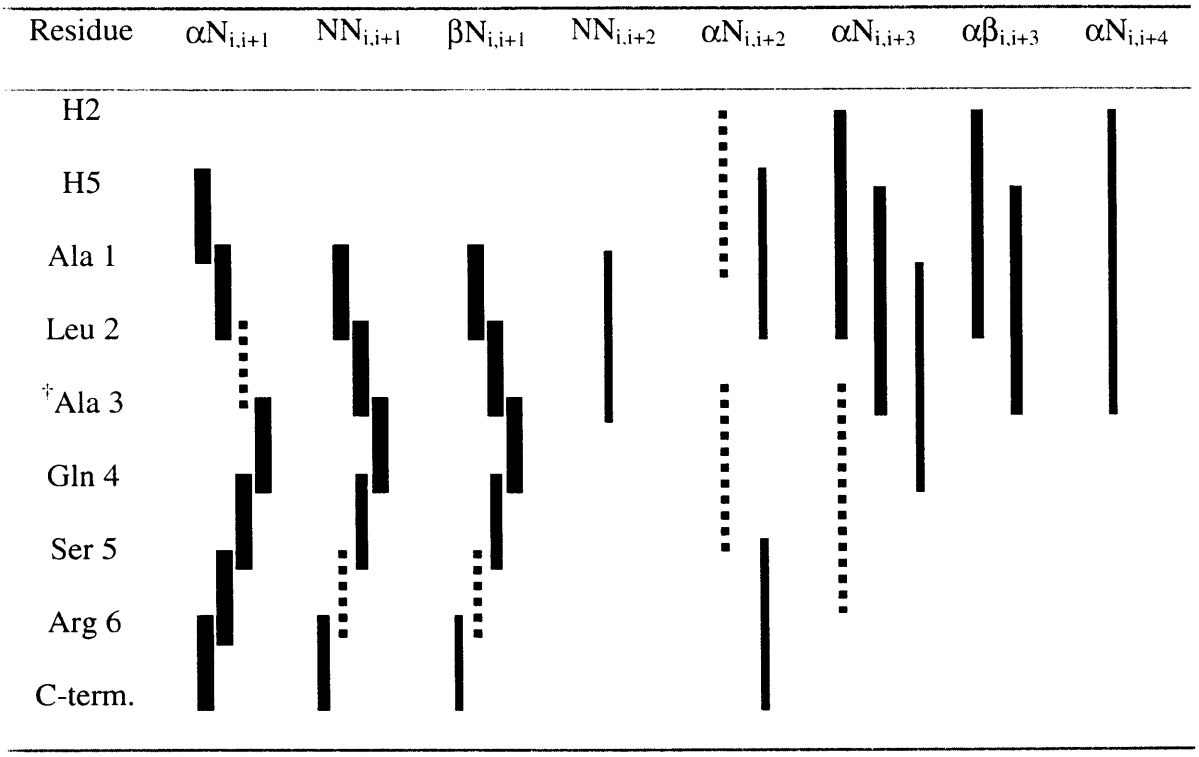
What can be expected of the NOEs found in SO_3Hel_1-peptide conjugates? If SO_3Hel_1 initiates frayed α helices, then the $\alpha N_{i,i+1}$ NOEs should be found with the same intensity throughout the peptide (since these occur in random coils as well as helices). The $NN_{i,i+1}$ and $\beta N_{i,i+1}$ NOEs should decrease toward the C-terminus, but may not disappear even if the helix is strongly frayed, since these pairs of protons have so much opportunity to be within NOE distance. Both $\alpha N_{i,i+2}$ and $NN_{i,i+2}$ NOEs should be weak, if observed at all. The intensities of the $\alpha N_{i,i+3}$, $\alpha\beta_{i,i+3}$ and $\alpha N_{i,i+4}$ NOEs should all decrease toward the C-terminus, and might disappear if the helix is strongly frayed.

The NOEs found in a sample SO_3Hel_1-peptide conjugate, $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$, whose design and use in amide hydrogen exchange will be described in the next chapter, are listed in table 4. The dagger at the third residue indicates that this alanine was labeled with ^{15}N in the α nitrogen to ensure that the resonances of the two alanines could be distinguished.

---

[53] Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161.

Table 4. Table of the inter-residue NOEs found in the ROESY spectrum of $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$ (600 MHz, 5 °C). The thickness of the line is proportional to the intensity of the NOE. A solid line indicates that the NOE was unambiguously observed; a dashed line indicates that an NOE was observed at the relevant position, but that it could not be assigned unambiguously because of overlap.



The $\alpha N_{i,i+1}$ NOEs are observed throughout the peptide with no decrease in intensity at the C-terminus. The $NN_{i,i+1}$ and $\beta N_{i,i+1}$ NOEs are also observed throughout the peptide, but their intensity decreases toward the C-terminus. Only one $NN_{i,i+2}$ is observed at the beginning of the helix, and it is appropriately weak. The $\alpha N_{i,i+2}$ NOEs at the N-terminus are also weak but, surprisingly, either one or two more NOEs of this type are seen at the C-terminus, one between Ser 5 and the C-terminal NH_2, and another possibly between Ala 3 and Ser 5 (the latter could also be an $\alpha N_{i,i+3}$ NOE between Ala 3 and Arg 6). These C-terminal $\alpha N_{i,i+2}$ NOEs will be discussed further below. The $\alpha N_{i,i+3}$ and $\alpha\beta_{i,i+3}$ are

medium strength near the SO_3Hel_1 template, their intensity decreases toward the middle of the peptide, and they disappear toward the C-terminus. The possible $\alpha N_{i,i+3}$ NOE between Ala 3 and Arg 6, which, as noted above, cannot be distinguished from an $\alpha N_{i,i+2}$ NOE between Ala 3 and Ser 5, will be discussed further below. Finally, a weak $\alpha N_{i,i+4}$ NOE is observed at the beginning of the helix.

4.3.3g Evaluation of the Evidence

Each of the helicity measures used in the five preceding sections, CD spectroscopy, α and NH proton chemical shifts, αH–NH coupling constants, amide hydrogen exchange, and NOEs, indicate the induction of helicity by SO_3Hel_1 in attached peptides with fraying toward the C-terminus. That the induced helicity is substantial can be shown by computing the global fractional helicity for the simple, homogeneous peptide $SO_3Hel_1AAAAAA\text{-}NH_2$. Given that $[\theta]_{222}$ for this peptide is −10,200 deg cm^2 /dmol res, and that $[\theta]_{222,100\%helix}$ and $[\theta]_{222,randomcoil}$ are −28,520 and −1,890 deg cm^2 / dmol res respectively (the values suggested for use with short template-peptide conjugates in section 2.5), then according to equation 4-1, the global fractional helicity of this peptide is about 0.31. The helicity increases somewhat when a terminal alanine is replaced by an arginine ($[\theta]_{222}$ = −11,950 deg cm^2 / dmol res, global fractional helicity = 0.38), just as a lysine at the same position results in an increase in the t/c ratio of $AcHel_1$-peptide conjugates.[40] Substantial helicity is even induced in a very heterogeneous peptide with two helix-breaking serine residues, $SO_3Hel_1SELSRL\text{-}NH_2$ ($[\theta]_{222}$ = −7,780 deg cm^2 /

dmol res, global fractional helicity = 0.22). Note that although these three peptides SO_3Hel_1-peptide conjugates are each only six residues long, and despite the CD measurements being taken at room temperature, all of these peptides are more helical that the C-peptide, an eleven residue peptide, is at 3 °C and optimal pH ($[\theta]_{222}$ = −7,100 deg cm^2 / dmol res; see chapter 3).

The NOEs measured in $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$ show the fraying effect explicitly. At the N-terminus, near the template, all of the NOEs expected for a helix are observed, and since there is an $\alpha N_{i,i+4}$ NOE, it can be asserted that this is specifically an α helix. However, if one ignores the possible $\alpha N_{i,i+3}$ NOE between Ala and Leu for the moment, none of the longer range i,i+3 or i,i+4 NOEs are present beyond the Gln 4 NH. This does not mean that the helix extends to Gln 4 and then stops abruptly. Rather, it is consistent with the concentration of states in which the helix extends beyond Gln 4 being below the threshold needed to detect NOEs at this distance. Regarding the pair of C-terminal NOEs seen in this peptide, the $\alpha N_{i,i+2}$ NOE between Ser 5 and the C-terminal NH_2 and the NOE that corresponds to either an $\alpha N_{i,i+2}$ NOE between Ala 3 and Ser 5 or an $\alpha N_{i,i+3}$ NOE between Ala 3 and Arg 6, two explanations are offered. First, they might be due to a helix capping structure that is specific to this peptide, or second they might be due to some residual structure that these residues occupy in the non-helical state. The latter possibility could be tested in future work by measuring the NOEs in $HHel_1AL^{\dagger}AQSR\text{-}NH_2$. If these $\alpha N_{i,i+2}$ NOEs appeared in $HHel_1AL^{\dagger}AQSR\text{-}NH_2$ as well, then they could not be due to a structure induced by SO_3Hel_1. The relevance of these NOEs will be taken up again at the end of chapter 5.

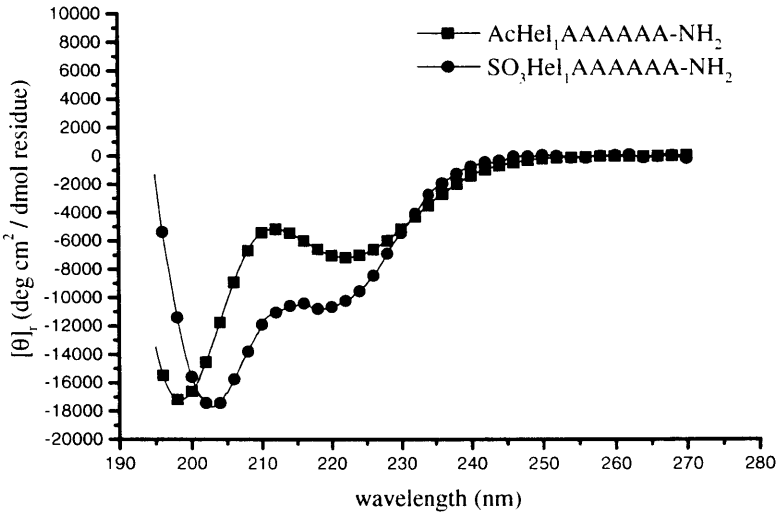
The SO_3Hel_1 template effectively induces helicity in attached peptides, but how does its potency as an initiator compare to that of $AcHel_1$? This question is answered in the following section.

4.3.4 Comparison of SO_3Hel_1 with $AcHel_1$

4.3.4a Qualitative

The effectiveness of SO_3Hel_1 as a helix initiating template can be compared qualitatively to that of $AcHel_1$ by comparing the template corrected CD spectra of $SO_3Hel_1AAAAAA\text{-}NH_2$ and $AcHel_1AAAAAA\text{-}NH_2$. (The correction of the former is described in the experimental section while the correction of the latter is described in ref. 35). These two CD spectra are shown in figure 9.

Figure 9. Comparison of the per residue CD spectra of $AcHel_1$- and $SO_3Hel_1AAAAAA\text{-}NH_2$ (both at 25 °C). The deeper minimum at 222 nm for the SO_3Hel_1 version of the peptide indicates its greater helicity.



It is apparent from the figure that $[\theta]_{222}$ is more negative for $SO_3Hel_1AAAAAA\text{-}NH_2$ than it is for $AcHel_1AAAAAA\text{-}NH_2$. The value of $[\theta]_{222}$ for the former is −10,200 deg cm^2 / dmol res while $[\theta]_{222}$ for the latter it is −7,180 deg cm^2 / dmol res. According to equation 4-1, and using the values quoted above for $[\theta]_{222,100\%helix}$ and $[\theta]_{222,randomcoil}$, these values of $[\theta]_{222}$ for short templated peptides correspond to fractional helicities 0.31 and 0.20. The helicity of the SO_3Hel_1-peptide conjugate is roughly 50% greater than that of the $AcHel_1$-peptide conjugate. This is a notable increase, but it is not clear that it is due to anything more than the elimination of the trans ⇄ cis equilibrium. To determine whether the sulfamate's negative charge or its multiple hydrogen bond acceptors add anything to helix initiation by SO_3Hel_1 it is necessary to approach the matter quantitatively.

4.3.4b Quantitative

In order to quantify SO_3Hel_1's ability to initiate helicity, its sum of helix-coil equilibrium constants must first be determined. The equilibria in SO_3Hel_1 should be much like the equilibria in $AcHel_1$ (shown in figure 8 of chapter 2), except for the lack of the acetamide trans ⇄ cis interconversion. If the $AcHel_1$-SO_3Hel_1 analogy is valid, then the helix-coil equilibria for an n residue SO_3Hel_1-peptide conjugate can be written as shown below.

$$s \underset{}{\overset{K_{1SO3}}{\rightleftharpoons}} e_0 \overset{K_{2SO3}}{\rightleftharpoons} e_1 \overset{s_1}{\rightleftharpoons} e_2 \overset{s_2}{\rightleftharpoons} \cdots \overset{s_n}{\rightleftharpoons} e_n$$

A letter indicates the conformation of the C8-C9 bond in each state, s for staggered and e for eclipsed. The e states also have a subscript indicating the number of hydrogen bonds that have formed, so e_0 is the state in which the C8-C9 bond is eclipsed but no hydrogen bonds have formed, e_1 is the state in which the amide NH at the template-peptide junction has formed the first hydrogen bond with the SO_3^- group, etc. The first two of the equilibria above, $s \rightleftarrows e_0$ and $e_0 \rightleftarrows e_1$, take place within the template while the attached peptide remains in the random coil state. Their equilibrium constants are K_{1SO3} and K_{2SO3}. The equilibria that follow involve helix formation to progressively greater degree, and their equilibrium constants are the s values of the residues that are joining the helix. Taking the template staggered-peptide random coil state as the reference state, the sum of equilibrium constants for an SO_3Hel_1-peptide conjugate is

$$Z'_{SO3} = 1 + \frac{e_0}{st} + \frac{e_1}{e_0} + \frac{e_2}{e_1} + \cdots + \frac{e_n}{e_{n-1}}$$

$$= 1 + K_{1SO3} + K_{1SO3}K_{2SO3} + K_{1SO3}K_{2SO3}s_1 + K_{1SO3}K_{2SO3}s_1s_2 + \cdots + K_{1SO3}K_{2SO3}s_1s_2\cdots s_n$$

Dividing both sides of this equation by $1 + K_{1SO3}$ and substituting B_{SO3} for $K_{1SO3}K_{2SO3} / (1 + K_{1SO3})$ and Z_{SO3} for $Z'_{SO3} / (1 + K_{1SO3})$ yields the simpler expression

$$Z_{SO3} = 1 + B_{SO3} + B_{SO3}(s_1 + s_1s_2 + \cdots + s_1s_2\cdots s_n) \qquad \textbf{4 - 2}$$

The constant B_{SO3} defines SO_3Hel_1's initiation efficiency, since it multiplies the sum of s value products. B_{SO3} must be determined to fulfill the purpose of this section. This can be achieved by noting that the sum of equilibrium constants in equation 4-2 can be translated to an expression for fractional helicity using equation 2-3 from chapter 2:

$$\text{fractional helicity} = \frac{B_{SO3}\left(\frac{1}{n}s_1 + \frac{2}{n}s_1s_2 + \cdots \frac{n}{n}s_1s_2\cdots s_n\right)}{Z_{SO3}} \qquad \textbf{4 - 3}$$

This equation can be combined with equation 4-1 to yield the following

$$\frac{[\theta]_{222,observed} - [\theta]_{222,randomcoil}}{[\theta]_{222,100\%helix} - [\theta]_{222,randomcoil}} = \frac{B_{SO3}\left(\frac{1}{n}s_1 + \frac{2}{n}s_1s_2 + \cdots \frac{n}{n}s_1s_2\cdots s_n\right)}{Z_{SO3}}$$

For $SO_3Hel_1AAAAAA\text{-}NH_2$, the above expression is

$$\frac{-10{,}200\,\frac{deg\,cm^2}{dmol\,res} - \left(-1{,}890\,\frac{deg\,cm^2}{dmol\,res}\right)}{-28{,}520\,\frac{deg\,cm^2}{dmol\,res} - \left(-1{,}890\,\frac{deg\,cm^2}{dmol\,res}\right)} = \frac{B_{SO3}\left(\frac{1}{6}s_A + \frac{2}{6}s_A^2 + \cdots s_A^6\right)}{1 + B_{SO3}\left(s_A + s_A^2 + \cdots s_A^6\right)}$$

Using the value cited in chapter 3 for the helix propensity of alanine, that is, $s_A = 1.02$, the above can be solved for B_{SO3}. This yields $B_{SO3} = 0.2$.

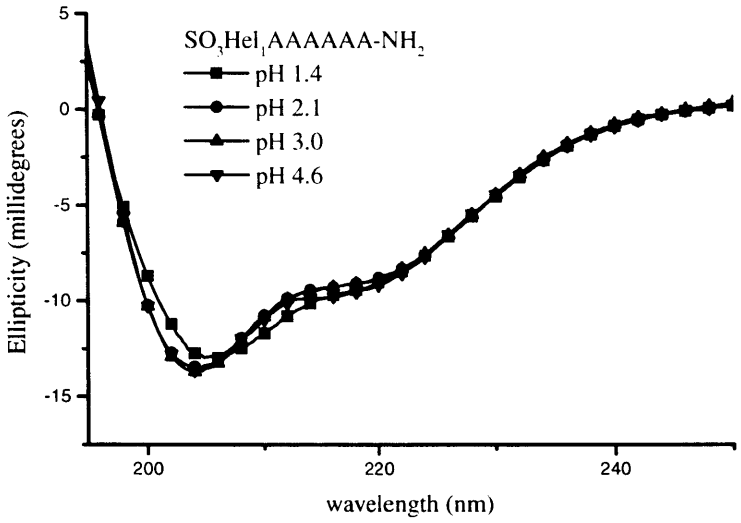
The quantity B_{SO3} from SO_3Hel_1 is analogous to B from $AcHel_1$ since these factors multiply the sums of helix propensity products (the expressions of the form $s_1 + s_1s_2 + \cdots + s_1s_2\cdots s_n$) in the sums of helix-coil equilibrium constants of SO_3Hel_1- and $AcHel_1$-peptide conjugates. As noted in chapters 2 and 3, B from $AcHel_1$ is 0.156. B_{SO3} is only about 30% larger than this, suggesting that the SO_3^- group only modestly improves helix initiation by the Hel_1 framework.

This is a surprising result. For all of the reasons cited in the design section, SO_3Hel_1 should have been a far better template than $AcHel_1$. Furthermore, it was shown in section 4.3.2a that $SO_3Hel_1NH_2$ occupied the e state to a greater degree than $AcHel_1CONHMe$. Since this is the helix initiating state in $AcHel_1$, this should have led to a further advantage for SO_3Hel_1. Two explanations are offered for SO_3Hel_1's performance.

First, it is possible that neither the improvement in hydrogen bond acceptor strength nor the improvement in helix dipole interactions were realized upon replacing the neutral acetamide of $AcHel_1$ with the negatively charged sulfamate of SO_3Hel_1. To test this prospect, the CD spectra for $SO_3Hel_1AAAAAA\text{-}NH_2$ were measured at pHs 4.6, 3.0, 2.1, and 1.4. Although sulfamates protonate on the nitrogen to create zwitterions, the presence of a positive charge on the nitrogen should negate any helix-favoring effects from the negative charge on the oxygen. As shown in figure 10, the CD spectrum of this template-peptide conjugate does not alter perceptibly over this pH range. Since it is likely that the sulfamate is protonated to at least some extent at the low end of this pH range (considering that the pK_as of mono- and dialkyl substituted sulfamates range from

1.3 to 2.3[54]), this evidence is consistent with the sulfamate's negative charge playing a minimal role in helix induction. It also raises the intriguing possibility that the helix stabilizing effects that are usually attributed to charge-helix dipole interactions do not, for some reason, operate in these N-terminally templated helices.

Figure 10. Dependence of the CD spectrum of $SO_3Hel_1AAAAAA\text{-}NH_2$ on pH. Note that the spectra change very little, even at the lowest pH. Measurements at pH 1.4 and 2.1 were made in solutions of perchloric acid; those at higher pHs were made in a phosphate buffer. Raw ellipticity measurements (instead of molar ellipticities) are plotted on the ordinate.



Second, it is possible that the improvement in hydrogen bonding geometry at the template-peptide junction was not realized upon replacing the acetamide (one acceptor site) with the sulfamate (three acceptor sites) because of the sp^3 hybridization of the sulfamate nitrogen. If the lone pair of N1 is directed toward H8, as suggested from this proton's unusual chemical shift in SO_3Hel_1 derivatives, then the SO_3^- group would be

[54] Spillane, W. J.; Hannigan, T. J.; Shelley, K. P. *J. Chem. Soc. Perkin Trans 2* **1982**, 19.

directed away from the core of the template. This may have offset any advantages offered by the sulfamate's multiple acceptor sites.

4.4 Summary and Conclusions

The goals of this work were to design and prepare a derivative of the Hel_1 core structure that would both be a stronger helix initiator and permit hydrogen exchange studies of its peptide conjugates. The new template, SO_3Hel_1, is only modestly successful in meeting the first goal. While it is a better helix initiator than $AcHel_1$, the improvement can be mostly attributed to eliminating $AcHel_1$'s trans $\rightleftarrows$ cis equilibrium. Despite its negative charge and multiple hydrogen bond acceptor sites, SO_3Hel_1's sulfamate group does not favor helicity much more than $AcHel_1$'s acetamide. SO_3Hel_1 is, however, entirely successful in meeting the second goal. Amide hydrogen exchange studies with SO_3Hel_1-peptide conjugates were alluded to in section 4.3.3e as a source of evidence of helicity. These studies will be presented in detail in the next chapter.

4.5 Experimental

Equipment. One dimensional ^{1}H-NMR spectra were measured at 500 MHz and ^{13}C-NMR spectra were measured at 125 MHz on Varian VXR500S and 501S spectrometers and processed using the Varian Instruments VNMR 3.1 software. Chemical shifts were measured relative to the reference signal of (trimethylsilyl) propionic-2,2,3,3-d_4 acid (TMSP). Most ^{1}H-NMR spectra were obtained using a 60° pulse width with a 2 s

acquisition time (40,000 points) and no delay between pulses. ^{1}H-NMR spectra for concentration determination were obtained using a 60° pulse width with a 4 s acquisition time and a 12 s delay between pulses. Between 128 and 512 transients were acquired, depending on the concentration of the sample (typically between 0.5 and 3 mM). CD spectra were measured on an Aviv 62DS circular dichroism spectrometer using 1, 5 or 10 mm strain free quartz cells (Hellma) depending on the concentration of the sample (usually 10 – 40 μM). CD spectra were processed with Aviv 62DS version 4.0s software. Analytical high performance liquid chromatography (HPLC) was performed on a Waters system consisting of two 501 pumps, a rheodyne injector, a model 660 automated gradient controller, a model 740 data module, a model 484 detector, and a Vydac 0.46 × 25 cm (218TP54) C_{18} reverse phase column. Flow rates were 1.0 mL/min. Preparative scale HPLC was performed on a Waters system consisting of a model 590 pump fitted with preparative heads, an Autochrome DPG/S pre-pump solvent mixer, a Rheodyne injector, a model 484 variable wavelength detector, and a Waters 2.5 × 10 cm radial compression column housed in a PrepLC 2.5 cm radial compression module (RCM). Flow rates for preparative HPLC were 12 mL/min. Detection in all uses of HPLC was carried out at 214 nm. The pHs or pDs of solutions were measured using either pH paper (colorpHast indicator strips pH 0 to 14 (EM reagents)) or a Cole-Parmer pH meter (model # 5982-00) and a 3.5 × 183 mm glass electrode with a calomel reference (model # 5990-30). The pH meter was referenced using pH 4.00 and 7.00 certified buffers. The H/D isotope effect for D_2O solutions was accounted for by adding 0.4 to the pD that was read off the meter ($pD = pD_{read} + 0.4$).[55] Mass spectra were obtained on a Hewlett-Packard HP5989B electrospray ionization mass spectrometer, courtesy of Dr. P. Wishnok and

[55] Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188.

Prof. S. Tannenbaum, from samples dissolved in 1:1 water: methanol with 0.1% acetic acid, detecting positive ions for $HHel_1$-peptide conjugates or negative ions for SO_3Hel_1-peptide conjugates.

Determination of Concentration. Concentrations were determined by quantitative integration of sample peaks against two or three internal standards. Known concentrations (~ 1 mM) of at least two of methanol, dioxane, and TMSP were added to D_2O solutions of the sample peptides (for example: 25 μL of a solution of 40 mM TSP in D_2O would be added to a solution of the sample in D_2O and the resulting solution diluted to 1 mL, to yield a solution known to be 1 mM in TSP; other standards could be added in the same way). The peak areas for single protons of the sample and the standards were determined and the peptide concentration calculated relative to each standard according to the equation below:

$$\frac{\text{peak area for a single sample proton}}{\text{peak area for a single standard proton}} \times \text{concentration of standard} = \text{peptide concentration}$$

The concentrations found according to each internal standard were then averaged to provide the mean sample concentration. The concentrations found according to the different standards usually differed by less than 10%. Once the concentrations of these samples were known, they could be diluted to known concentrations for CD spectroscopy.

Two Dimensional NMR spectroscopy. The NOESY spectrum (in D_2O) $SO_3Hel_1CONH_2$ and the TOCSY spectra (in 9:1 H_2O:D_2O) for $HHel_1AQSFLR\text{-}NH_2$ and $SO_3Hel_1AQSFLR\text{-}NH_2$ were acquired at 25 °C and 5 °C on the 500 MHz Varian spectrometers described above. For the NOESY spectrum, the spectral width was 3000

Hz (6 ppm) in both dimensions, and 64 t_1 increments were obtained each consisting of 16 transients. The mixing time was 400 ms. For the TOCSY spectra, the spectral width was 5000 Hz (10ppm) in both dimensions and 32 t_1 increments were obtained each consisting of 32 transients. The H_2O peak was suppressed by 1 s of presaturation. The mixing time was 60 ms. Data for both NOESY and TOCSY spectra were processed with gaussian weighting in both dimensions. The TOCSY and ROESY spectra for SO_3Hel_1-$AL^{\dagger}AQSR$-NH_2 were kindly obtained by Dr. S. Pochapsky on a Bruker DRX-600 600 MHz NMR spectrometer in 9:1 H_2O:D_2O at 5 °C. The spectral width was 4800 Hz in both dimensions and 512 t_1 increments were measured consisting of either 16 (for the TOCSY) or 128 (for the ROESY) transients. The mixing time was either 60 ms (for TOCSY) or 250 ms (for ROESY). The H_2O peak was suppressed by the WATERGATE method.

Correction of the CD Spectra of $HHel_1$- and SO_3Hel_1-peptide Conjugates for Template Contributions. The CD spectra of $HHel_1$- and SO_3Hel_1-peptide conjugates, like the CD spectra of $AcHel_1$-peptide conjugates, were considered to arise from independent contributions due to the template and the peptide attached to the template. The CD spectra of template-peptide conjugates could therefore be corrected for contributions due to the template by subtracting the molar CD spectrum of the appropriate template (shown in section 4.3.2) from the molar CD spectrum (as opposed to the per-residue CD spectrum) of the template-peptide conjugate. The corrected molar CD spectrum of the template-peptide conjugate, now free from contributions due to the template, was then divided by the number of residues to obtain the per residue CD

spectrum. The molar ellipticities from 270 to 195 nm from the CD spectra of $HHel_1CONH_2$ and $SO_3Hel_1CONH_2$ are listed in appendix 1.

Aggregation. None of the peptides described above show any of the obvious signs of aggregation, such as broadening of peaks in NMR spectra. In addition, the CD spectra of $SO_3Hel_1AAAAAA\text{-}NH_2$ and $SO_3Hel_1SELSRL\text{-}NH_2$ are invariant upon 100-fold dilution (~ 300 μM to ~ 3 μM). Similarly, the NMR spectrum of $SO_3Hel_1AQFASR\text{-}NH_2$, a peptide with a sequence very similar to those of $SO_3Hel_1AQSFLR\text{-}NH_2$ and $SO_3Hel_1AL^{+}AQSR\text{-}NH_2$ is unchanged upon 10 fold dilution (~ 2 mM to ~ 200 μM), and its CD spectrum is further unchanged upon another 10 fold dilution (~ 200 μM to ~ 20 μM). The above results are consistent with SO_3Hel_1-peptide conjugates such as the ones for which data are presented above being monomeric up to at least millimolar concentrations.

Synthesis of $HHel_1$-peptide Conjugates. The solid phase peptide synthesis and deprotection / cleavage of $HHel_1$-peptide conjugates (as the C-terminal acids) proceeded much the same as the synthesis of $AcHel_1$-peptide conjugates described in the previous chapter, except that $BocHel_1CO_2H$ replaced $AcHel_1CO_2H$ at the stage where the template was coupled. TFA deprotection / cleavage produced mixtures of $HHel_1$-peptide conjugates and the un-templated peptides (since $BocHel_1CO_2H$, like $AcHel_1CO_2H$, was used as the limiting reagent in the last coupling). These could be easily separated by HPLC as $HHel_1$-peptide conjugates, like $AcHel_1$-peptide conjugates, had much longer retention times than the corresponding free peptides. The calculated vs. observed masses and HPLC retention times for the $HHel_1$-peptide conjugates are recorded in table 5. Note

that $HHel_1NH_2$ was prepared by coupling $BocHel_1CO_2H$ directly to the resin followed by deprotection / cleavage.

Synthesis of SO_3Hel_1-peptide Conjugates by Sulfamation of $HHel_1$-peptide Conjugates. A typical sulfamation reaction began with dissolving an $HHel_1$-peptide conjugate in freshly prepared 10% aqueous $NaHCO_3$ to between 2 and 5 mg / mL. This solution was treated with a large excess of SO_3-pyridine complex (~ 100 eq, usually between 60 and 100 mg). The SO_3-pyridine complex dissolved slowly in the solution as it hydrolyzed and reacted. After stirring for 12 h at 25 °C, the reaction would not proceed any further (usually about 30% unreacted starting material remained), so the product was purified by preparative HPLC. The reaction mixture, which was approximately neutral by this time, could be directly applied to the HPLC column with no intervening work-up. Note that SO_3Hel_1-peptide conjugates always had somewhat longer retention times than the corresponding $HHel_1$-peptide conjugates, and their separation was facile. A specific sample procedure is given below for the preparation of $SO_3Hel_1ALQEAA\text{-}NH_2$ from $HHel_1ALQEAA\text{-}NH_2$.

$HHel_1ALQEAA\text{-}NH_2$ (6.6 μmol, 5.5 mg) were dissolved at 25 °C in 1 mL of 10% aqueous $NaHCO_3$, and this solution was treated with 100 mg (580 μmol) of SO_3-pyridine complex and the resulting suspension was vigorously stirred. After 3 h the reaction mixture had become clear. After another 12 h, HPLC showed the reaction to be about 70% complete. The reaction mixture was injected directly onto the preparative HPLC for purification. This yielded 3.3 mg (3.6 mmol, 55%) of $SO_3Hel_1ALQEAA\text{-}NH_2$.

The calculated vs. observed masses and HPLC retention times for the SO_3Hel_1-peptide conjugates are recorded for all new compounds in table 5.

Table 5. Characterization of $HHel_1$ and SO_3Hel_1 derivatives. The retention times are for a 5% – 100% CH_3CN (remainder 0.1% TFA in H_2O) gradient over 40 min on the preparative HPLC system described above.

Peptide	observed m/z (calculated m/z)	retention time
$HHel_1NH_2$	$(M+H)^+$: 256 (256.1)	8.2 min
$SO_3Hel_1NH_2$	M^-: 334 (334.1)	8.7 min
$HHel_1AAAAAA\text{-}NH_2$	$(M+H)^+$: 682 (682.3)	7.7 min
$SO_3Hel_1AAAAAA\text{-}NH_2$	M^-: 760 (760.3)	10.5 min
$HHel_1AAAAAR\text{-}NH_2$	$(M+H)^+$: 767 (767.4)	7.4 min
$SO_3Hel_1AAAAAR\text{-}NH_2$	M^-: 845 (845.3)	10.0 min
$HHel_1SELSRL\text{-}NH_2$	$(M+H)^+$: 941 (941.5)	11.2 min
$SO_3Hel_1SELSRL\text{-}NH_2$	M^-: 1019 (1019.4)	14.0 min
$HHel_1SEAQAL\text{-}NH_2$	(M+H)+: 855 (855.4)	9.0 min
$SO_3Hel_1SEAQAL\text{-}NH_2$	M-: 933 (933.3)	10.0 min
$HHel_1ALQEAA\text{-}NH_2$	(M+H)+: 838.2 (838.4)	9.3 min
$SO_3Hel_1ALQEAA\text{-}NH_2$	M-: 917.0 (917.4)	9.7 min
$HHel_1AQSFLR\text{-}NH_2$	$(M+H)^+$: 958 (958.5)	10.6 min
$SO_3Hel_1AQSFLR\text{-}NH_2$	M^-: 1036 (1036.4)	13.6 min
$HHel_1AL^{\dagger}AQSR\text{-}NH_2$	$(M+H)^+$: 883 (883.5)	8.4 min
$SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$	M^-: 961 (961.4)	9.9 min

Appendix 1. Molar Ellipticities of $SO_3Hel_1CONH_2$ and $HHel_1CONH_2$ for Correction of CD Spectra of SO_3Hel_1-peptide and $HHel_1$-peptide Conjugates

Table 6. Molar ellipticities in deg cm^2 / dmol

Wavelength	[θ], $SO_3Hel_1CONH_2$	[θ], $HHel_1CONH_2$
270	85.744	175.57782
269	32.419	63.911818
268	-8.584	7.5968182
267	-38.105	-13.38018
266	-56.986	-19.03018
265	-66.067	-45.37418
264	-66.19	-110.2312
263	-58.196	-124.6242
262	-42.925	-91.01118
261	84.286	5.9588182
260	134.604	150.60582
259	156.552	207.78182
258	33.656	351.64082
257	-102.298	310.51982
256	-87.174	172.13182
255	-184.051	224.94382
254	-183.89	248.59882
253	-450.698	392.79682
252	-670.141	399.83882
251	-657.775	456.36082
250	-909.98	462.69382
249	-973.827	536.41382
248	-1250.101	558.57482
247	-1205.17	633.54382
246	-1243.21	760.86482
245	-1359.555	803.84182
244	-1584.366	779.42982
243	-1658.35	787.64582
242	-1785.02	634.54682
241	-1849.332	440.15782
240	-1825.516	236.55382
239	-1840.672	51.740818
238	-1661.363	-133.1732
237	-1336.893	-280.1132
236	-793.889	-446.7462

235	-157.339	-572.2082
234	540.494	-791.5562
233	1367.43	-1018.051
232	2515.939	-1380.296
231	3667.141	-1687.343
230	4832.552	-2032.63
229	6039.028	-2459.768
228	7208.088	-2825.299
227	8271.667	-3316.05
226	9243.83	-3952.146
225	9829.73	-4740.693
224	10141.779	-5494.658
223	10375.529	-6360.763
222	10036.48	-7085.43
221	9499.445	-7962.164
220	8865.31	-8657.296
219	7973.037	-9367.777
218	7122.415	-9965.991
217	6462.761	-10454.41
216	5936.073	-10624.93
215	5216.483	-10818.95
214	5060.596	-10693.18
213	5199.145	-10032.27
212	5511.355	-9116.84
211	5822.738	-7658.89
210	6656.373	-5578.685
209	7394.517	-3308.613
208	8429.472	-1446.547
207	9360.18	840.10882
206	10338.085	3452.7488
205	11401.184	6226.1998
204	12377.797	8848.9838
203	12897.742	11557.246
202	13303.976	13694.175
201	13754.809	15543.175
200	13609.266	17236.012
199	13752.352	18854.752
198	13517.785	20355.733
197	13152.1	21992.643
196	12845.558	23036.151
195	12677.903	24293.09
194	12728.886	25568.606
193	13078.253	26997.192

192	13805.754	28346.84
191	14991.136	29521.374
190	16714.148	30424.614

Appendix 2. An Abbreviated Template for Helix Initiation

The sequence SO_3ProAla contains most of the non-hydrogen atoms of SO_3Hel_1, except for part of the second pyrrolidine and the thiomethylene bridge (it lacks only C7, C8, C9, and S10). If it were able to adopt a conformation analogous to that of SO_3Hel_1, this simple dipeptide sequence could serve as a helix initiation sequence. To test this possibility, H-ProAlaSELSRL-NH_2 and SO_3ProAlaSELSRL-NH_2 were prepared (by the same solid phase peptide synthesis techniques and sulfamation chemistry as the other compounds; mass spec. for H-ProAlaSELSRL-NH_2 M^- = 869.42 observed, 870.49 calculated; mass spec. for SO_3ProAlaSELSRL-NH_2, M^- = 949.41 observed, 949.44 calculated). The values of $[\theta]_{222}$ for these two compounds are: $[\theta]_{222}$ = −970 deg cm^2 / dmol res for H-ProAlaSELSRL-NH_2 and $[\theta]_{222}$ = −3860 deg cm^2 / dmol res for SO_3ProAlaSELSRL-NH_2. Compare these to $[\theta]_{222}$ = −3215 deg cm^2 / dmol res for $HHel_1$SELSRL-NH_2 and $[\theta]_{222}$ = −7780 deg cm^2 / dmol res for SO_3Hel_1SELSRL-NH_2. $[\theta]_{222}$ is definitely more negative for the SO_3ProAla derivative than for its random coil model, indicating that it may induce some structure, but it is not much more negative for the SO_3ProAla derivative than for the $HHel_1$ derivative. Further experiments will be needed to establish whether and to what extent the SO_3ProAla motif initiates helicity.

Chapter 5. The Amide Hydrogen Exchange Technique and Its Use in Determining the Importance of Non-Canonical Interactions in Peptide Helicity.

5.1 Introduction

In the preceding chapter, peptide conjugates of a new N-terminal helix initiating template, SO_3Hel_1, were introduced and shown to be helical. The initiation constant of this template, B_{SO3}, was found to be 0.2, which is about 100 times larger than the spontaneous initiation constant, σ (recall from chapter 2 that σ is about 2×10^{-3}). In this chapter, SO_3Hel_1 is employed in a study of the amide hydrogen exchange technique[1] for assigning fractional site helicities to peptides. As noted in section 2.5, the determination of fractional site helicity first involves measurement of the rate constants, k_{obs}, for NH → ND exchange at each of a helical peptide's backbone amides. Each amide's measured rate constant is then divided into its intrinsic exchange rate constant, k_{int}, the exchange rate constant observed for the amide when it is completely solvent-exposed and not intramolecularly hydrogen bonded. The resulting quotient is called a protection factor (PF). Under the assumption that the exchange rates of backbone amide NH groups that are imbedded within a helical structure is zero, the protection factor is inversely related to the mole fraction of each amide NH that is not hydrogen bonded in a helix, f_{nhb}.[2,3,4,5] The

---

[1] Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1984**, *16*, 521.
[2] Zhou, H. X.; Hull, L. A.; Kallenbach, N. R.; Mayne, L.; Bai, Y.; Englander, S. W. *J. Am. Chem. Soc.* **1994**, *116*, 6482.

required fractional site helicity is simply $1 - f_{nhb}$. This series of relationships is summarized below

$$PF = \frac{k_{int}}{k_{obs}} = \frac{1}{f_{nhb}} = \frac{1}{1 - (\text{fractional site helicity})}$$

The slowing of amide hydrogen exchange by hydrogen bond formation is discussed in more detail in section 5.3.1.

Three problems are addressed in this chapter. First, to achieve a useful measure of peptide site helicity, obtaining an accurate value for k_{int} is critical. For peptides in which helicity is initiated with SO_3Hel_1, the same peptide attached to the non-initiating template $HHel_1$ (the synthetic precursor of SO_3Hel_1) provides a model system for measuring k_{int} values that is not helical, but is otherwise nearly identical to the system of interest. The values of k_{int} determined in $HHel_1$-peptide conjugates will be compared to a method for calculating k_{int} using literature data. Second, as seen in section 2.4, helices initiated by N-terminal templates fray monotonically from the N- to the C-terminus. This feature can be rigorously tested using the fractional site helicities measured in SO_3Hel_1-peptide conjugates. These two studies serve as a preamble for a third pilot study of a novel and very general protocol for exploring the full scope of context dependence in helix formation. Because the peptides designed for this pilot study are also used to address the first two problems, the subject of this pilot study is briefly introduced below.

---

[3] Rohl, C. A.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 7760.
[4] Shalongo, W.; Dugad, L.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 8288.
[5] Mayne, L.; Englander, S. W.; Qiu, R.; Yang, J.; Gong, Y.; Spek, E. J.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1998**, *120*, 10643.

5.1.1 Canonical and Non-Canonical Interactions in Peptide Helicity

The studies to date in which it has been attempted to elucidate the interactions that influence peptide helicity have largely (and fruitfully) focused on coulombic interactions between two charged residues or a charged residue and the helix dipole, N- or C-terminal capping interactions, and interactions between the side chains of residues that are on the same face of the helix (i to i+3 or i to i+4 interactions). The incorporation of corrections for such interactions, which will henceforth be referred to as canonical interactions, has greatly improved algorithms for predicting peptide helicity,[6,7,8,9,10,11,12] but these are not the only types of interactions that can affect helix formation. For example, evidence for i to i+1 interactions between adjacent residues was presented in chapter 3, and there is also evidence that coil-stabilizing i to i+2 hydrophobic interactions might exist.[13] Furthermore, interactions between the helix barrel and the side-chains of several residues (see section 2.6) have been proposed to affect helicity. The incorporation of corrections for such interactions, which will be referred to as non-canonical interactions, into the algorithms for predicting peptide helicity has only recently begun,[14] but how pressing is the need for these corrections? This depends on two factors: the magnitude of the interactions and their prevalence. If non-canonical interactions usually modulated s values by, say, less than 25 cal/mol then they could probably be safely ignored. If they

[6] Scheraga, H. A.; Vásquez, M. *Biopolymers* **1988**, *27*, 41.
[7] Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605.
[8] Muñoz, V.; Serrano, L. *Nature: Struct. Biol.* **1994**, *1*, 399.
[9] Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 3396.
[10] Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623.
[11] Shalongo, W.; Stellwagen, E. *Proteins: Struct. Funct. Gen.* **1997**, *28*, 467.
[12] Andersen, N. H.; Tong, H. *Protein Sci.* **1997**, *6*, 1920.
[13] Bundi, A.; Andreatta, R. H.; Wüthrich K. *Eur. J. Biochem.* **1978**, *91*, 201.
[14] Lacriox, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173.

only occurred under very rare circumstances then a model that ignored them would only result in the occasional outlier. If they were neither weak nor rare, however, then accounting for them could significantly improve the accuracy of helicity prediction.

The work in the third section of this chapter is an attempt to ascertain the importance of non-canonical interactions to peptide helicity using NH → ND exchange. Peptides with helicity induced from the N-terminus by SO_3Hel_1 are chosen for this work because, as noted in section 2.4, their helix-coil equilibria are much simpler than those of peptides with free initiation are. The amide hydrogen exchange technique is chosen for this work because it is capable of providing more detailed information about a peptide's helicity than any other technique. In the design phase of this pilot study, a group of eight SO_3Hel_1 linked hexapeptide amides was selected that met the following conditions. The internal residues of the peptides (residues 2 through 5) were spaced from the template by an alanine residue and were C-terminally capped by an arginine residue to ensure adequate solubility. The four internal residues were chosen from a set of five amino acids: Ala, Gln, Ser, Phe, and Leu. None of these have charged side chains, thus ensuring that effects due to charge-helix dipole interactions would be absent. The set includes amino acids that are commonly regarded as helix formers (Ala, Gln, Leu) as well as helix breakers (Phe and Ser), and it spans much of the polar and hydrophobic ranges of the amino acids. Finally, the sequences were planned to avoid known i to i+3 or i to i+4 canonical interactions. The eight sequences resulting from these criteria are:

AAQSFR	AASLQR
ALAQSR	ALSAFR
AQFASR	AQSFLR
ASFLAR	ASLQFR

According to the web-based helicity prediction program Helix 1.5, which incorporates all of the currently known canonical interactions (see ref. 12 or http://weber.u.washington.edu/~nielshan/helix/index.html), the only one of the above sequences that has a potential canonical interaction is ALSAFR, where L and F are positioned i and i+3 relative to each other.

A simple, preliminary test of the significance of non-canonical interactions within this group of peptides is whether the t/c ratios of their $AcHel_1$ conjugates can be calculated with reasonable accuracy using the s values from the literature listed in table 2 of chapter 2. The results from this test set the stage for an analysis of the protection factor data from the SO_3Hel_1 conjugates. An attempt will be made to fit the measured protection factors using a model that does not account for any context dependencies. If the error in the model's fit exceeds what one could possibly expect from the errors in the measurement of the protection factors, and since no known canonical interactions can take place in the peptides, then the excess error can be attributed to the presence of non-canonical interactions. Even if no evidence for non-canonical interactions is obtained, this is a pilot study, and no generalizations to larger databases would be appropriate. However, if strong evidence for non-canonical interactions were obtained from this study of only eight compounds with only six of the twenty amino acids represented, then this study

would have demonstrated that rigorous helicity-predicting algorithms cannot be constructed that omit such interactions. More refined studies could then be designed to characterize the structural origins and energetics of non-canonical interactions. The details of the design, results and conclusions of this study follow in section 5.4.

5.2 Amide Hydrogen Exchange

Pioneered by Linderstrøm-Lang and co-workers in the 1950s,[15,16,17,18,19,20] the study of isotopic exchange between water and a protein's backbone amide hydrogens has long been used to probe the degree shielding from solvation of NH protons in the interiors of folded proteins. Protection factors of many orders of magnitude are typically observed in such experiments. More recently, this technique has been refined by the application of modern ^{1}H-NMR spectroscopy to the monitoring of the exchange process, which allows the resolution of the exchange kinetics of individual amides. As noted in section 2.5, the foundation of the technique is a comparison between the exchange rate observed for a particular amide NH when in a structured state and the same amide's intrinsic exchange rate (its exchange rate when it is fully exposed to solvent). With the aim of improving the accuracy of the method, Englander and co-workers[21] have carried out extensive studies on the exchange properties of short peptides, culminating in tabulated data and a simple algorithm for calculating the pH-dependent intrinsic exchange rates for amides in

---

[15] Hvidt, A.; Linderstrøm-Lang, K. *Biochim. et Biophys. Acta* **1954**, *14*, 574.
[16] Hvidt, A.; Linderstrøm-Lang, K. *Biochim. et Biophys. Acta* **1955**, *16*, 168.
[17] Hvidt, A.; Linderstrøm-Lang, K. *Biochim. et Biophys. Acta* **1955**, *18*, 308.
[18] Krause, I. M.; Linderstrøm-Lang, K. *Compt. rend. trav. lab. Carlsberg Ser. Chim.* **1955**, *29*, 367.
[19] Linderstrøm-Lang, K. In *Symposium on Protein Structure*, Neuberger, A. ed. Methuen: London, **1958**.
[20] Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. *Protein Sci.* **1997**, *6*, 1101.

any peptide sequence. This algorithm has provided an invaluable foundation for the quenched and pulse-labeling hydrogen exchange techniques for studying protein folding pathways, and the native state hydrogen exchange technique for studying the dynamics of the native state.[22,23,24]

We are interested in using the fractional site helicities measured by the amide hydrogen exchange technique to probe the effects of amino acid composition and sequence on the helicity of short peptides. In such studies, the accuracy of the values used for intrinsic exchange rates is critical. Protection factors for the amides in peptide helices (where the fractional site helicity of residues is typically between 0 and 0.9) are likely to be between 1 and 10 or 20. In this range, an error of as little as 30% in an amide's k_{int} could lead to a major inaccuracy in the fractional site helicity assigned to that amide. Take for example an amide whose actual k_{int} is 0.10 min^{-1} and whose k_{obs} is 0.09 min^{-1}. The actual protection factor of the amide should be 1.11, corresponding to a fractional site helicity of 0.09; however, if k_{int} were mistakenly believed to be 0.13 min^{-1}, the protection factor would be 1.44, corresponding to a fractional site helicity of 0.30. Here the error of 30% in k_{int} introduces an error of a factor of 300% in fractional site helicity. If such an error were made consistently, then a peptide that is mostly in the random coil state could mistakenly be believed to be substantially helical (or vice versa if k_{int} were mistakenly underestimated). Such considerations are not nearly as important to the qualitative interpretation of hydrogen exchange data in proteins, where the protection factors are often > 1000. In this range, an error of a factor of two in a protection factor,

---

[21] Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 75.
[22] Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. *Curr. Opin. Struct. Biol.* **1996**, *6*, 18.
[23] Raschke, T. M.; Marqusee, S. *Curr. Opin. Biotech.* **1998**, *9*, 80.
[24] Clarke, J.; Itzhaki, L. S. *Curr. Opin. Struct. Biol.* **1998**, *8*, 112.

say 1000 vs. 2000, would not result in a misassignment of structure, as both of these protection factors would indicate highly structured regions.

The ideal model system in which to measure intrinsic exchange rates for the amides in a helical peptide would be the identical peptide that was somehow constrained to be in a random coil conformation. For medium and long peptides that have detectable helical character due to spontaneous helix initiation, this model is unrealizable. A given peptide is either helical or it is not. By contrast, the amide hydrogen exchange rate constants in short peptides linked to an efficient helix-initiating template, such as SO_3Hel_l, can be compared to those in the identical peptides attached to similar non-initiating templates, such as $HHel_l$. In this section, results are reported of a comparison between the exchange rates derived from the Englander algorithm with those measured in $HHel_l$-peptide conjugates of the peptides listed above. First, though, the mechanistic basis of the exchange process, the experimental variables that affect exchange rates, the experimental measurement protocol, and the nature and scope of the Englander algorithm are all reviewed.

5.2.1 The Mechanism of Amide Hydrogen Exchange

In aqueous solution, the NH proton of peptide groups, like any other labile proton, exchanges with the protons of solvent water. This reaction is dominated by specific acid and base catalysis,[25,26] but general acid/base catalysis has been observed under some

[25] Berger, A.; Loewenstein, A.; Meiboom, S. *J. Am. Chem. Soc.* **1959**, *81*, 62.
[26] Englander, S. W.; Poulsen, A. *Biopolymers* **1969**, *7*, 379.

conditions.[27,28,29] General catalysis appears to be minimal for aqueous phosphate,[29] the buffer used in these studies. Water catalysis also makes a very small, pH independent contribution to the reaction rate.[21,30,31] The overall rate law is[21]:

$$\text{Exchange Rate} = \left(k_{acid}[D^+] + k_{base}[OD^-] + k_{water}\right)[\text{amide}]$$

$$= k_{total}[\text{amide}]$$

where k_{total} is the pseudo-first order rate constant for the reaction that would be found by fitting observed data to an exponential decay equation. The three constituent rate constants of k_{total} can be determined by measuring k_{total} as a function of pD. Base catalysis is much more efficient ($k_{base} \sim 10^{10}$) than acid catalysis ($k_{acid} \sim 10^{1.5}$), which is in turn much more efficient than water catalysis ($k_{water} \sim 10^{-2.5}$). The minimum in the reaction rate with respect to pH is thus usually between pH 3.0 and 3.5, depending on the amide's substituents (the half-time for exchange is between 30 min and 2 h for typical, solvent exposed peptide amides at the pH minimum and 5 °C).

The acid and base catalyzed mechanisms for the hydrogen exchange of peptide amides are illustrated in figure 1. The base catalyzed reaction is believed to occur by a rate-limiting deprotonation of the amide NH by hydroxide to produce the imidate ion, followed by reprotonation.[25] The acid catalyzed mechanism has been more controversial, as it has been difficult to establish whether the relevant pathway involves O- or N-

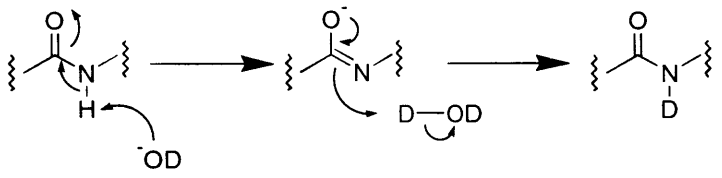
---

[27] Klotz, I. M.; Frank, B. H. *J. Am. Chem. Soc.* **1965**, *87*, 2721.
[28] Leichtling, B. H.; Klotz, I. M. *Biochemistry* **1966**, *5*, 4026.
[29] Wang, W.-H.; Cheng, C.-C. *Bull. Chem. Soc. Jap.* **1994**, *67*, 1054.
[30] Englander, J. J.; Calhoun, D. B.; Englander, S. W. *Anal. Biochem.* **1979**, *92*, 517.
[31] Gregory, R. B.; Crabo, L.; Percy, A. J.; Rosenberg, A. *Biochemistry* **1983**, *22*, 910.

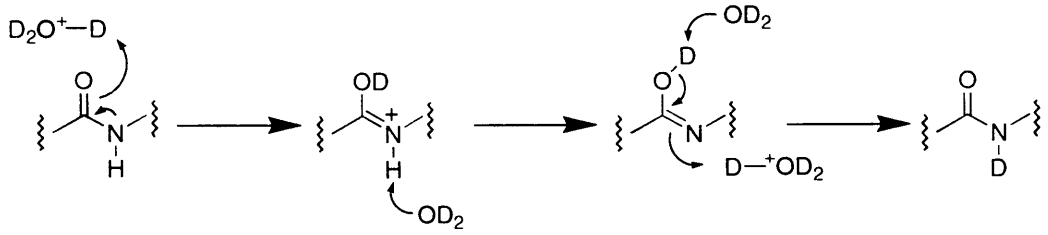
protonation.[32,33] It is currently believed that in secondary amides, acid catalyzed exchange occurs by the mechanism shown in figure 1 in which the amide is first O-protonated to form an imidic acid intermediate.[3,33,34] Deprotonation followed by tautomerization yields the exchanged amide. (Note that primary amides, unlike peptide amides, appear to exchange through an N-protonation mechanism).

Figure 1. Base and acid catalyzed amide hydrogen exchange mechanisms.

Base Catalyzed Exchange



Acid Catalyzed Exchange



[32] Cross, D. G.; Brown, A.; Fisher, H. F. *J. Biol. Chem.* **1976**, *251*, 1785.
[33] Perrin, C. L. *Acc. Chem. Res.* **1989**, *22*, 268.
[34] Tüchsen, E.; Woodward, C. *J. Mol. Biol.* **1985**, *185*, 421.

5.2.2 Factors that Affect Amide Hydrogen Exchange

In addition to the strong pH dependence noted above, the intrinsic rates of hydrogen exchange for peptide backbone NHs depend on at least four other variables as well. First, they depend on temperature. The activation energies for acid, base, and water catalyzed exchange are about 14, 17, and 19 kcal/mol respectively.[21] This translates to exchange rates increasing by a factor of about 3 when the temperature is increased from 5 °C to 20 °C. Second, the intrinsic exchange rates depend on ionic strength (μ). The extent of this dependence is apparently a function of local charge. Ionic strength effects are small for amides associated with neutral residues in mostly uncharged peptides, where k_{int} only increases by about 15% as μ is increased from 0.1 to 0.5 M.[21] For amides associated with lone charged residues in otherwise neutral peptides, k_{int} increases by about 30% over the same range.[21] Larger salt effects on amide exchange rates have been observed, but only in proteins,[34,35] peptide polyelectrolytes (poly-Lys),[36] and a highly charged short peptide[37] (sequence: H_3N^+-ISMSEEDLLNAK-COO^-, with six charges in twelve residues). Third, peptide amide exchange rates depend on the lengths of the peptides in which they are measured.[21] Amides in shorter peptides exchange more quickly than amides in longer peptides. This effect amounts to a factor of about 2 between the exchange rates of Ac-Ala_3-NHMe and poly-DL-alanine (average degree of polymerization = 28). Fourth, and most importantly, peptide amide exchange rates depend strongly on the types of residues from which the CO and NH portions of the amide are derived (the CO and NH donors,

[35] Christoffersen, M.; Bolvig, S.; Tüchsen, E. *Biochemistry* **1996**, *35*, 2309.
[36] Kim, P. S.; Baldwin, R. L. *Biochemistry* **1982**, *21*, 1
[37] Koide, S.; Jahnke, W.; Wright, P. E. *J. Biomol. NMR* **1995**, *6*, 305.

respectively).[21,38] For example, residues with electron withdrawing side chains, such as serine, cysteine, or asparagine, lower the amide's pK_a and therefore cause the reaction rate minimum to shift to lower pH. In addition, residues with bulky substituents, such as phenylalanine, leucine, or valine, retard exchange perhaps by sterically shielding amides from the solvent. The effect of primary structure can change amide exchange rate constants by more than a factor of ten, or shift the pH minimum of the reaction by more than 0.5 pH units.

5.2.3 The Englander Model for Calculating Intrinsic Amide Hydrogen Exchange Rate Constants

In the model due to Englander and co-workers,[21] k_{int} values for peptide amides can be computed as functions of the residues from which the amide CO and NH groups were derived and peptide length. In this model, reference rates for the acid, base, and water catalyzed exchange reactions were measured in alanine peptides, and these reference rates were taken to be additively modulated by substituent effects due to having a residue other than alanine as the source of the amide CO or NH. Thus for an amide X-CO-NH-Y where residue X donates the CO group and Y donates the NH group,

[38] Molday, R. S.; Englander, S. W.; Kallen, R. G. *Biochemistry* **1972**, *11*, 150.

$$\log k_{acid} = \log k_{acid,ref} + (\text{CO effect of X})_{acid} + (\text{NH effect of Y})_{acid}$$

$$\log k_{base} = \log k_{base,ref} + (\text{CO effect of X})_{base} + (\text{NH effect of Y})_{base}$$

$$\log k_{water} = \log k_{water,ref} + (\text{CO effect of X})_{base} + (\text{NH effect of Y})_{base}$$

Note that the substituent effects on the water catalyzed reaction are the same as those on the base catalyzed reaction. The total pseudo-first order rate constant for the exchange reaction in a deuterated medium, such as that used in this work, is

$$k_{int,calc} = k_{acid}[D^+] + k_{base}[OD^-] + k_{water}$$

The substituent effects of the naturally occurring amino acids were measured in compounds of the form AcNH-X-CONHMe by taking the effect of X on the AcN<u>H</u> and CON<u>H</u>Me exchange reactions to be the same as the NH and CO effects of X in real peptides. The effect of peptide length on the exchange rate constant has been accounted for by using different sets of values for $k_{acid,ref}$, $k_{base,ref}$, and $k_{water,ref}$ for short and long peptides. The reference rate constants for short peptides were obtained from $AcAla_3NHMe$ and those for long peptides from poly-DL-alanine (PDLA; average degree of polymerization = 28). The effect of ionic strength can be taken into account by using $AcAla_3NHMe$ or PDLA reference rates measured at either high salt (0.5 M NaCl) or low salt conditions (around 0.1 M NaCl). Finally, the effect of temperature can be taken into account using the activation energies for acid, base, and water catalyzed exchange.

The hypothesis on which this model rests, that the side-chain effects on log k_{acid}, log k_{base}, and log k_{water} are additive, was tested by Englander and co-workers[21] by comparing predicted and measured exchange rate constants for the amides in three short peptides at

several pHs and 5 °C. Good agreement was observed in most cases, although significant deviations were found for the N-terminal residues. Koide, et al.[37] have also tested the model by comparing the predicted and measured exchange rates in a twelve residue peptide at pH 7.0 and 20 °C. They found that the predicted exchange rates could deviate from the measured by as much as a factor of two.

Accuracy to within a factor of two in k_{int} is easily good enough for studies of hydrogen exchange in proteins, where protection factors are often on the order of several thousand or larger. However, it is not accurate enough for studies of peptide helicity, where such errors could be as large as the overall slowing of the exchange reaction by helix formation. In the following sections, the measured intrinsic exchange rate constants of the amide NHs in the eight $HHel_1$-peptide conjugates are compared to the predicted intrinsic exchange rate constants from the Englander model. This will show whether the deviations between the predicted and model exchange rate constants found by Koide, et al. are specific to their particular peptide, or whether this level of error is inherent in the model.

5.2.4 Comparison of Exchange Rate Constants Measured in $HHel_1$-peptide Conjugates to Those Calculated from the Literature Model

5.2.4a Technique for Measuring Exchange Rate Constants

In order to make an amide by amide comparison of predicted and measured intrinsic exchange rate constants, a method is needed that can resolve the exchange reactions of several amides in any given peptide is needed. The only method capable of such discrimination is NMR, where separate resonances can be observed for each amide. Therefore, amide exchange rate constants are measured in this work by dissolving the peptides of interest in buffered D_2O solutions and tracking the disappearance of the amide peaks as the exchange reaction replaces the protons of the amide NH groups with deuterons from the solvent. The decay of the peak's intensity with time is fitted to a single exponential decay to obtain the exchange rate constant of interest. If two peaks overlap, the decay of their combined intensities is fitted to a double exponential decay equation to obtain both of the pertinent rate constants, although it might not then be obvious which rate corresponds to which amide. The assignment of the rate constants obtained from double exponential decays requires external information (see the experimental section). Once the rate data have been obtained, the experimental error that attends their measurement can be assessed by an analysis of variance on duplicate or triplicate determinations of the rate constants. It should be noted here, though, that the

variabilities of the rate constants measured by this technique are proportional to their magnitudes, as is apparent from inspection of the data from the measurement of individual exchange rate constants in table 15 of appendix 1. A logarithmic transformation, therefore, is applied to the rate constants to stabilize the variance before all analyses.

This NMR based method is best for measuring the rates of reactions whose half-lives are greater than 5 minutes, as faster reactions are too close to completion after the dead time of the experiment (~ 10 to 15 min). The exchange reactions were therefore carried out near their pD minima in a 200 mM potassium phosphate buffer (low salt conditions in the Englander model) adjusted to around pD 3.0 and at 5 °C to ensure that the reaction rates were in the optimal range for the method.

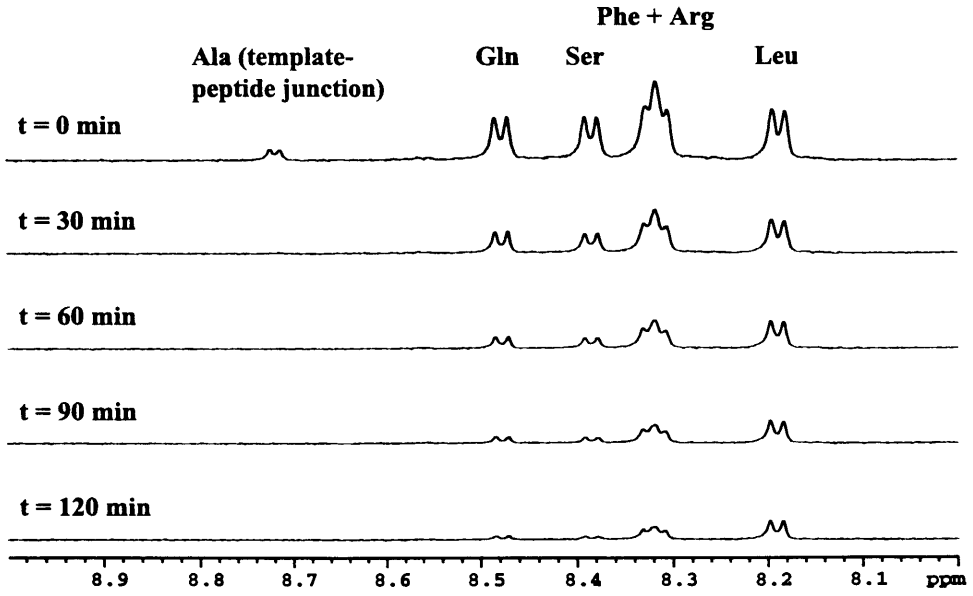
5.2.4b Intrinsic Exchange Rate Constants Measured in $HHel_1$-peptide Conjugates

The eight $HHel_1$-peptide conjugates for which exchange rate constants were measured are:

$HHel_1A^{\dagger}AQSFR\text{-}NH_2$	$HHel_1AASLQR\text{-}NH_2$
$HHel_1AL^{\dagger}AQSR\text{-}NH_2$	$HHel_1ALSAFR\text{-}NH_2$
$HHel_1AQFASR\text{-}NH_2$	$HHel_1AQSFLR\text{-}NH_2$
$HHel_1ASFL^{\dagger}AR\text{-}NH_2$	$HHel_1ASLQFR\text{-}NH_2$

(The †A residues in some of the above $HHel_1$-peptide conjugates are alanines with the α amine isotopically labeled with ^{15}N, to ensure that the two alanine amide resonances in the 1H-NMR spectra of these compounds could be distinguished. As it happened, the alanine amide resonance at the template peptide junction had a unique chemical shift, so the labeling was not necessary for every peptide.) A set of spectra representative of those used to determine the k_{int} values of the amides in $HHel_1$-peptide conjugates is shown in figure 2. In particular, the spectra shown in the figure are from the exchange reaction of $HHel_1AQSFLR\text{-}NH_2$ (pD 3.0, 5 °C).

Figure 2. Decay of amide NH peak intensities in the ^{1}H-NMR spectrum of $HHel_1AQSFLR\text{-}NH_2$ in 200 mM phosphate buffer, pD 3.0, 5 °C.



The Gln, Ser, and Leu amide resonances[39] are well separated and the k_{int} for each of these amides can be determined simply by fitting their integrals from successive times to

[39] All peak assignments were made using total correlation spectroscopy (TOCSY).

single exponential decay equations. The amide peaks of Phe and Arg, in contrast, overlap so that their rate constants have to be determined simultaneously by fitting their integrals to a double exponential decay equation. This is straightforward, but the assignment of the two resulting k_{int} values to the two amides does not follow directly from the fitting. Other information has to be used in order to make the assignment. In this case, it can be made on the basis of a priori expectations: the Phe amide, because of the bulky aromatic side chain, is expected to exchange more slowly than the Arg amide, so the smaller of the two rate constants is assigned to the Phe amide. Other assignment issues from double exponential decay fits can be similarly resolved- see the experimental section for further details. Notice that the signal for the amide at the template peptide junction is already very weak in the first spectrum at t = 0, indicating that the reaction progresses substantially during the dead time of the experiment (~ 10 - 15 min). While this amide exchanges too quickly for its k_{int} to be accurately measured by this method, it can be estimated to be around 0.1 min^{-1} from the limited data available. Also, notice that no signal whatsoever is seen for the C-terminal amide. This complete lack of data allows only the conclusion that the k_{int} for this amide must be much larger than 0.1 min^{-1}.

The measurement of the k_{int} values of the amides in the other seven $HHel_1$-peptide conjugates proceeded in much the same way. About half of the amide resonances were well enough isolated to determine their k_{int} values from single exponential fits while the rest were obtained from double exponential fits. In only one case did three resonances coincide- the Ser, Gln, and Arg of $HHel_1$ASLQFR-NH_2. None of the k_{int} values for these residues could be determined. In addition, the amides at the template-peptide junctions and the C-termini always exchanged too quickly to measure their k_{int} values. In all, the

k_{int} values for 37 amides in the eight $HHel_1$-peptide conjugates (five from each peptide except for $HHel_1$ASLQFR-NH_2 for which only two k_{int} values could be obtained) were successfully measured either in duplicate or triplicate. The average values of k_{int} and ln k_{int}, and the standard errors in ln k_{int} are listed in table 1. The data for all of the individual exchange reactions are recorded in table 15 of appendix 1.

Table 1. Amide exchange rate constants for $HHel_1$-peptide conjugates measured at pD 3.0 ± 0.15 and 5 °C. In each cell of the table is given the type of residue to which the amide NH belongs, the exchange rate constant k_{int}, and in parenthesis ln k_{int} and the standard error in ln k_{int} (either ± 0.09 or ± 0.07 depending on whether two or three replicate measurements were made).

$HHel_1$-Peptide	Exchange Rate Constants by Amide Position					
	1st	2nd	3rd	4th	5th	6th
AAQSFR	Ala ~ 0.1 min^{-1}	ǂAla 0.0217 min^{-1} (-3.83±0.07)	Gln 0.0170 min^{-1} (-4.08±0.07)	Ser 0.0215 min^{-1} (-3.84±0.07)	Phe 0.0140 min^{-1} (-4.27±0.07)	Arg 0.0165 min^{-1} (-4.10±0.07)
AASLQR	Ala ~ 0.1 min^{-1}	Ala 0.0196 min^{-1} (-3.93±0.07)	Ser 0.0180 min^{-1} (-4.02±0.07)	Leu 0.0064 min^{-1} (-5.05±0.07)	Gln 0.0117 min^{-1} (-4.45±0.07)	Arg 0.0123 min^{-1} (-4.40±0.07)
ALAQSR	Ala ~ 0.1 min^{-1}	Leu 0.0073 min^{-1} (-4.93±0.09)	ǂAla 0.0157 min^{-1} (-4.16±0.09)	Gln 0.0142 min^{-1} (-4.25±0.09)	Ser 0.0226 min^{-1} (-3.79±0.09)	Arg 0.0207 min^{-1} (-3.88±0.09)
ALSAFR	Ala ~ 0.1 min^{-1}	Leu 0.0062 min^{-1} (-5.09±0.07)	Ser 0.0126 min^{-1} (-4.37±0.07)	Ala 0.0124 min^{-1} (-4.39±0.07)	Phe 0.0100 min^{-1} (-4.60±0.07)	Arg 0.0141 min^{-1} (-4.26±0.07)
AQFASR	Ala ~ 0.1 min^{-1}	Gln 0.0187 min^{-1} (-3.98±0.07)	Phe 0.0097 min^{-1} (-4.64±0.07)	Ala 0.0102 min^{-1} (-4.58±0.07)	Ser 0.0201 min^{-1} (-3.91±0.07)	Arg 0.0165 min^{-1} (-4.10±0.07)
AQSFLR	Ala ~ 0.1 min^{-1}	Gln 0.0155 min^{-1} (-4.17±0.09)	Ser 0.0159 min^{-1} (-4.14±0.09)	Phe 0.0098 min^{-1} (-4.62±0.09)	Leu 0.0050 min^{-1} (-5.30±0.09)	Arg 0.0116 min^{-1} (-4.46±0.09)

ASFLAR	Ala	Ser	Phe	Leu	†Ala	Arg
	~ 0.1 min^{-1}	0.0206 min^{-1}	0.0099 min^{-1}	0.0045 min^{-1}	0.0155 min^{-1}	0.0132 min^{-1}
		(-3.88±0.07)	(-4.62±0.07)	(-5.41±0.07)	(-4.17±0.07)	(-4.32±0.07)
ASLQFR	Ala	Ser	Leu	Gln	Phe	Arg
	~ 0.1 min^{-1}	-	0.0064 min^{-1}	-	0.0088 min^{-1}	-
			(-5.05±0.07)		(-4.73±0.07)	

The standard errors for ln k_{int} reported in the table correspond to roughly 7% and 9% error in k_{int}, which speaks well of the method's reproducibility. The largest of the exchange rate constants in table 1, those associated with the amides of Ser and Ala, are about four times larger than the smallest, those associated with the amides of Leu and Phe. This shows how influential neighboring side chains can be on an amide's exchange rate.

5.2.4c Evidence for the Lack of Structure in $HHel_1$-peptide Conjugates

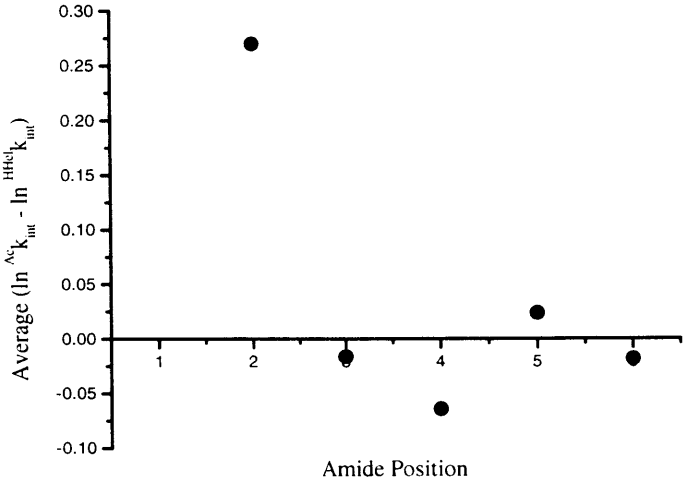
To check whether the nature of the N-terminal group had a significant impact on the measured exchange rate constants, the k_{int} values measured in $HHel_1A^{\dagger}AQSFR\text{-}NH_2$, $HHel_1AL^{\dagger}AQSR\text{-}NH_2$, $HHel_1AQSFLR\text{-}NH_2$, and $HHel_1\text{-}ASFL^{\dagger}AR\text{-}NH_2$ were compared to those measured in the corresponding peptides in which a simple acetyl group replaced $HHel_1$: $AcA^{\dagger}AQSFR\text{-}NH_2$, $AcAL^{\dagger}AQSR\text{-}NH_2$, $AcAQSFLR\text{-}NH_2$, and $AcASFL^{\dagger}AR\text{-}NH_2$. The k_{int} values for the amides in these peptides ($^{Ac}k_{int}$) were measured in the same way as for the $HHel_1$-peptide conjugates, and the averages (for duplicate data) or the lone values (for singleton data) are reported in table 2. The corresponding $^{HHel}k_{int}$ values determined in $HHel_1$-peptides are included for comparison, and in figure 3 the average

differences ln $^{Ac}k_{int}$ – ln $^{HHel}k_{int}$ are plotted as a function of amide position. As in the $HHel_1$-peptide conjugates, the exchange reaction of the first amide (the amide between the acetyl group and the first alanine) occurred too quickly to be measured. The k_{int} of the second amide tends to be somewhat larger (~ 20% on average) in the acetylated peptides than in the $HHel_1$-peptide conjugates. All of the other amides in the acetylated peptides have exchange rate constants that are slightly smaller (by ~ 9% on average) but comparable to those of their counterparts in the $HHel_1$-peptide conjugates.

Table 2. Comparison of exchange rate constants at pD 3.0 ± 0.15 and 5 °C determined in Ac-peptides and $HHel_1$-peptide conjugates.

Peptide	N-terminus	Amide Exchange Rate Constants (min^{-1}) by Position					
		1st	2nd	3rd	4th	5th	6th
A†AQSFR	Ac	-	0.0242	0.0134	0.0234	0.0116	0.0144
	$HHel_1$	~ 0.1	0.0217	0.0170	0.0215	0.0140	0.0165
AL†AQSR	Ac	-	0.0090	0.0141	0.0121	0.0234	0.0197
	$HHel_1$	~ 0.1	0.0073	0.0157	0.0142	0.0226	0.0207
AQSFLR	Ac	-	0.0220	0.0171	0.0075	0.0048	0.0130
	$HHel_1$	~ 0.1	0.0155	0.0159	0.0098	0.0050	0.0116
ASFL†AR	Ac	-	0.0224	0.0099	0.0035	0.0147	0.0121
	$HHel_1$	~ 0.1	0.0206	0.0099	0.0045	0.0155	0.0132

Figure 3. Average difference between ln $^{Ac}k_{int}$ and ln $^{HHel}k_{int}$ (both measured in 200 mM phosphate, pD 3.0 ± 0.15, 5 °C) as a function of amide position.



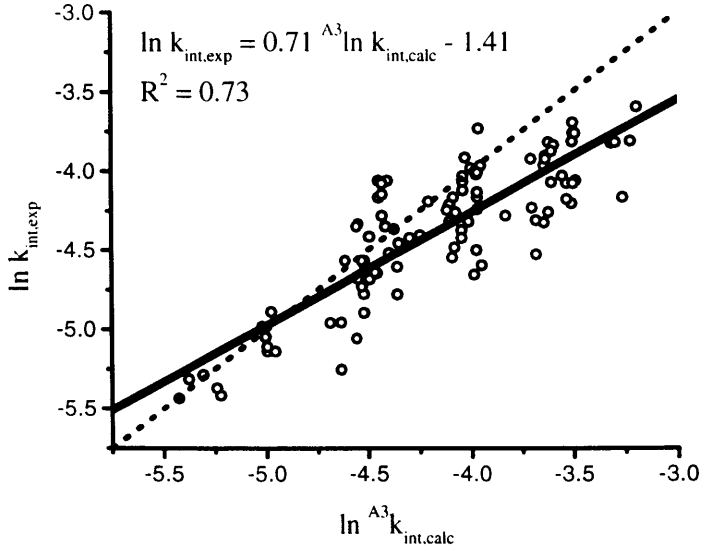
The only position at which substantial and consistent differences are observed between $^{HHel}k_{int}$ and $^{Ac}k_{int}$ is the position closest to the N-terminal capping group. This indicates that the nature of the N-terminus could have a significant effect on the hydrogen exchange rates of the amides closest to it. (Similar N-terminal effects were observed by Koide, et al., who found particularly large deviations between measured exchange rate constants and exchange rate constants calculated using the Englander model for the N-terminal residues in the peptide H_3N^+-ISMSEEDLLNAK-COO^-.[37]) The agreement between $^{HHel}k_{int}$ and $^{Ac}k_{int}$ at all other positions, however, is very good, indicating that the nature of the N-terminus does not affect the exchange reactions of amides that are more than two residues away from it.

5.2.5 Comparison of Measured with Calculated Values of k_{int}

Each particular measurement of each amide's experimental k_{int} ($k_{int,exp}$) was compared with the k_{int} that was calculated at the relevant pD using the alanine tripeptide reference rates ($^{A3}k_{int,calc}$). (Note that the individual measured exchange rate constants listed in table 15 of appendix 1, not the average exchange rate constants in table 1, were used for the comparison.) The natural logarithms of $k_{int,exp}$ are plotted versus the natural logarithms of $^{A3}k_{int,calc}$ in figure 4. The natural logarithms of the exchange rate constants are plotted instead of the rate constants because the errors in the natural logarithms should be normally distributed, while the error in the rate constants themselves are not. The line of best fit through the data is shown as a solid line in the figure. Its equation is:

$$\ln k_{int,obs} = 0.71 \times \ln {}^{A3}k_{int,calc} - 1.41$$

Figure 4. Plot of ln $^{A3}k_{int,calc}$ vs. ln $k_{int,exp}$. The solid line is the line of best fit through the data (equation given in the upper left). For comparison, a dotted line is shown for the ideal case where ln $k_{int,exp}$ = ln $^{A3}k_{int,calc}$.



The plot shows that there is a moderate correlation between the observed and calculated data, and this is reflected in the coefficient of determination of the line of best fit (R^2 = 0.73). However, there is considerable scatter. The measured and predicted exchange rate constants differ by more than 65% for 18 out of 101 measurements (the difference between their natural logarithms ≥ 0.5), and by more than 35% for 36 out of 101 measurements (the difference between their natural logarithms ≥ 0.3). Excluding the exchange rate constants determined for the amides at the second position, those that were found to differ between $HHel_1$- and Ac-peptides, does not improve the correlation (R^2 excluding 2nd amides = 0.72). In addition, most of the data points in the plot fall below line for the ideal case in which ln $k_{int,exp}$ = ln $^{A3}k_{int,calc}$, shown as a dotted line in figure 4. This shows that ln $k_{int,exp}$ has a strong tendency to be less negative than ln $^{A3}k_{int,calc}$ (ln $k_{int,exp}$ – ln $^{A3}k_{int,calc}$ = 0.21 on average), and that the exchange reaction therefore usually occurs more slowly than predicted based on the calculations using tripeptide reference rates. Using the PDLA reference rates instead of the alanine tripeptide reference rates to compute $k_{int,calc}$ does little to improve the agreement between the measured and calculated exchange rates. The amount of scatter remains about the same, but instead of being generally too large, the calculated exchange rate constants are generally too small.

5.2.6 Conclusions Regarding the Calculation of Intrinsic Amide Exchange Rate Constants from the Englander Model.

The calculated values of k_{int} have two shortcomings: first, the exchange rate constants for the amides in the $HHel_1$-peptide conjugates are generally smaller than those

calculated using tripeptide reference rates but larger than those calculated using PDLA reference rates, and second, the measured rate constants correlate only moderately well with the calculated rate constants. The first problem has an obvious rationalization. Since the $HHel_1$-peptide conjugates are intermediate in length between $AcAla_3NHMe$ and PDLA, the exchange rate constants of the amides in the $HHel_1$-peptide conjugates are intermediate between those calculated using reference rates from $AcAla_3NHMe$ and PDLA. If this were correct, a more elaborate length correction would have to be incorporated into the model for calculating exchange rate constants. The second problem, the scatter between the measured and calculated exchange rate constants, requires a more involved explanation. Perhaps the substituent effects on amides in real peptides are somewhat different than in the model compounds (the acetylated amino acid N-methyl amides) from which the CO and NH effects used in the computation of $^{A3}k_{int,calc}$ were derived. Alternatively, it could be that interactions between side-chains cause deviations from the strict additivity of their effects. Either way, further study is clearly required for a more complete understanding of the factors that influence amide exchange rate constants.

In summary, computed exchange rate constants do not quantitatively reproduce experimentally measured exchange rate constants. The deviation can probably be attributed to a combination of length effects and substituent effects that are not well enough accounted for in the model. The accuracy of the exchange rate constants calculated using the model is certainly adequate for studies of protein structure, where the observed error in k_{int} is likely to be small compared to the protection factors. However,

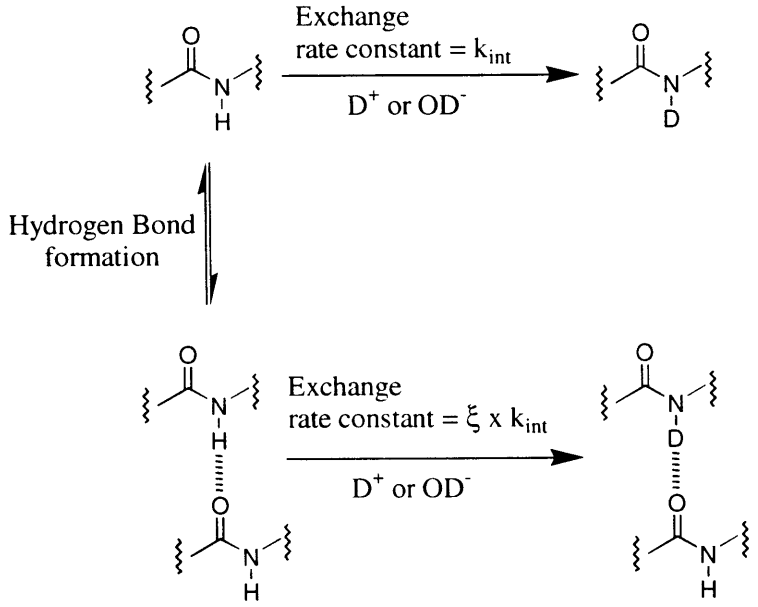
the use of computed intrinsic exchange rate constants could introduce substantial errors into studies of peptide helicity.

5.3 Helix Fraying in SO_3Hel_1-peptide Conjugates Studied by Amide Hydrogen Exchange

With the values of k_{int} from $HHel_1$-peptide conjugates in hand, the exchange rate constants measured in SO_3Hel_1-peptide conjugates can be measured and used to calculate protection factors. Since the helices induced by SO_3Hel_1 in linked peptides should be increasingly frayed from the N- to the C-terminus, the fractional site helicities, and therefore the protection factors, in SO_3Hel_1-peptide conjugates should decrease from the N- to the C-terminus. This condition is discussed in detail and checked experimentally in the present section.

5.3.1 The Effects of Hydrogen Bonding on Amide Hydrogen Exchange

Figure 5. The effect of hydrogen bonding on amide hydrogen exchange.



It has already been noted that hydrogen bond formation between amide NHs and non-solvent acceptors can slow hydrogen exchange. This is because the protonation and deprotonation steps of the hydrogen exchange reaction all require that the amide be accessible to the catalytic species. The formation of structures by peptides or proteins that inhibit this access can slow hydrogen exchange.[1,19,40,41] For example, hydrogen bond formation between an amide NH and a non-solvent hydrogen bond acceptor impedes both acid and base catalyzed hydrogen exchange by reducing the amide's capacity to hydrogen bond to the catalytic species (D^+ and OD^-) and by adding the cost of disrupting the

[40] Hvidt, A; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287.

hydrogen bond to the activation barrier for the proton's removal. Similarly, hydrogen bond formation between the amide CO and a non-solvent hydrogen bond donor impedes acid catalyzed hydrogen exchange (although it does not slow the base catalyzed reaction).[3] The effect of hydrogen bond formation on amide hydrogen exchange is summarized in figure 5. This figure depicts an equilibrium between two states of some peptide or protein, one unstructured in which an amide of interest is not hydrogen bonded and the other structured in which the amide is hydrogen bonded. In the limit where the hydrogen bond forming reaction is much faster than the exchange reaction (known as the EX2 limit),[40] so that structure formation is at equilibrium while exchange occurs, the rate of amide hydrogen exchange is:

$$\text{Exchange Rate} = k_{int}[\text{non-hydrogen bonded form}] + \xi \times k_{int}[\text{hydrogen bonded form}]$$

$$= k_{int} \times f_{nhb} \times [\text{total amide}] + \xi \times k_{int} \times f_{hb} \times [\text{total amide}]$$

$$= (k_{int} \times f_{nhb} + \xi \times k_{int} \times f_{hb})[\text{total amide}]$$

where f_{nhb} is the fraction of the population of amides that is not hydrogen bonded, f_{hb} is the fraction that is hydrogen bonded, k_{int} is the intrinsic exchange rate constant, and ξ is the fraction of the intrinsic rate at which amides in the hydrogen bonded state exchange. If the amide of interest is completely unable to exchange its proton with solvent while in the structured, hydrogen-bonded form, then ξ approaches zero and the expression for the exchange rate simplifies to

---

[41] Barksdale, A. D.; Rosenberg, A. *Methods Biochem. Anal.* **1982**, *28*, 1

$$\text{Exchange Rate} = k_{int} \times f_{nhb} \times [\text{total amide}]$$

$$= k_{obs} \times [\text{total amide}]$$

where the apparent rate constant for the exchange reaction is simply $k_{obs} = k_{int} \times f_{nhb}$. Given k_{int}, then f_{nhb} can be determined through the simple relation:

$$f_{nhb} = \frac{k_{int} \times f_{nhb}}{k_{int}} = \frac{k_{obs}}{k_{int}}$$

Although the relation of f_{nhb} to the structural equilibrium is obvious and direct (f_{nhb} is smaller when the peptide or protein is more structured) amide hydrogen exchange data are usually reported in terms of protection factors, which are inversely related to f_{nhb}:

$$PF = \frac{k_{int}}{k_{obs}} = \frac{1}{f_{nhb}}$$

In terms of the structural equilibrium, then

$$PF = \frac{1}{f_{nhb}} = \frac{[\text{unstructured} + \text{structured states}]}{[\text{unstructured states}]}$$

Protection factors increase as the population of the structured, hydrogen bonded state increases.

The application of the amide hydrogen exchange technique to the quantitative study of peptide helicity is straightforward.[2,3,4,5,42,43] Since hydrogen bonding occurs during helix formation, amides in helices are protected from exchange, and the fractional helicity of a

---

[42] Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1992**, *31*, 1263.
[43] Rohl, C. A.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8435.

given residue is reflected in the protection factor of the subsequent residue's amide (recall that hydrogen bonding of the i^{th} amide requires that the $i\text{-}1^{th}$ set of ϕ,ψ torsions be in the helical range). That is, in most cases (see following paragraph for exceptions) the protection factor of residue i's amide is related to residue i-1's degree of helicity as follows:

$$PF_i = \frac{1}{f_{nhb,i}} = \frac{1}{f_{rand,i-1}} = \frac{\text{all states}}{\text{states in which residue i - 1 is not helical}} = \frac{1}{1 - f_{hel,i-1}} \qquad \boxed{5\text{-}1}$$

where $f_{rand,i-1}$ is the fractional site random coil character of residue i-1 and $f_{hel,i-1}$ is the fractional site helicity of residue i-1. Hydrogen exchange data can potentially provide a tremendous amount of information about a peptide's helicity, as the protection factor for each individual amide measures the level of helicity at that site. Thus, one can determine not only the overall fractional helicity but also the distribution of helicity throughout a peptide by knowing the protection factors of each amide.

Before leaving this discussion of the effect of hydrogen bonding on amide hydrogen exchange rates, two issues need to be discussed. First, amides in helical peptides can be differentially protected from acid and base catalyzed exchange.[3] Recall that the NH groups of amides at the N-terminus of a helix are not hydrogen bonded, although their CO groups are. These amides are protected from acid catalyzed exchange, since the mechanism for acid catalyzed exchange requires that both the amide CO and NH be accessible to solvent, but they are not protected from base catalyzed exchange. The protection data of the acid and base catalyzed exchange reactions of most helical peptides have to be considered separately in order to account for this complication. Fortunately, though, in systems in which helicity is induced from the N-terminus (as in SO_3Hel_1-

peptide conjugates), situations in which an amide's carbonyl is hydrogen bonded while its NH is not are extremely unlikely. Amides in such systems can therefore be safely assumed to be equally protected from both types of exchange, and the acid and base catalyzed exchange reactions can be considered together. Second, in order for PF to be equal to $1/f_{nhb}$, ξ (the fraction of k_{int} at which hydrogen bonded amides exchange) has to be nearly 0. If it is not, then

$$PF = \frac{k_{int}}{k_{obs}} = \frac{k_{int}}{f_{nhb} \times k_{int} + f_{hb} \times \xi \times k_{int}} \neq \frac{1}{f_{nhb}} \text{ or } \frac{1}{f_{rand}}$$

There is evidence that ξ is very small for amides in protein helices, as such amides, whether on the protein surface or in the protein interior, have been shown to exchange extremely slowly (PFs $\geq$ 1000) in several systems.[44,45,46] Studies of small molecules have shown only a thirty-fold rate decrease in the exchange rate for an intramolecularly hydrogen bonded primary amide indicating a somewhat higher, though still small, value for ξ of ~ 0.03.[47] Both of the preceding statements justify the assumption that ξ is small, and therefore that PF = 1 / f_{nhb}. One should bear in mind, however, that small peptide helices are fluctuating structures, much less solid than protein helices, where the helix is enforced by the rest of the protein's structure, or rigid small molecules. Small peptide helices might be more susceptible to breathing motions whose amplitude is large enough to allow catalytic species to invade the i to i+4 CO – NH hydrogen bond and initiate exchange reactions. Such behavior, if observed, should be most pronounced in residues

---

[44] Kuwajima, K.; Baldwin, R. L. *J. Mol. Biol.* **1983**, *169*, 299.
[45] Wemmer, D.; Kallenbach, N. R. *Biochemistry* **1983**, *22*, 1901.
[46] Wand, A. J.; Roder, H.; Englander, S. W. *Biochemistry* **1986**, *25*, 1107.
[47] Perrin, C. L.; Dwyer, T. J.; Rebek, Jr. J.; Duff, R. J. *J. Am. Chem. Soc.* **1990**, *112*, 3122.

with small side chains (such as Gly, Ala, and Ser) and would lead to helicities calculated from exchange data being underestimated.

5.3.2 Protection Factors in SO_3Hel_1-peptide Conjugates According to Zimm-Bragg Model

The protection factor for an amide at a given site in a helical peptide is a function of the fractional helicity at that site, which is in turn a function of the helix initiation and propagation parameters as defined in the Zimm-Bragg model described in chapter 2. As discussed there, these functions can be cumbersome but are greatly simplified for N-terminally templated peptides, such as are used in this study. For an n residue SO_3Hel_1-peptide conjugate, the protection factor for residue i's amide in terms of the template initiation constant (B_{SO3}) and the helix propensities of the constituent residues (s values) is

$$PF_i = \frac{1}{f_{nhb,i}} = \frac{1}{f_{rand,i-1}} = \frac{1}{\dfrac{\text{sum of eq. constants for states in which residue i - 1 is not helical}}{\text{sum of all equilibrium constants}}}$$

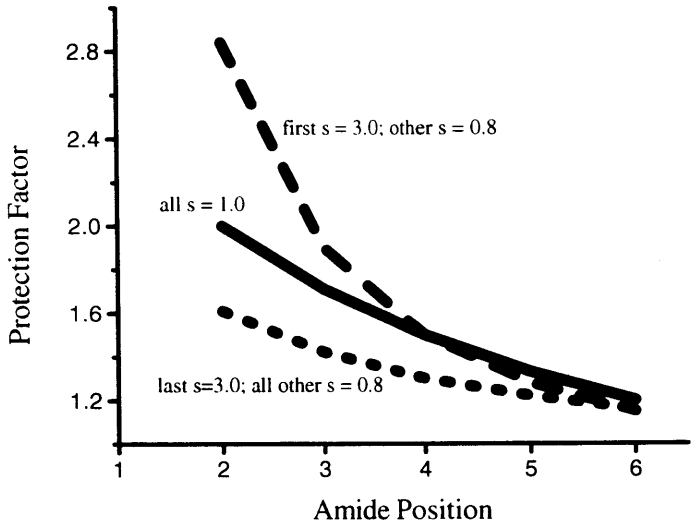
$$= \frac{\text{sum of all equilibrium constants}}{\text{sum of eq. constants for states in which amide i is not h - bonded}}$$

$$= \frac{1 + B_{SO3} + B_{SO3}s_1 + B_{SO3}s_1s_2 + \cdots + B_{SO3}s_1s_2\cdots s_n}{1 + B_{SO3} + B_{SO3}s_1 + \cdots + B_{SO3}s_1s_2\cdots s_{i-2}} \qquad \boxed{5\text{-}2}$$

The sum of equilibrium constants in the numerator is the one developed for SO_3Hel_1-peptide conjugates in section 4.3.4b (equation 4-2). The sum of equilibrium constants in

the denominator includes only those terms from the numerator that correspond to states in which the amide of residue i is not hydrogen bonded. Note that in stating that $PF_i = 1 / f_{rand,i-1}$ it has been implicitly assumed that $\xi \rightarrow 0$.

Figure 6. Plots of expected protection factors vs. amide position in a six residue SO_3Hel_1-peptide conjugate. Three cases are shown: a homopeptide case, in which all s = 1 (solid line), a case in which the first s value is 3, and all other s values are 0.8 (upper dashed line), and a case in which the last s value is 3 and all other s values are 0.8 (lower dashed line). In all cases B_{SO3} is set to 0.2, the value found in chapter 4.



Given the expression in equation 5-2, and given that $B_{SO3} = 0.2$ as determined in chapter 4, what can be expected from the protection factors for SO_3Hel_1-peptide conjugates? The protection factors that would be observed at positions two through six for three cases are plotted in figure 6 (the first position is not considered because k_{int} values for these amides could not be determined from the $HHel_1$-peptide conjugates). The solid line shows the homopeptide case, where all of the s values are 1. The protection factors decrease regularly from 2.0 at position 2 to 1.2 at position 6. The upper dashed line shows the case where the first s value is very high (s = 3.0) and all of the other s values are low (s = 0.8). Here, the protection factor at position 2 is large but it

decreases strongly with amide position, so that by position 4 the protection factor is about the same in this case as for the homopeptide case. The protection factor at position 6 is 1.15, slightly less than that in the homopeptide case. The lower dashed line shows the case where the last s value is very high (s = 3.0) and all of the other s values are low (s = 0.8). The protection factor at position 2 in this case is smaller than the corresponding protection factors in the other two cases, but it decreases less as a function of amide position than in the other two cases. The ratio of the first to the last protection factor in this case is 1.4, compared to 1.7 in the homopeptide case and 2.5 for the case with the large first protection factor. Three features of figure 6 should noted before the measured protection factors from the SO_3Hel_1-peptide conjugates are examined. First, in all cases the protection factors decrease monotonically from the N- to the C-terminus, as they must for N-terminally initiated helices. Second, the protection factor for the last amide is almost the same in all three cases. The cases are most easily distinguished by the behavior of their protection factors near the N-terminus. Third, the largest protection factor calculated in these cases is only 2.8. This indicates that protection factors in SO_3Hel_1-peptide conjugates are in the range where their interpretation is very sensitive to errors in k_{int}.

5.3.3 Measured Exchange Rate Constants in SO_3Hel_1-peptide Conjugates

The exchange rate constants of the amides in the eight SO_3Hel_1-peptide conjugates in this study (k_{obs}) were measured by the same method and under the same conditions as those in the corresponding $HHel_1$-peptide conjugates. The exchange reaction of the C-

terminal amide was still far too fast in these peptides to be measured, but the exchange reaction of the amide at the template-peptide junction was easily measurable. Also, the peaks in the amide regions of NMR spectra of SO_3Hel_1-peptide conjugates were much better dispersed than in the NMR spectra of the $HHel_1$-peptide conjugates, so more exchange rate constants could be determined by fitting to single exponential decays. In all, the k_{obs} values for 48 amides in the eight SO_3Hel_1-peptide conjugates (six from each peptide) were successfully measured either in triplicate or duplicate, 38 by single exponential decay and 10 by double exponential decay. The average values of k_{obs} and ln k_{obs}, and the standard errors in ln k_{obs} are listed in table 3. The data for all of the individual exchange reactions are recorded in table 16 of appendix 1.

Table 3. Amide exchange rate constants for SO_3Hel_1-peptide conjugates at pD 3.0 ± 0.15 and 5 °C. In each cell of the table is given the type of residue to which the amide NH belongs, the exchange rate constant k_{obs}, and in parenthesis ln k_{obs} and the standard error in ln k_{obs} (either ± 0.05 or ± 0.04 depending on whether two or three replicate measurements were made).

SO_3Hel_1-Peptide	Exchange Rate Constants by Amide Position					
	1st	2nd	3rd	4th	5th	6th
A†AQSFR	Ala 0.0129 min^{-1} (-4.35±0.05)	Ala 0.0113 min^{-1} (-4.48±0.05)	Gln 0.0126 min^{-1} (-4.37±0.05)	Ser 0.0182 min^{-1} (-4.01±0.05)	Phe 0.0119 min^{-1} (-4.43±0.05)	Arg 0.0147 min^{-1} (-4.22±0.05)
AASLQR	Ala 0.0128 min^{-1} (-4.36±0.04)	Ala 0.0086 min^{-1} (-4.75±0.04)	Ser 0.0123 min^{-1} (-4.40±0.04)	Leu 0.0041 min^{-1} (-5.51±0.04)	Gln 0.0105 min^{-1} (-4.56±0.04)	Arg 0.0104 min^{-1} (-4.57±0.04)
AL†AQSR	Ala 0.0070 min^{-1} (-4.95±0.05)	Leu 0.0013 min^{-1} (-6.68±0.05)	Ala 0.0096 min^{-1} (-4.64±0.05)	Gln 0.0088 min^{-1} (-4.73±0.05)	Ser 0.0127 min^{-1} (-4.36±0.05)	Arg 0.0112 min^{-1} (-4.49±0.05)
ALSAFR	Ala 0.0088 min^{-1} (-4.74±0.04)	Leu 0.0018 min^{-1} (-6.33±0.04)	Ser 0.0077 min^{-1} (-4.86±0.04)	Ala 0.0107 min^{-1} (-4.54±0.04)	Phe 0.0089 min^{-1} (-4.72±0.04)	Arg 0.0108 min^{-1} (-4.53±0.04)

AQFASR	Ala	Gln	Phe	Ala	Ser	Arg
	0.0085 min^{-1}	0.0037 min^{-1}	0.0055 min^{-1}	0.0093 min^{-1}	0.0216 min^{-1}	0.0133 min^{-1}
	(-4.77±0.05)	(-5.59±0.05)	(-5.18±0.05)	(-4.68±0.05)	(-3.83±0.05)	(-4.32±0.05)
AQSFLR	Ala	Gln	Ser	Phe	Leu	Arg
	0.0106 min^{-1}	0.0049 min^{-1}	0.0113 min^{-1}	0.0069 min^{-1}	0.0042 min^{-1}	0.0089 min^{-1}
	(-4.55±0.04)	(-5.31±0.04)	(-4.48±0.04)	(-4.98±0.04)	(-5.48±0.04)	(-4.73±0.04)
ASFL†AR	Ala	Ser	Phe	Leu	Ala	Arg
	0.0106 min^{-1}	0.0066 min^{-1}	0.0034 min^{-1}	0.0026 min^{-1}	0.0126 min^{-1}	0.0094 min^{-1}
	(-4.55±0.05)	(-5.01±0.05)	(-5.52±0.05)	(-5.95±0.05)	(-4.38±0.05)	(-4.67±0.05)
ASLQFR	Ala	Ser	Leu	Gln	Phe	Arg
	0.0080 min^{-1}	0.0083 min^{-1}	0.0039 min^{-1}	0.0084 min^{-1}	0.0077 min^{-1}	0.0114 min^{-1}
	(-4.83±0.04)	(-4.79±0.04)	(-5.50±0.04)	(-4.78±0.04)	(-4.86±0.04)	(-4.47±0.04)

The errors in ln k_{obs} (0.04 and 0.05) reported in this table correspond to roughly 4% and 5% error in k_{obs}. The k_{obs} values in table 3 are uniformly less than the k_{int} values in table 1, consistent with the presence of helicity throughout the SO_3Hel_1-peptide conjugates.

5.3.4 Measured Protection Factors in SO_3Hel_1-peptide Conjugates

The protection factors for all of the amides in the SO_3Hel_1-peptide conjugates for which accurate k_{int} values could be measured in the $HHel_1$-peptide conjugates can be calculated using the k_{int} and k_{obs} reported in tables 1 and 3 respectively. This yields 37 protection factors, all of which are listed in table 4 (followed by their natural logarithms

and the standard error in the natural logarithms in parenthesis.[48] The average protection factors at each position are listed at the bottom of the table.

Table 4. Protection factors for SO_3Hel_1-peptide conjugates at pD 3.0 ± 0.15 and 5 °C. In each cell of the table is given the type of residue to which the amide NH belongs, the protection factor (PF), and in parenthesis, ln PF and the standard error in ln PF (which is the square root of the sum of the squared standard errors of ln k_{int} and ln k_{obs}).

SO_3Hel_1-Peptide	Protection Factors by Amide Position					
	1st	2nd	3rd	4th	5th	6th
A†AQSFR	Ala	Ala	Gln	Ser	Phe	Arg
	-	1.92	1.35	1.18	1.17	1.12
		(0.65±0.09)	(0.30±0.09)	(0.17±0.09)	(0.16±0.09)	(0.12±0.09)
AASLQR	Ala	Ala	Ser	Leu	Gln	Arg
	-	2.27	1.47	1.57	1.12	1.19
		(0.82±0.08)	(0.38±0.08)	(0.45±0.08)	(0.11±0.08)	(0.17±0.08)
AL†AQSR	Ala	Leu	Ala	Gln	Ser	Arg
	-	5.78	1.63	1.61	1.77	1.84
		(1.76±0.10)	(0.49±0.10)	(0.48±0.10)	(0.57±0.10)	(0.61±0.10)
ALSAFR	Ala	Leu	Ser	Ala	Phe	Arg
	-	3.45	1.64	1.16	1.12	1.30
		(1.24±0.08)	(0.49±0.08)	(0.15±0.08)	(0.12±0.08)	(0.27±0.08)
AQFASR	Ala	Gln	Phe	Ala	Ser	Arg
	-	5.03	1.72	1.10	0.93	1.24
		(1.62±0.09)	(0.54±0.09)	(0.09±0.09)	(-0.07±0.09)	(0.21±0.09)
AQSFLR	Ala	Gln	Ser	Phe	Leu	Arg
	-	3.14	1.40	1.43	1.20	1.31
		(1.14±0.10)	(0.34±0.10)	(0.36±0.10)	(0.18±0.10)	(0.27±0.10)

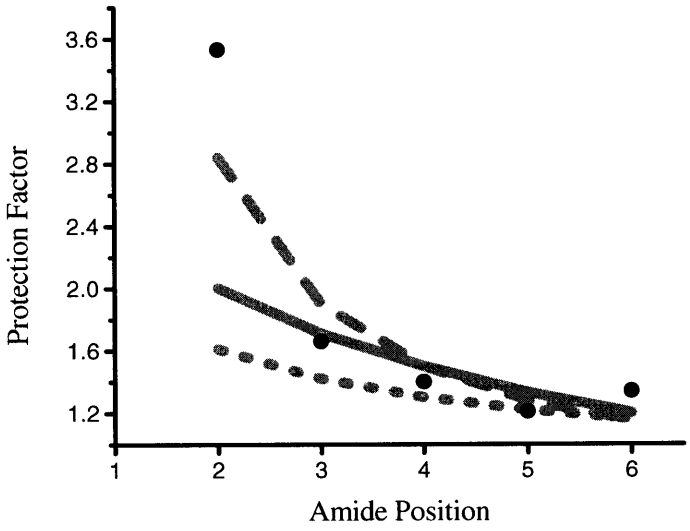
[48] The protection factors are quotients of rate constants. Since the errors in the rate constants are proportional the magnitude of the rate constants, the same must be true of the protection factors. The natural logarithms are given since the error in these ought to be normally distributed.

ASFL AR	Ala	Ser	Phe	Leu	Ala	Arg
	-	3.10	2.47	1.72	1.23	1.41
		(1.13±0.09)	(0.90±0.09)	(0.55±0.09)	(0.21±0.09)	(0.35±0.09)
ASLQFR	Ala	Ser	Leu	Gln	Phe	Arg
	-	-	1.58	-	1.14	-
			(0.46±0.08)		(0.13±0.08)	
Average	-	3.53	1.66	1.40	1.21	1.34

The average protection factors decrease consistently for the second through the fifth amides, as they should if SO_3Hel_1 induced frayed helices in attached peptides. A modest increase in the average protection factor is seen at the sixth amide. This amide might have been expected to behave differently, as it is always associated with arginine, the only charged residue in the peptide, but why this should occur will require further study. If one excludes the protection factors for the last residue, then in 18 of 21 consecutive pairs of protection factors, the more C-terminal one is less than the more N-terminal one. Considering that there is about 10% error in the reported protection factors (the errors in ln PF are around ±0.1), this is in excellent agreement with the expectation that the protection factors should decrease monotonically from the N- to the C-terminus.

The average protection factors are plotted against amide position in figure 7. Also shown in the figure for comparison are lines (in gray) from the three cases of figure 6 (all s = 1; s = 3 for the first amide while all other s = 0.8; and s = 3 for the last amide while all other s = 0.8).

Figure 7. Plot of average protection factors from SO_3Hel_1-peptide conjugates vs. amide position (black circles). The gray lines represent what would be expected if all s = 1 (solid line), if the first s = 3 and all other s = 0.8 (upper dashed line), and if the last s = 3.0 and all other s = 0.8 (lower dashed line)



The plot in figure 7 shows that there is a large drop in the average protection factor between the second and third amides, much like that observed in the case where the first s = 3 and the rest of the s = 0.8 (the upper dashed line). The large average protection factor at the second position is due mostly to the very large protection factors observed in the SO_3Hel_1-peptide conjugates in which the second residue is either Leu or Gln. Whether these residues cause the first s value to be particularly large when they occupy this position, or whether the behavior can be explained in some other way, will be discussed in detail in section 5.4.

5.4 Non-Canonical Interactions and Peptide Helicity

This section presents the results of a preliminary use of the amide hydrogen exchange technique in SO_3Hel_1-peptide conjugates to study the way in which context dependent effects influence peptide helicity. In particular, the importance of non-canonical interactions, that is, interactions that have not been widely studied in the context of peptide helicity, such as interactions between side-chains and the helix barrel or i to i+1 or i to i+2 interactions between side-chains, is addressed as follows. The protection factor data for the eight SO_3Hel_1-peptide conjugates from the last section will be fit to the Zimm-Bragg model for N-terminally initiated peptide helicity as presented in section 2.4. The model will not include any corrections to account for interactions of any kind. There should be two sources of error in the fit: measurement error from the experimental determination of the protection factors, and model error from the model's inability to account for context dependent effects that might influence peptide helicity. If it can be shown that the total error in the fit exceeds what could have been expected from the known measurement error in the protection factors (recall that the error in ln PF is about ± 0.1), and if it can be shown that canonical interactions are not the source of the model error, then it must be that non-canonical interactions occur in, and significantly affect the helicity of, the SO_3Hel_1-peptide conjugates. Before engaging in this analysis of the protection factor data, however, the helicities of the eight peptides is examined in a more familiar setting: using the t/c ratios of their $AcHel_1$-peptide conjugates.

5.4.1 The Use of t/c Ratios of $AcHel_1$-peptide Conjugates to Aid the Interpretation of Protection Factor Data from SO_3Hel_1-peptide Conjugates

The eight $*AcHel_1$-peptide conjugates (where $*AcHel_1$ has the acetyl methyl and carbonyl carbons labeled with ^{13}C) corresponding to the SO_3Hel_1-peptide conjugates used in this study were prepared and their t/c ratios measured.[49] What can these t/c ratios reveal about the presence of interactions that affect helix formation in these peptides? Recall the formula for the t/c ratio that was developed in chapter 2:

$$\left(\frac{t}{c}\right) = A + B + B(s_1 + s_1s_2 + s_1s_2s_3 + \cdots) \qquad \boxed{5\text{-}3}$$

where A and B are constants characteristic of the $AcHel_1$ template (0.832 and 0.156 respectively), s_1 is the s value for the type of residue that occupies the first position in the peptide of interest, s_2 is the s value for the type of residue that occupies the second position, etc. For the $*AcHel_1$-peptide conjugates, t/c ratios can be calculated using equation 5-3 and the s values from table 2 of chapter 2. These s values were measured in systems designed to be interaction free. One would therefore expect the $*AcHel_1$-peptide conjugates' measured t/c ratios to match the calculated t/c ratios if the $*AcHel_1$-peptide conjugates were also interaction free. The t/c ratios predicted with each set of s values from table 2 of chapter 2 can be compared to the experimentally measured t/c ratio in table 5.

---

[49] The labeled $*AcHel_1$ template was used so that, so that t/c ratios could be measured from the ^{13}C-NMR spectra in case the ^1H-NMR spectra of these template-peptide conjugates were too crowded to measure accurate t/c ratios. This turned out to be unnecessary, as the ^1H-NMR spectra of the $AcHel_1$-peptide conjugates were well dispersed. However, it should be noted that in all cases the t/c ratios measured in the ^1H- and ^{13}C-NMR spectra were very similar.

Table 5. Comparison of experimentally measured t/c ratios with predictions using the various sets of s values in table 2 of chapter 2. All experimental t/c ratios were measured from the ^{1}H-NMR spectra obtained at 5 °C, pD 3.0 ± 0.15 (t/c ratios were measured in the same buffer as used in the exchange experiments).[50]

*AcHel$_1$-Peptide	Measured t/c ratio	Calculated t/c Ratios using s values from			
		Scheraga	Baldwin	Stellwagen	Kallenbach
AAQSFR	1.77 (± 0.13)	1.94	2.07	2.36	2.31
AQFASR	1.76 (± 0.13)	1.98	1.56	2.00	1.59
AASLQR	2.68 (± 0.16)	1.93	2.14	2.34	2.65
ALAQSR	2.19 (± 0.14)	2.06	2.20	2.65	2.66
ALSAFR	2.10 (± 0.06)	1.99	1.76	2.15	2.14
ASFLAR	1.83 (± 0.16)	1.92	1.47	1.84	1.64
AQSFLR	1.60 (± 0.07)	1.89	1.50	1.63	1.54
ASLQFR	1.53 (± 0.09)	1.88	1.51	1.60	1.67
RMSD	-	0.30	0.37	0.32	0.43

The root mean square deviation (RMSD) of the calculated t/c ratios from the experimentally measured t/c ratios are 2.5 to 3.5 times more than the average experimental error in the measured t/c ratios (~ 0.12). This is a preliminary indication that the helix propensities of at least some of the residues in these AcHel$_1$-peptide conjugates might be affected by interactions.

A more rigorous method for determining whether the t/c ratio data show the influence of context dependent helix propensities is the recently introduced permutation test.[51] The

[50] The t/c ratios found from the ^{13}C-NMR spectra were in good agreement with those in table 6 (these are, in the same order as in the table: 1.76, 1.91, 2.84, 2.57, 2.11, 1.75, 1.67, 1.51).
[51] Zerkowski, J. A.; Powers, E. T.; Kemp, D. S. *J. Am. Chem. Soc.* **1997**, *119*, 1153.

permutation test is founded on the observation that for two peptides whose sequences are related by a permutation, the difference in their helicities is dictated by the difference between the largest and smallest s values. Take for example two pairs of $AcHel_1$-peptide conjugates where the sequences of each pair's members differ only by a permutation. In the first pair, neither member's helicity is affected by any context dependent interactions while in the second pair, the first member houses an interaction that promotes helicity whereas the second member does not. The difference in the t/c ratios (Δt/c) of the first pair's $AcHel_1$-peptide conjugates should not be any larger than one would predict using the scales of intrinsic s values in table 2 of chapter 2. The difference in the t/c ratios of the second pair's $AcHel_1$-peptide conjugates, though, should be larger than one would predict using these s values, and this would serve as an observable indication of the helix stabilizing interaction.

The eight peptides studied in this work can be organized into four pairs whose sequences are related by permutation: $A^{\dagger}AQSFR$ and AQFASR, AASLQR and $AL^{\dagger}AQSR$, ALSAFR and $ASFL^{\dagger}AR$, and AQSFLR and ASLQFR. These data are arranged for the permutation test in table 6.

Table 6. Difference in experimental t/c ratios (Δt/c) for permutation-related *$AcHel_1$-peptide conjugates compared to the difference in the corresponding calculated t/c ratios.

*$AcHel_1$-Peptide Pairs	Δt/c ratios from				
	Experimental	Scheraga	Baldwin	Stellwagen	Kallenbach
AAQSFR-AQFASR	-0.33	-0.04	0.51	0.36	0.72
AASLQR-ALAQSR	-0.92	-0.13	-0.06	-0.31	-0.01
ALSAFR-ASFLAR	0.59	0.07	0.29	0.31	0.50
AQSFLR-ASLQFR	0.30	0.01	-0.01	0.03	-0.13

The Δt/c ratios calculated from Scheraga's s values are not accurate for any of the pairs. While the sign of the difference is always correct, the magnitude is always far too small, consistent with context dependent interactions affecting helix formation in at least one member of each pair. The Δt/c ratios calculated from all of the other sets of s values share the following pattern: Δt/c is large enough but has the wrong sign for the first pair, Δt/c has the right sign but is much too small for the second pair, Δt/c has the right sign but is too small for the third pair, and Δt/c tends to be close to 0 and therefore much too small for the last pair. This pattern is also consistent with helix stabilizing or destabilizing interactions being present in the *$AcHel_1$-peptide conjugates. With this independent information in hand, the analysis of the protection factor data from the SO_3Hel_1 conjugates can be addressed.

5.4.2 The Analysis of Protection Factor data from SO_3Hel_1-peptide conjugates

Recall that the protection factor for any amide in an n residue SO_3Hel_1-peptide conjugate is given in terms of SO_3Hel_1's initiation constant, B_{SO3}, and the s values of the residues in the attached peptide, s_1, s_2, s_3, etc, by equation 5-2. For convenience, that equation is reproduced here:

$$PF_i = \frac{1 + B_{SO3} + B_{SO3}s_1 + B_{SO3}s_1s_2 + \cdots + B_{SO3}s_1s_2\cdots s_n}{1 + B_{SO3} + B_{SO3}s_1 + \cdots + B_{SO3}s_1s_2\cdots s_{i-2}}$$

An equation such as the one above can be written for each of the protection factors listed in table 4 for the eight SO_3Hel_1-peptide conjugates. Thus, each protection factor is a function of B_{SO3} and the s values of the residues in the peptide. If the protection factors could be measured with perfect accuracy, and if helix formation in the SO_3Hel_1-peptide conjugates were not affected by interactions, these equations could be solved exactly to provide B_{SO3} and a set of s values. However, experimental error in the measurement of protection factors (henceforth measurement error), and error due to inadequacies of the model (henceforth model error) inevitably intrude to degrade the fit of the model to the measured protection factors. Under these circumstances, rather than searching for the set of parameters that provides an exact solution to the equations, one searches for the set of parameters that minimizes the differences between the measured and calculated protection factors. The means by which the data are fit, the relationship between the error in the fit and the influence of non-canonical interactions on helix formation in the SO_3Hel_1-peptide conjugates, and a discussion of the results from some simulations of the data fitting process are presented in the three following sections.

5.4.2a Determination of the Optimal set of Zimm-Bragg Parameters

The template initiation constant, B_{SO3}, and the helix propensities of the six types of residues in the peptides described above can be determined by finding the set of these parameters that optimizes the fit of the calculated to the measured protection factors. Practically, this is achieved using a least squares approach; that is, by minimizing the sum of the squares of the differences (the SSQ) between the calculated and the experimental protection factors:

$$SSQ_{PF}([B_{SO3}, s_A, s_F, s_L, s_Q, s_R, s_S]_{optimal}) =$$

$$\text{Minimum}\left[\sum_{\text{all PFs}} (\text{experimental PF - calculated PF})^2\right]$$

The function SSQ_{PF} consists of a sum of terms of the form (experimental PF – calculated PF)^2, one for each measured protection factor.

The above is the most obvious form for the least squares process to take, but it is not the best for three reasons. First, it requires that the errors in the protection factors be independent of their magnitudes. The protection factors are quotients of rate constants, the errors of which are dependent on their magnitudes. A logarithmic transformation is accordingly appropriate to stabilize the variance. Second, the parameters that are varied during the minimization are not bounded, and it could happen that one of them might be assigned a negative value. This is physically nonsensical; the parameters are all equilibrium constants, and must therefore be positive. It can be remedied by replacing

the s values in the SSQ with $e^{\ln s}$, which is equivalent to s except in that it cannot be negative. The values of ln s can then be treated as the parameters that are varied to minimize SSQ. Third, the fitting process can be simplified by assigning to some of the parameters values determined from previous work. Therefore, B_{SO3} is set to 0.2, the value found for it in section 4.3.4. After applying the above remedies, the revised least squares approach can be expressed as follows:

$$SSQ_{lnPF}([B_{SO3} = 0.2, \ln s_A, \ln s_F, \ln s_L, \ln s_Q, \ln s_R, \ln s_S]_{optimal}) = \quad \boxed{5-4}$$

$$\text{Minimum}\left[\sum_{\text{all PFs}} (\ln[\text{experimental PF}] - \ln[\text{calculated PF}])^2\right]$$

Any optimization algorithm can be used to find the minimum of SSQ_{lnPF} with respect to its parameters. In this work, a modification of Powell's method as implemented by Mathematica 3.0 (Wolfram Research Inc, 1996) is used, and the optimization is always started with all the ln s values set to 0. The Mathematica program used for all data fitting is reproduced in appendix 2. As with all multiparameter optimization problems, estimation of the errors in the parameters can be complicated. This issue is addressed by taking advantage of a relationship that should exist between the SSQ_{lnPF} and a χ^2 statistic:

$$\chi^2_{\ln PF} = \sum_{\text{all PFs}} \frac{(\ln[\text{experimental PF}] - \ln[\text{calculated PF}])^2}{sd^2_{lnPF}} = \frac{SSQ_{lnPF}}{sd^2_{lnPF}}$$

where sd^2_{lnPF} is the common standard deviation for all of the protection factors. The minimum of χ^2_{lnPF} (and therefore of SSQ_{lnPF}) should be a point in the space defined by the model parameters that is surrounded by a series of ellipsoids defined by the set of equations $\chi^2_{lnPF} = SSQ_{lnPF}/sd^2_{lnPF}$ = constant, where the constant is greater than the

minimum value of χ^2_{lnPF}. Each ellipsoid encloses a joint confidence region for all the parameters with a confidence coefficient that depends on the difference between the constant and the minimum value of χ^2_{lnPF} and the dimensionality of the parameter space.[52] This procedure requires only that the residuals of the fit, that is the values of ln [experimental PF] – ln [calculated PF], be normally distributed and independent of the experimental protection factors' magnitude, and the value of SSQ_{lnPF} must be the global minimum (as opposed to a merely local minimum).

5.4.2b Relationship between the Error in the Fit and the Presence of Non-Canonical Interactions in the SO_3Hel_1-peptide Conjugates

The quality of the calculated data's fit to the experimental data with the optimal parameter set, as measured by the value of SSQ_{lnPF} at its minimum, reflects the composite measurement and model error. It was noted above that SSQ_{lnPF} is related to a χ^2 variable, in this case with 31 degrees of freedom (one for each measured protection factor used in the data fitting (37), less one for each parameter determined(6)). A χ^2 variable with 31 degrees of freedom is expected to be around 31 if only known measurement errors contribute to it.[53] Since the measurement error, sd_{lnPF}, according to the data in table 4 should be at most 0.1 in ln PF, and since $SSQ_{lnPF} = \chi^2_{lnPF} \times sd^2_{lnPF}$, this means that SSQ_{lnPF} should be around 0.31. If SSQ_{lnPF} is much larger than 0.31, then there is reason to suspect that model error is contributing to it in addition to measurement error. For

[52] Press, W. H.; Flannery, B. P.; Teukolvsky, S. A. Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing* pp. 529-538; Cambridge University Press: Cambridge, New York, Melbourne, **1986**.

example, the probability that SSQ_{lnPF} could be inflated to 0.70 (or $\chi^2_{lnPF} > 70$) just by random fluctuations in the data due to measurement error is less than 10^{-4}. If a minimum value of SSQ_{lnPF} as large as 0.70 were observed, and if the model error could not be explained in terms of canonical interactions, the one could confidently conclude that model error due to non-canonical interactions was the source of the excess in SSQ_{lnPF}.

5.4.2c Response of the Results from the Fitting of Exchange Rates to Measurement and Model Error Ascertained Using Simulated Data Sets

Before the real protection factors from the SO_3Hel_1-peptide conjugates are analyzed, the effects of measurement and model error on the results of the analysis will be evaluated using the results from fits to simulated sets of protection factors. The full details of the analysis are available in appendix 3; this section just summarizes the results.

In the first simulation, protection factors for the 2nd through the 6th amides of the eight SO_3Hel_1-peptide conjugates (where peptide = A†AQSFR, AASLQR, AL†AQSR, ALSAFR, AQFASR, AQSFLR, ASFL†AR, and ASLQFR) were calculated using equation 5-2, with $B_{SO3} = 0.2$, and the s values determined by Scheraga and co-workers: $s_A = 1.07$, $s_F = 1.09$, $s_L = 1.14$, $s_Q = 0.98$, $s_R = 1.03$, $s_S = 0.76$. This yielded a set of 40 protection factors, to which the fitting procedure outlined in section 5.4.2a above was applied. The minimum found for SSQ_{lnPF} was 7.6×10^{-6}, and the parameters at the minimum were all within 2% of the input parameters. This simulation served merely as a

[53] More exactly, a χ^2 variable with 31 deg. of freedom has a 98% chance of falling between 15.6 and 52.2

positive control- it showed that the minimization method returned the correct parameters in the absence of experimental or model error.

In the second simulation, the protection factors for the eight SO_3Hel_1-peptide conjugates were calculated as above, but before the data were fit about 10% random error was added to each protection factor.[54] The data were then fit to the protection factors with the synthesized error. This process, from the generation of error laden protection factors to the fitting of the data, was repeated ten times, to yield ten different sets of minimum SSQ_{lnPF}s and optimal s values. The median of the values of SSQ_{lnPF} found in the simulations was 0.36, about what one would expect given that SSQ_{lnPF} / sd^2_{lnPF} is a χ^2 variable, and the high and low values found for SSQ_{lnPF} were 0.44 and 0.22. The medians of the s values found at the SSQ_{lnPF} minima were all about equal to the input s values. However, there was a striking feature of the data: the s values found in individual cases occasionally could vary a great deal from their input values. For example, the input value for s_L was 1.14. The median value found for s_L was 1.08, but the high and low values found were 1.70 and 0.87. This was most pronounced for s_R, where the input value was 1.03, and the high and low values were 5.33 and 0.01.

In the third simulation, the effect of increasingly common context dependencies in s values was studied on top of the effect of measurement error. This simulation actually consisted of three groups of trials. In the first group, the protection factors were calculated using the s values as before, except that in two situations different s values were used (when Ser preceded Arg, the s_S was decreased to 0.51, and when Leu was the second residue, s_L was increased to 1.71). In the second group, there were two more

[54] A normally distributed random number with mean = 0 and standard deviation = 0.1 was added to the natural logarithm of the protection factor.

situations in which different s values were used in addition to the two above. In the third group, two more situations in which different s values were used were added to the preceding four. All of the groups of trials included 10% random error in the protection factors as well as the various context dependent s values. The median values found for the SSQ_{lnPF}s of the groups increased as the number of context dependent s values increased, as shown in table 7.

Table 7. Dependence of SSQ_{lnPF} found in minimizations of simulated sets of protection factors on the number of context dependent s values included in the calculation of said simulated protection factors. All simulations include 10% random error in the protection factors.

Trial	number of context dependent s values	median SSQ_{lnPF} (max, min)
(from preceding paragraph)	0	0.36 (0.22, 0.44)
1	2	0.38 (0.25, 0.61)
2	4	0.59 (0.44, 1.03)
3	6	0.85 (0.60, 1.24)

Three conclusions can be drawn from the simulations summarized above. First, that the method can return correct s values under ideal conditions. Second, when there is even 10% error, the values of the individual parameters at the SSQ_{lnPF} minimum may deviate significantly from their true value, although they ought to be close for the most part. Third, context dependence in s values increases the minimum value of SSQ_{lnPF} found from fitting a set of protection factors beyond what it would be from random error alone.

5.4.3 Fit of Experimental Protection Factors from SO_3Hel_1-peptide Conjugate

The set of parameters (s_A, s_F, s_L, s_Q, s_R, and s_S) were found that minimized SSQ_{lnPF} for the 37 experimentally determined protection factors of the SO_3Hel_1-peptide conjugates. These s values and the value of SSQ_{lnPF} at the minimum are listed in table 8.

Table 8. s values at SSQ_{lnPF} minimum for the 37 experimentally determined protection factors from table 4.

SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
2.82	1.35	1.83	2.27	0.92	0.09	0.37

Recall that the value of SSQ_{lnPF} expected on the basis of the measurement error is 0.31. The actual minimum value for SSQ_{lnPF} (2.82) is about 900% larger than this value. This indicates that the data are heavily influenced by non-measurement sources of error. Another indication that this is the case comes from using the s values from the table above to calculate the t/c ratios of the *$AcHel_1$-peptide conjugates from section 5.4.1. These calculated t/c ratios are consistently and substantially larger than the measured t/c ratios, as shown in table 9.

Table 9. Comparison of experimental and calculated t/c ratios for *$AcHel_1$-peptide conjugates using the s values found from the fitting of protection factors of SO_3Hel_1-peptide conjugates. Experimental t/c ratios are taken from table 5.

Peptide	t/c_{exp}	t/c_{calc}
*$AcHel_1A^{\dagger}AQSFR\text{-}NH_2$	1.77 (± 0.13)	2.03
*$AcHel_1AASLQR\text{-}NH_2$	1.76 (± 0.13)	2.06
*$AcHel_1AL^{\dagger}AQSR\text{-}NH_2$	2.68 (± 0.16)	3.15
*$AcHel_1ALSAFR\text{-}NH_2$	2.19 (± 0.14)	2.56
*$AcHel_1AQFASR\text{-}NH_2$	2.10 (± 0.06)	2.41
*$AcHel_1AQSFLR\text{-}NH_2$	1.83 (± 0.16)	1.91
*$AcHel_1ASFL^{\dagger}AR\text{-}NH_2$	1.60 (± 0.07)	2.21
*$AcHel_1ASLQFR\text{-}NH_2$	1.53 (± 0.09)	1.93
RMSD	-	**0.35**

The high value of SSQ_{lnPF} and the inability of the optimal set of s values to account for the t/c ratios of the *$AcHel_1$-peptide conjugates show that there must certainly be something other than measurement error that contributes to the total error in the fit. What could the additional source or sources of error be? Two rationalizations are offered below.

First, the excess error could be attributed to model error, the only other source of error considered thus far. Since only one canonical interaction (the potential i to i+3 Leu-Phe interaction in ALSAFR) can occur in the set of peptides under study, it is unlikely that the excess in SSQ_{lnPF} is due to canonical interactions. Non-canonical interactions therefore have to cause most of the model error. A potential non-canonical interaction

was already apparent from the examination of the average protection factors in section 5.3.4. There it was noted that the dependence of the average protection factors on amide position fit the pattern expected if the first s value were large while the subsequent s values were small. This was particularly striking in the protection factors of the SO_3Hel_1-peptide conjugates in which Leu or Gln were the second residues. Perhaps Leu and Gln are able to interact with the template, which has a structure much like the helix barrel, in such a way as to increase the s value of the residue preceding them. This Leu or Gln side-chain to main-chain interaction would qualify as a non-canonical interaction, and it could explain at least some of the excess error found at the SSQ_{lnPF} minimum.

The second, alternative rationalization is that perhaps the exchange rates can be affected by factors other than peptide helicity. This would introduce a source of error due to neither measurement nor the model. If this were the case, perhaps the fit of the model to the data could be improved by attempting to correct the protection factors. This avenue is explored in the following sections.

5.4.4 Fit of Model to Corrected Protection Factors from SO_3Hel_1-peptide Conjugates

Which protection factors might require correction and why? The comparison of exchange rate constants measured in $HHel_1$-peptide conjugates with those measured in the corresponding Ac-peptides revealed that the rate constants measured for the second amide could be affected by the nature of the group at the N-terminus. In fact, it has been found that the exchange rate constant of the second amide in SO_3ProProAQSFLR-NH_2 and SO_3ProProASFL†AR-NH_2, where the SO_3ProPro unit replaces $HHel_1$ or Ac at the N-

terminus, is smaller than the same position's exchange rate constant measured in the corresponding $HHel_l$ or Ac-peptides (see appendix 4). Since the $SO_3ProPro$ unit cannot induce helices the way SO_3Hel_l can (it is prevented from assuming the helix initiating conformation by steric repulsions between the two pyrrolidine rings), this suggests that something about the $SO_3ProPro$ unit, possibly the sulfamate's negative charge, decreases the k_{int} of the amides at the second position. Protection factors that are calculated for this position from the larger k_{int} values found in $HHel_l$-peptide conjugates would consequently be too large, and would overestimate the fractional site helicity. In addition, it was noted in section 5.3.1 that, since short peptide helices are not as rigid as protein helices, exchange might not be completely prevented in helical states for the residues with the smallest side-chains. Protection factors calculated for the alanine and serine residues in the SO_3Hel_l-peptide conjugates might therefore be too small. If either or both of these effects were significant, it would mean that a third type of error due to faulty assumptions about the relationship between the protection factors and fractional site helicity (henceforth assumption error) were contributing to SSQ_{lnPF}. Corrections for such errors are added to the model in this section. If the corrections applied to the protection factors result in the optimal SSQ_{lnPF} decreasing to the level expected from measurement error, it would indicate that there may be no model error- further work would be required to establish that the corrections were required for the reasons given. If the optimal SSQ_{lnPF} were still too large to be due solely to measurement error, it would indicate that there was definitely significant model error from non-canonical interactions.

5.4.4a Correction of Assumption Error in Protection Factors

The protection factors affected by assumption errors can be corrected by adjusting them with a constant multiplicative factor or, since a logarithmic transformation is required for the fitting process, by adjusting their natural logarithms with a constant additive term:

$$PF_{adj} = c \times PF_{exp}$$

or

$$\ln PF_{adj} = \ln c + \ln PF_{exp}$$

where PF_{adj} is the adjusted protection factor, PF_{exp} is the experimental protection factor and c is the constant. The adjusted protection factors can be substituted for the experimental protection factors in the SSQ_{lnPF} expression, so that terms with adjusted protection factors become:

$$(\ln [PF_{adj}] - \ln [PF_{calc}])^2 = (\ln [PF_{exp}] + \ln c - \ln [PF_{calc}])^2$$

The value of the adjustment term, ln c, can then be determined by treating it as a parameter in the SSQ_{lnPF} minimization procedure.

5.4.4b Determination of Which of the Corrections Are Significant

The application of corrections to the protection factors of three types of amides (the amides of the second residues, the amides of alanine residues, and the amides of serine residues) was advocated above. However, each correction applied amounts to another parameter in the model, and another step closer to the trivial situation in which the 37 experimentally measured protection factors are fit perfectly by a model with 37 parameters. In order to avoid an unregulated expansion in the number of model parameters, a compelling statistical justification must be required for the incorporation of each correction. In this case, a significant improvement in the fit of the model to the data, where the gauge for goodness-of-fit is the adjusted multiple correlation coefficient (R^2_{adj}, a correlation coefficient adjusted for the number of model parameters), because of applying a given correction constitutes the necessary statistical justification.

The need for each of the three types of corrections is accordingly assessed as follows. First, the model without any corrections is fitted to the data with B_{SO3} set to 0.2, and the R^2_{adj} is calculated. A correction is then added to the model, the model is refitted and a new R^2_{adj} is calculated. If the new R^2_{adj} is significantly larger than the old R^2_{adj}, the correction is kept. If not, the correction is discarded. The results for each correction are summarized in table 10.

Table 10. The effect of various corrections on the quality of the protection factor fit. "PF2" signifies the correction for the amides at the second position, "Ala" signifies the correction for the amides of alanine residues, and "Ser" signifies the correction for the amides of serine residues.

Corrections	number of parameters	SSQ_{lnPF}	R^2_{adj}	Result
none	6	2.82	0.48	-
PF2	7	1.52	0.71	PF2 correction kept
PF2 + Ala	8	1.06	0.79	Ala correction kept
PF2 + Ala + Ser	9	0.99	0.80	Ser correction discarded

The fit of the model to the data is vastly improved upon the application of a correction to the second amides' protection factors (R^2_{adj} increases from 0.48 to 0.71), and improved again upon application of the correction to the alanine amides' protection factors (R^2_{adj} = 0.79). In contrast, the correction for the serine amides' protection factors results in the barest increase in R^2_{adj} (to 0.80). The first two corrections are retained while the last is discarded, so that the final form for the model fitting is

$$SSQ_{lnPF}([B_{SO3} = 0.2, \ln s_A, \ln s_F, \ln s_L, \ln s_Q, \ln s_R, \ln s_S, \ln c_{PF2}, \ln c_{Ala}]_{optimal}) =$$

$$\text{Minimum}\left[\sum_{\text{all PFs}} (\ln[\text{experimental PF}] - \ln[\text{calculated PF}])^2\right] \qquad \boxed{5 - 5}$$

where $\ln c_{PF2}$ and $\ln c_{Ala}$ are the corrections applied to the second and alanine amides' protection factors such that

$$\ln PF2_{adj} = \ln PF2_{exp} + \ln c_{PF2}$$

and

$$\ln PF_{Ala\ amide,\ adj} = \ln PF_{Ala\ amide,\ exp} + \ln c_{Ala}$$

5.4.5 Fit of the Model to the Protection Factors of the SO_3Hel_1-peptide Conjugates with Corrections Incorporated

The set of s values (s_A, s_F, s_L, s_Q, s_R, and s_S) were found that minimized SSQ_{lnPF} for the 37 experimentally determined protection factors of the SO_3Hel_1-peptide conjugates with the correction factors c_{PF2} and c_{Ala} (eight parameters in all). The s values, the correction factors, and the value of SSQ_{lnPF} at the minimum are listed in table 11. Also listed are the upper and lower bounds of the 90% confidence region along each parameter axis determined as described in section 5.4.2a and ref. 52. (The construction of confidence regions in this manner is permitted because, as shown in appendix 5, the residuals of the fit are normal and independent of the magnitude of the protection factors, and the fit very likely represents a global minimum in SSQ_{lnPF}.)

Table 11. SSQ_{lnPF}, s values, and correction factors for the best fit of the model to the SO3Hel1-peptide protection factors. Upper and lower bounds of the 90% confidence region for the s values and correction factors are also reported.

Parameter	SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S	c_{PF2}	c_{Ala}
Best Fit	1.06	1.06	1.67	2.47	0.91	1.30	0.27	0.57	1.39
Upper Bound, 90% CR	-	1.47	4.50	4.78	1.47	197	1.14	0.68	1.65
Lower Bound, 90% CR	-	0.76	0.62	1.27	0.57	0.01	0.06	0.47	1.16

The minimum value of SSQ_{lnPF}, even with the corrections applied to the protection factors, is still too large for it to have arisen entirely from measurement error. With the corrections included as part of the model, SSQ_{lnPF} is related to a χ^2 variable with 29 degrees of freedom (one degree of freedom for each of 37 protection factors, less one for each of the s values and each of the correction factors). As such, SSQ_{lnPF} should have been around 0.29. The probability that random fluctuations due to measurement error could have made it as large as 1.06 is on the order of 10^{-10}.

According to table 11, the confidence region encompasses a broad range for all of the s values. This is expected, given the variability of the s values observed in the simulations of section 5.4.2c and appendix 3. Despite these s values being ill defined, the t/c ratios calculated using them for the eight *$AcHel_1$-peptide conjugates are quite accurate, as shown in table 12.

Table 12. Comparison of experimental and calculated t/c ratios for *$AcHel_1$-peptide conjugates using the s values found from the fitting of the corrected protection factors of SO_3Hel_1-peptide conjugates. All experimental t/c ratios were measured from the ^{1}H-NMR spectrum obtained at 5 °C, pD 3.0 ± 0.15.

Peptide	t/c_{exp}	t/c_{calc}
*$AcHel_1A^{\dagger}AQSFR\text{-}NH_2$	1.77 (± 0.13)	1.70
*$AcHel_1AASLQR\text{-}NH_2$	1.76 (± 0.13)	1.74
*$AcHel_1AL^{\dagger}AQSR\text{-}NH_2$	2.68 (± 0.16)	2.63
*$AcHel_1ALSAFR\text{-}NH_2$	2.19 (± 0.14)	2.24
*$AcHel_1AQFASR\text{-}NH_2$	2.10 (± 0.06)	1.99
*$AcHel_1AQSFLR\text{-}NH_2$	1.83 (± 0.16)	1.80

*$AcHel_1$ASFL†AR-NH_2	1.60 (± 0.07)	1.91
*$AcHel_1$ASLQFR-NH_2	1.53 (± 0.09)	1.79
RMSD	-	**0.15**

The agreement between the calculated and measured t/c ratios is excellent for all of the $AcHel_1$-peptide conjugates except for the two in which serine is the second residue. The disparity in these two cases is so outstanding that it fosters the impression that a serine in the second position might be able to interact with the SO_3Hel_1 template in a way that is not possible with the $AcHel_1$ template. This point will be taken up again in section 5.4.6.

The optimal set of s values found from the fit to the corrected protection factors is compared to previously reported s values in table 13. The s values determined in this section agree about as well with the literature values as the literature values agree with each other. It would be imprudent to make any assertions about the s values beyond this, given the breadth of the s values' confidence region. However, one point is worth a remark. The s value of alanine has been the subject of some contention.[2,55,56,57,58,59,60,61] This is the best determined of all of the s values in this work, with a best fit value of 1.06 and varying only from 0.76 to 1.47. This value of s_A compares favorably to the values determined from studies of $AcHel_1$-peptide conjugates[58] (1.02) and to the value found by

---

[55] Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci.* **1989**, *86*, 5286.
[56] Vila, J.; Williams, R. L.; Grant, J. A.; Wójcik, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci.* **1992**, *89*, 7821.
[57] Padmanabhan, S.; York, E. J.; Gera, L.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 8604
[58] Kemp, D. S.; Oslick, S. L.; Allen, T. J. *J. Am. Chem. Soc.* **1996**, *118*, 4249.
[59] Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci.* **1996**, *93*, 4025.
[60] Renold, P.; Tsang, K. Y.; Shimizu, L. S.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 12234.
[61] Yang, J.; Zhao, K.; Gong, Y.; Vologodskii, A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1998**, *120*, 10646.

Scheraga[62] (1.07) from studies of random peptide copolymers. It does not compare favorably to the values determined by Baldwin,[10] Stellwagen,[63,64] and Kallenbach[65] (1.6 – 1.9) in designed, medium-sized peptides.

Table 13. Comparisons of s values obtained by several different groups with those determined in the fit to the corrected protection factors.

	This Work	Scheraga	Baldwin	Stellwagen	Kallenbach
s_A	1.06	1.07	1.64	1.81	1.92
s_F	1.67	1.09	0.26	0.79	0.26
s_L	2.47	1.14	0.84	1.03	1.00
s_Q	0.91	0.98	0.60	0.58	0.45
s_R	1.30	1.03	1.10	1.94	0.91
s_S	0.27	0.76	0.39	0.29	0.48

The set of s values that minimizes SSQ_{lnPF} with corrected protection factors seem to faithfully reflect the factors that influence helicity in the eight peptide sequences studied in this chapter. This is evident from the comparison of calculated and measured t/c ratios in table 12. Nevertheless, SSQ_{lnPF} for this fit is still far too large to be due solely to measurement error; even correction of the protection factors could not reduce SSQ_{lnPF} enough to make it explicable without appealing to model error. In the final section, speculations as to the nature of the non-canonical interactions that could be responsible for this model error are presented.

---

[62] Wójcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121.
[63] Park, S.-H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 7048.
[64] Park, S.-H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 12901.

5.4.6 Discussion

One of the goals of this chapter is to determine whether the causes of s value context dependence extend significantly beyond the canonical interactions. The means to this goal is a proposed link between the dependence of s values on non-canonical interactions and the quality of the fit achievable on the protection factor data by a model that disregards such interactions. It has now been shown that the minimum SSQ_{lnPF} from the fit of the SO_3Hel_1-peptide conjugates' protection factors, even when they are corrected for possible assumption error, is too large to be due only to measurement error. Model error must be the cause of the excess in SSQ_{lnPF}, and since the peptides were designed to avoid known canonical interactions, the model error must be due to non-canonical interactions. This is in accord with the results of the permutation test carried out on the corresponding $*AcHel_1$-peptide conjugates in section 5.4.1. Is it possible to determine which peptides harbor the interactions that cause the model error using the protection factor data? This study was not designed to identify particular interactions, so it must be emphasized that any attempt to do so is merely speculation. Nevertheless, the urge to rationalize this study's results in terms of particular physical events will be indulged below.

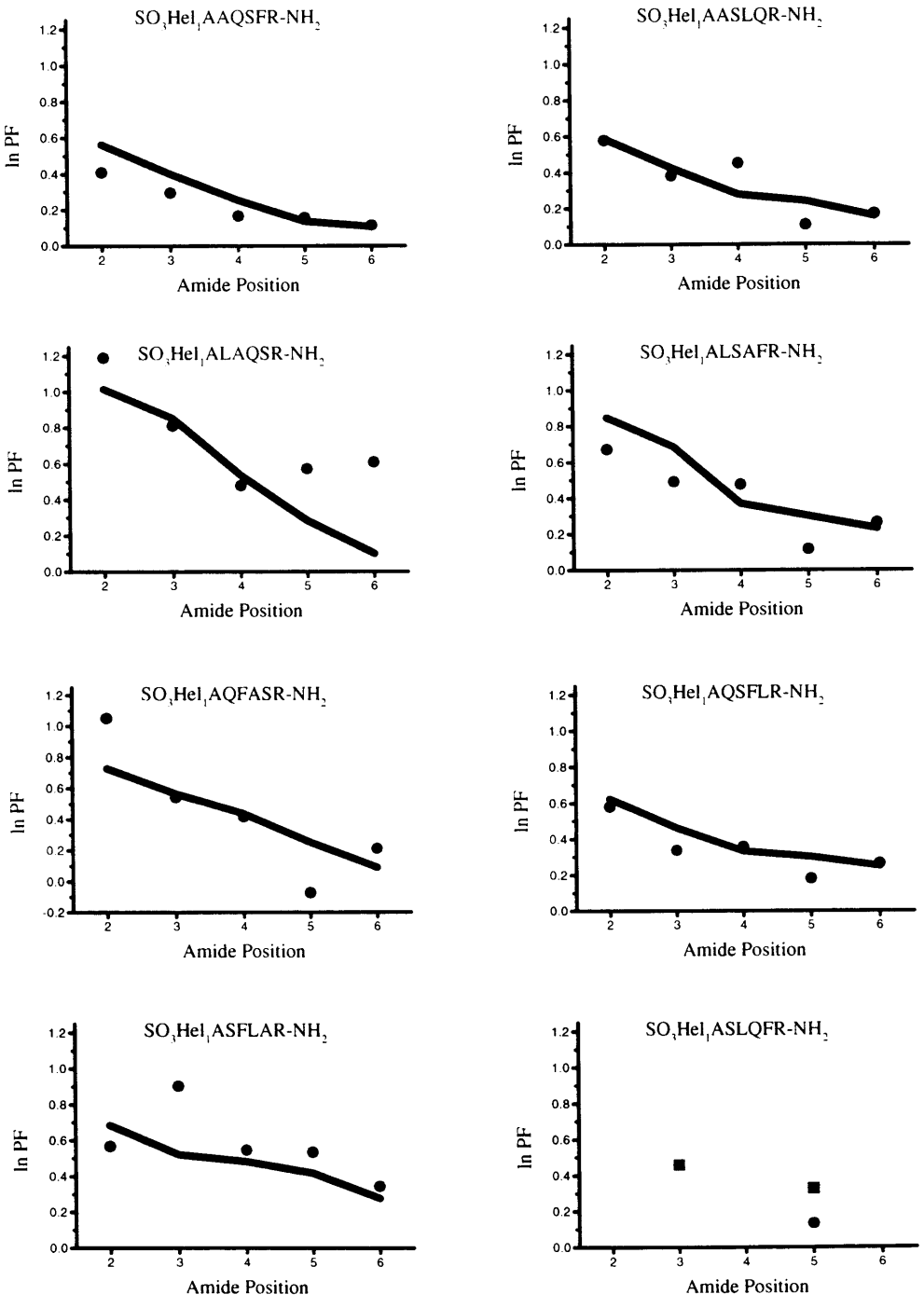
On the basis of the large drop between the second and third protection factors when the second protection factor is uncorrected, it has already been speculated that Leu and Gln might interact with the template is such a way as to increase the first s value. This is one possibility. Other possibilities can be entertained by considering the results of the fit in which the protection factors for the second amides and the amides associated with alanine residues are corrected (the fit from section 5.4.5). The natural logarithms of the

---

[65] Yang, J.; Spek, E. J.; Gong, Y.; Zhou, H.; Kallenbach, N. R. *Protein Sci.* **1997**, *6*, 1264.

measured (filled circles) and calculated (solid line or squares) protection factors from the fit using corrections are plotted in figure 8 on the following page. (For the purposes of interpreting these plots, it should be noted that the corrections applied to the second and alanine amides' protection factors are included in the measured ln PFs, not the calculated ln PFs). The largest deviations between the observed and calculated ln PFs occur at the C-terminus of $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$ (ln PF_{exp} – ln PF_{calc} = 0.29 for Ser and 0.51 for Arg), the Phe amide of $SO_3Hel_1ASFL^{\dagger}AR\text{-}NH_2$ (ln PF_{exp} – ln PF_{calc} = 0.38), and the Ser and Gln amides of $SO_3Hel_1AQFASR\text{-}NH_2$ (ln PF_{exp} – ln PF_{calc} = -0.33 and 0.32 respectively).

Figure 8. Calculated (line or filled squares) vs. measured protection factors (filled circles; adjusted as appropriate by correction factors).

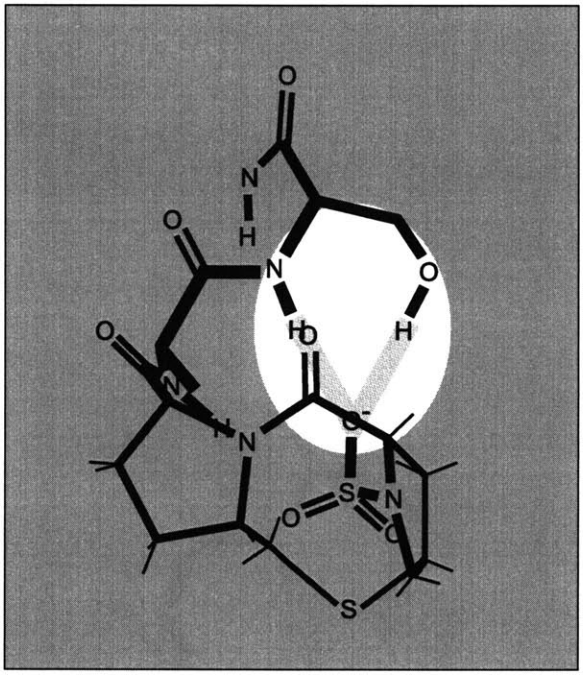


Since the protection factors of $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$ have the largest negative impact on the quality of the fit of any single peptide's protection factors, $AL^{\dagger}AQSR$ probably harbors the strongest interaction observed in any of the peptides. As shown in figure 8, the largest deviations of the measured from the calculated protection factors in this peptide are positive ($PF_{exp} > PF_{calc}$) and occur toward the C-terminus. This is consistent with the surprising NOEs found toward the C-terminus of $SO_3Hel_1AL^{\dagger}AQSR\text{-}NH_2$ in chapter 4, section 4.3.3f. This unexpected retention of helicity at the C-terminus suggests that the helix stabilizing effect in $AL^{\dagger}AQSR$ should involve the arginine residue, but an interaction between the side-chains of arginine and leucine or alanine (the residues at the i-3 and i-4 positions) is unlikely. It may be that the very small alanine side-chain at the i-3 position allows an interaction between the arginine side-chain and the helix barrel that larger side-chains inhibit. Such interactions have been demonstrated for lysine with NOE and other evidence and offered as explanations for the high helicities of peptides with sequences based on A_2KA_2 repeats.[59] Since $AL^{\dagger}AQSR$ is the only peptide with an alanine at the critical i-3 position, $AL^{\dagger}AQSR$ could benefit from this interaction.

The protection factors of $SO_3Hel_1ASFL^{\dagger}AR\text{-}NH_2$ also notably upset the protection factor fitting, and figure 8 shows that the measured protection factors are consistently larger than the calculated. More striking, though, is the discrepancy between the experimental and calculated t/c ratios that $*AcHel_1ASFL^{\dagger}AR\text{-}NH_2$ shares only with $*AcHel_1ASLQFR\text{-}NH_2$ (see section 5.4.5). Such a discrepancy could be conveniently explained if, through some common feature, both $ASFL^{\dagger}AR$ and ASLQFR were able to interact with SO_3Hel_1 in a way that is not possible with $AcHel_1$. The most conspicuous common feature of these two peptides is the serine at the second position and the only

difference between SO_3Hel_1 and $AcHel_1$ is the substitution of a sulfamate for an acetyl group. This suggests that a helix-stabilizing hydrogen bond might form between the serine hydroxyl and the SO_3Hel_1 sulfamate. This might occur as pictured in figure 9.

Figure 9. Hydrogen bonding between the side-chains of serines at the second position in SO_3Hel_1-peptide conjugates and the sulfamate of the SO_3Hel_1 template.



This interaction could not be as strong in the corresponding peptide with the $AcHel_1$ template, since helix formation in this template already requires the amide carbonyl to hydrogen bond with both of the first two amide NHs (see figure 1 of chapter 4). It should be pointed out that this interaction is different from the interactions between the serine side chain and the main chain that have been proposed to explain serine's high N-capping propensity. When acting as an N-cap, the serine hydroxyl serves as a hydrogen bond acceptor for the N-terminal amide NHs that are not hydrogen bonded to amide

carbonyls[66]; here the serine hydroxyl is a donor and, together with its amide NH, forms a bifurcated hydrogen bond with a sulfamate oxygen. Such a serine-sulfamate interaction would increase the serine's s value (as opposed to its N-cap propensity), and thus to increase the helicity of any SO_3Hel_1-peptide conjugate with serine in the second position beyond that observed in other systems, such as $AcHel_1$-peptide conjugates, that lack this interaction.

The influence of $SO_3Hel_1AQFASR\text{-}NH_2$ on the protection factor fitting could be ascribed to a Gln-Arg i to i+4 interaction. Although i to i+4 Gln-Arg interactions themselves have never been demonstrated, similar i to i+4 Glu^0-Arg interactions have.[67] Two features of the data oppose this attribution. First, if this were true, $SO_3Hel_1AQSFLR\text{-}NH_2$ should also show the signs of this interaction, but it does not. Second, as mentioned above, a helix-stabilizing interaction that involves C-terminal residues should result in positive deviations of the experimental from the calculated protection factors toward the C-terminus, whereas the positive deviation in $SO_3Hel_1AQFASR\text{-}NH_2$ occurs at the N-terminus (see figure 8), suggesting that the context dependent helix propensity should belong to glutamine or phenylalanine. Perhaps an i to i+1 interaction occurs between these two residues, or the Gln residue is able to interact with the template.

---

[66] Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21.
[67] Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668.

5.5 Conclusions

Conclusions relating to the issues addressed in each of the three sections of this chapter can be drawn. First, it was shown that the intrinsic exchange rate constants calculated from the model proposed by Englander and co-workers[21] sometimes deviated from measured exchange rate constants by more than 60%, and often deviated by more than 30%. This error, while acceptable for studies of protein folding, cannot be tolerated in studies of peptide helicity. We have introduced a new means for determining the intrinsic exchange rate constants for use in studies of peptide helicity in which helices are initiated by the SO_3Hel_1 template. The intrinsic exchange rate constants are measured in $HHel_1$-peptide conjugates, which are identical to the SO_3Hel_1-peptide conjugates in every way except for the group present at the N-terminus (a free amine in $HHel_1$ vs. a sulfamate in SO_3Hel_1). Second, it was shown that the protection factors, with the exception of the last protection factor (which is always associated with a charged residue), decrease consistently from the N- to the C-terminus. This is an explicit demonstration that SO_3Hel_1 induces helices that fray monotonically from the N- to the C-terminus. Third, it was shown using both the protection factors of a set of eight SO_3Hel_1-peptide conjugates and the t/c ratios from the eight corresponding $*AcHel_1$-peptide conjugates that non-canonical interactions significantly affect peptide helix formation. The exact nature of these interactions could not be determined from this work, but their occurrence is strongly suggested by the data. Further work to explicitly identify non-canonical interactions and quantify their energetics should enable models for peptide helicity to be adjusted for their effects.

5.6 Experimental

Equipment. One dimensional ^{1}H-NMR spectra were measured at 500 MHz and ^{13}C-NMR spectra were measured at 125 MHz on Varian VXR500S and 501S spectrometers and processed using the Varian Instruments VNMR 3.1 software. Chemical shifts were measured relative to the reference signal of (trimethylsilyl) propionic-2,2,3,3-d_4 acid (TMSP). Most ^{1}H-NMR spectra were obtained using a 60° pulse width with a 2 s acquisition time (40,000 points) and no delay between pulses. ^{1}H-NMR spectra for t/c ratio measurement were obtained using a 60° pulse width with a 4 s acquisition time and a 12 s delay between pulses. Between 128 and 512 transients were acquired, depending on the concentration of the sample (typically between 0.5 and 3 mM). Analytical high performance liquid chromatography (HPLC) was performed on a Waters system consisting of two 501 pumps, a rheodyne injector, a model 660 automated gradient controller, a model 740 data module, a model 484 detector, and a Vydac 0.46 × 25 cm (218TP54) C_{18} reverse phase column. Flow rates were 1.0 mL/min. Preparative scale HPLC was performed on a Waters system consisting of a model 590 pump fitted with preparative heads, an Autochrome DPG/S pre-pump solvent mixer, a Rheodyne injector, a model 484 variable wavelength detector, and a Waters 2.5 × 10 cm radial compression column housed in a PrepLC 2.5 cm radial compression module (RCM). Flow rates for preparative HPLC were 12 mL/min. Detection in all uses of HPLC was carried out at 214 nm unless otherwise specified. The pDs of solutions were measured using a Cole-Parmer pH meter (model # 5982-00) and a 3.5 × 183 mm glass electrode with a calomel reference (model # 5990-30). The pH meter was referenced using pH 4.00 and 7.00

certified buffers. The H/D isotope effect for D_2O solutions was accounted for by adding 0.4 to the pD that was read off the meter ($pD = pD_{read} + 0.4$).[68] pDs were recorded at room temperature, although exchange rate measurements were made at 5 °C. Since the first pK_a of H_3PO_4 decreases by 0.075 between 25 °C and 5 °C,[69] 0.075 was subtracted from the recorded pDs to account for the temperature. Mass spectra were obtained courtesy of Dr. P. Wishnok and Prof. S. Tannenbaum on a Hewlett-Packard HP5989B electrospray ionization mass spectrometer from samples dissolved in 1:1 water: methanol with 0.1% acetic acid, detecting positive ions for $HHel_1$-peptide conjugates or negative ions for SO_3Hel_1-peptide conjugates. The simulated data of section 5.4.2c and appendix n were generated and all minimizations of SSQ_{lnPF} functions were performed using Mathematica 3.0 (Wolfram Research Inc, 1996) on a PC platform.

Two Dimensional NMR spectroscopy. TOCSY spectra (in 9:1 H_2O:D_2O) were acquired for all compounds on the 500 MHz Varian spectrometers described above at 5 or 25°C. The spectral width was 5000 Hz (10ppm) in both dimensions and 32 t_1 increments were obtained each consisting of 32 transients. The mixing time was 60 ms and the H_2O peak was suppressed by 1 s of presaturation. Data were processed with gaussian weighting in both dimensions.

Measurement of t/c Ratios. The t/c ratios for $*AcHel_1$-peptide conjugates were measured from ^{1}H-NMR spectra as described in chapter 1, and from ^{13}C-NMR spectra as described in chapter 3. All t/c ratios were measured in the buffer used for the amide exchange reactions at 5°C.

---

[68] Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188.
[69] Bates, R. G. *J. Res. Natl. Bur. Stand. (U. S.)* **1951**, *47*, 127.

Amide Hydrogen Exchange Reactions: Procedure. The deuterated exchange reaction buffer was prepared by dissolving 0.02 mmol of KH_2PO_4 in D_2O, removing the solvent and repeating twice. The resulting 0.02 mmol of KD_2PO_4 was diluted to 200 mM with D_2O (~ 100 mL), and the solution was adjusted to pD 3.12 with CF_3COOD. Peptides were prepared for exchange studies by desalting them as necessary (by re-purifying them by preparative HPLC) and lyophilizing them at least twice from deionized H_2O. The peptide of interest (2 – 5 mg) and ~ 5 mL of buffer were then placed in separate 15 mL Falcon tubes and chilled on ice for 10 min. About 1 mL of the cold buffer was then added to the cold peptide in the Falcon tube and the suspension was mixed until it dissolved by alternately taking it into and expelling it from a pipette. Dissolution was usually complete within 1 min. The solution, typically 1 to 4 mM in peptide, was then placed in an NMR tube and inserted into a probe that had been pre-equilibrated to 5°C. The solution was allowed to equilibrate with the probe (10 min), and NMR spectra (30 scans, 2 s acquisition time, no delay; the total time taken to measure a single spectrum = 1 min) were acquired on a regular schedule thereafter: for $HHel_1$-peptide conjugates and Ac-peptides spectra were acquired every 15 min for 120 to 180 min, while for SO_3Hel_1-peptide conjugates spectra were acquired every 20 min for 180 to 240 min (all of these times include the 1 min required to measure the NMR spectra). These time ranges usually covered at least 3 half-times except for some Leu amides, which had the slowest exchange rates. After the exchange reaction was complete, the peptide solution was warmed to room temperature and its pD was checked. The pD was of the peptide solution was typically slightly more acidic than the original buffer, probably because of residual trifluoroacetic acid from the purification by preparative HPLC. If the pD was

less than 2.88 or more than 3.18, the results from the exchange reaction were not used. The pDs of all the exchange reactions for the $HHel_1$- and SO_3Hel_1-peptide conjugates are recorded in appendix 1.

Amide Hydrogen Exchange Reactions: Data Fitting. After processing, the integrals for each amide proton's peak (relative to the integral of a non-exchangeable hydrogen's peak) at each time point were measured, either alone or in pairs, depending on whether the peak overlapped inextricably with another amide's peak. The data were then fit to either a single or a double exponential decay equation:

$$\text{Integral}(t) = C \times e^{-k_{XH} \times t}$$

or

$$\text{Integral}(t) = C_1 \times e^{-k_{XH1} t} + C_2 \times e^{-k_{XH2} t}$$

where C, C_1, and C_2 are constants, k_{XH}, k_{XH1}, and k_{XH2} are the exchange rate constants, and t is time. Note that an additive constant is left out; this should not have a large impact on the fit as this constant should be very small. It corresponds to the integral that an amide peak would have after the exchange reaction reached equilibrium, where the H/D distribution for the amide is about the same as the H/D distribution in the solvent, about 100:1. Since the exchange reactions were run for 3 half-times, so that the final data point should have had an intensity about 12.5% of the first data point, the additive constant should have amounted to no more than 10% of the integral at the final data point (1% of the integral at the first data point). The error in the integral was assumed to be multiplicative, so the objective of the least squares fitting was to minimize the sum of the squared differences of the experimental and calculated ln (Integral):

$$SSQ_{lnInt}([C, k_{XH}]_{optimal}) = \text{Minimum} \sum_{\text{all time points}} (\ln[\text{exp.integral}] - \ln[\text{calc.integral}])^2$$

The effect of experimental error on this protocol for double exponential fitting was simulated as follows. Given assumed values for C_1, C_2, k_{XH1}, and k_{XH2}, "true" values for the integral were computed for 9 time points spaced 15 min apart, covering 120 min. Error was added by multiplying the true integral by a factor e^{ε} where ε was a number randomly generated from a normal distribution centered at 0 with a standard deviation 0.02. This corresponds to a 2% error in the integral, roughly what was observed experimentally. In five trials where the assumed $(k_{XH1}, k_{XH2})_{true}$ = (0.0200 min^{-1}, 0.0050 min^{-1}) the fit resulted in $(k_{XH1}, k_{XH2})_{best\ fit}$ = (0.0201 min^{-1}, 0.0049 min^{-1}), (0.0202 min^{-1}, 0.0051 min^{-1}), (0.0175 min^{-1}, 0.0057 min^{-1}), (0.0189 min^{-1}, 0.0050 min^{-1}), (0.0205 min^{-1}, 0.0046 min^{-1}). In five trials where the difference between the k_{XH1} and k_{XH2} was less, $(k_{XH1}, k_{XH2})_{true}$ = (0.0150 min^{-1}, 0.0100 min^{-1}) the fit resulted in $(k_{XH1}, k_{XH2})_{best\ fit}$ = (0.0140 min^{-1}, 0.0106 min^{-1}), (0.0143 min^{-1}, 0.0102 min^{-1}), (0.0123 min^{-1}, 0.0120 min^{-1}), (0.0127 min^{-1}, 0.0117 min^{-1}), (0.0148 min^{-1}, 0.0100 min^{-1}). In both cases, the best-fit exchange rate constants match well with the true exchange rate constants.

Once they had been determined by the above fitting process, k_{XH1} and k_{XH2} from the double exponential decay had to be assigned to individual amides. As noted in section 5.2.4b, this could usually be done based on a priori expectations; occasionally, however, the assignment had to be based on the assumed similarity of the exchange rate constants of amides in similar environments. For example, in $HHel_1AAQSLR\text{-}NH_2$, the amides of Arg and the second Ala overlap, and there is no reason to expect one to be larger than the other is. The rate constant for the second Ala in $HHel_1A^{\dagger}AQSFR\text{-}NH_2$ could be

unambiguously determined by single exponential decay. Since the amides of the second alanines of $HHel_1A^{\dagger}AQSFR\text{-}NH_2$ and $HHel_1AASLQR\text{-}NH_2$ are in very similar environments, the exchange rate constants from the double exponential decay of the Ala and Arg resonances are assigned so that the exchange rate constants of the second alanines in $HHel_1AASLQR\text{-}NH_2$ and $HHel_1A^{\dagger}AQSFR\text{-}NH_2$ are similar. The exchange rate constants determined from double exponential decay fitting for random coil peptides were assigned as follows.

For $HHel_1$ and acetylated peptides: AAQSFR: Phe and Ser (larger assigned to Ser). AASLQR: Ala and Arg (more similar to Ala of AAQSFR assigned to Ala), Leu and Ser (larger assigned to Ser). ALAQSR: Gln and Ser (larger assigned to Ser), Arg with one of the two doublets from $\alpha^{15}N$ Ala (more similar to the other doublet from $\alpha^{15}N$-Ala assigned to Ala). ALSAFR: Leu and Ala (larger assigned to Ala), Phe and Arg (larger assigned to Arg). AQFASR: Gln and Ala (more similar to Gln of AQSFLR assigned to Gln). AQSFLR: Phe and Arg (larger assigned to Arg). ASFLAR: Ser and Phe (larger assigned to Ser).

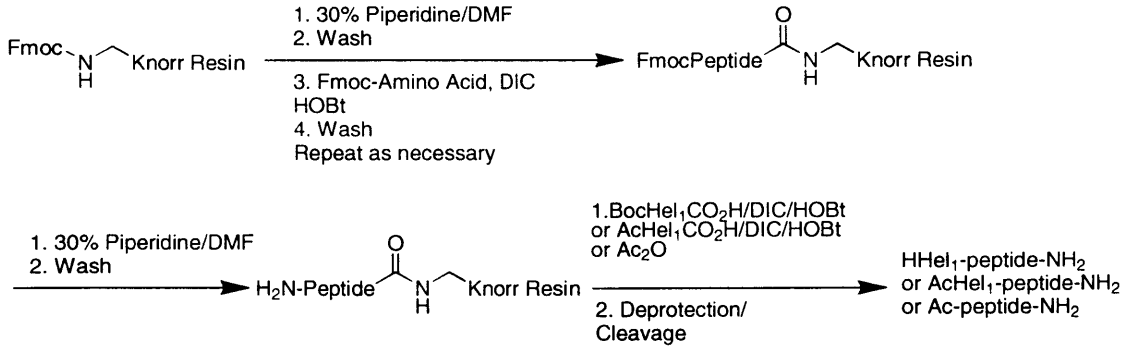
Overlapping amides were more rare in SO_3Hel_1-peptide conjugates. Where they did occur, however, they could not necessarily be assigned according to what would be expected from amides in random coils. However, in ambiguous cases the expected properties of protection factors could be used to make the assignments; that is, an assignment was judged correct if it led to a more monotonic set of protection factors than its opposite assignment would have, or if it led to two PFs > 1 while its opposite led to one or both PFs < 1. The exchange rate constants determined from double exponential decay fitting for SO_3Hel_1-peptide conjugates were assigned as follows: AAQSFR: Ser

and Phe (larger assigned to Ser). ALAQSR: Ser and Arg (nearly identical, but the larger were assigned to Ser). ALSAFR: Ser and Ala (larger assigned to Ala for improved monotonicity). AQFASR: Ser and Arg (larger assigned to Ser; opposite assignment would have led to an Arg PF << 1). ASLQFR: Phe and Leu (larger assigned to Phe).

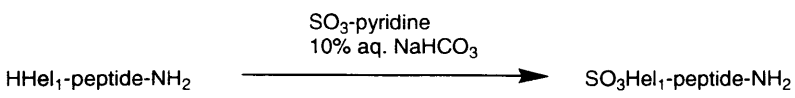
Synthesis: Description

$HHel_1$- and SO_3Hel_1-peptide conjugates were prepared according to the procedure detailed in chapter 4. Fmoc(^{15}N)Ala-OH was used to introduce the ^{15}N label where desired. Acetylated peptides were synthesized from precursors to the $HHel_1$-peptide conjugates; at the last step, the appropriate resin bound peptides were treated with acetic anhydride rather than an activated form of $BocHel_1$. The syntheses are summarized in schemes 1 and 2. All peptides were characterized by NMR and mass spectroscopy.

Scheme 1. Peptide synthesis



Scheme 2. Sulfamation of $HHel_1$-peptide conjugates



Synthesis: Procedures

The solid-phase peptide synthesis procedures used in the preparation of all compounds have been described in chapter 3. The preparation of $HHel_1$-peptide conjugates and their sulfamation to produce SO_3Hel_1-peptide conjugates have been described in chapter 4. The preparation of *$AcHel_1$-peptide conjugates, where the template acetyl group is doubly labeled with ^{13}C, has been described in chapter 3. The acetyl analogs of $HHel_1$-peptide conjugates were prepared by treating the resin bound peptides with 10% acetic anhydride in DMF instead of $BocHel_1CO_2H$ / DIC / HOBt at the step before the deprotection / cleavage. The acetyl capped peptides were then deprotected / cleaved and purified just as their $BocHel_1$ capped analogs were.

The calculated vs. observed masses and HPLC retention times for the $HHel_1$-, SO_3Hel_1-, $AcHel_1$-, and Ac-peptides are summarized below in table 14.

Table 14. Characterization of $HHel_1$-, SO_3Hel_1-, $AcHel_1$-, Ac- peptides. The retention times are for a 5% – 100% CH_3CN (remainder 0.1% TFA in H_2O) gradient over 40 min on the preparative HPLC system described above.

Peptide	End group	observed m/z (calculated m/z)	retention time
A†AQSFR	$HHel_1$	$(M+H)^+$: 917 (917.4)	9.0 min
	SO_3Hel_1	M^-: 995 (995.4)	10.2
	$AcHel_1$	$(M+H)^+$: 961 (961.4)	12.7
	Ac	$(M+H)^+$: 721 (721.4)	7.3
AASLQR	$HHel_1$	$(M+H)^+$: 883 (882.5)	8.4
	SO_3Hel_1	M^-: 960 (960.4)	11.0
	$AcHel_1$	$(M+H)^+$: 926 (926.5)	13.6

AL†AQSR	$HHel_1$	$(M+H)^+$: 883 (883.5)	8.4
	SO_3Hel_1	M^-: 961 (961.4)	9.9
	$AcHel_1$	$(M+H)^+$: 927 (927.5)	12.7
	Ac	$(M+H)^+$: 687 (687.4)	7.6
ALSAFR	$HHel_1$	$(M+Na)^+$: 923 (923.5)	11.2
	SO_3Hel_1	M^-: 979 (979.4)	12.6
	$AcHel_1$	$(M+H)^+$: 945 (944.5)	14.2
AQFASR	$HHel_1$	$(M+H)^+$: 916 (916.4)	9.7
	SO_3Hel_1	M^-: 995 (994.4)	12.5
	$AcHel_1$	$(M+H)^+$: 961 (960.4)	13.2
AQSFLR	$HHel_1$	$(M+H)^+$: 958 (957.5)	10.6
	SO_3Hel_1	M^-: 1036 (1036.4)	13.6
	$AcHel_1$	$(M+H)^+$: 1002 (1002.5)	16.9
	Ac	$(M+H)^+$: 763 (762.4)	9.0
ASFL†AR	$HHel_1$	$(M+H)^+$: 902 (902.5)	10.3
	SO_3Hel_1	M^-: 980 (980.4)	12.9
	$AcHel_1$	$(M+H)^+$: 946 (945.5)	14.0
	Ac	$(M+H)^+$: 706 (705.4)	9.4
ASLQFR	$HHel_1$	$(M+H)^+$: 959 (958.5)	9.9
	SO_3Hel_1	M^-: 1036 (1036.4)	12.8
	$AcHel_1$	$(M+H)^+$: 1003 (1002.5)	13.7

Appendix 1. Tables of Raw Exchange Data

Table 15. Amide exchange rate constants for $HHel_1$-peptide conjugates. All exchange rate constants are in min^{-1}.

$HHel_1$-Peptide	Exchange Rate Constants by Amide Position					
	1st	2nd	3rd	4th	5th	6th
AAQSFR	Ala	†Ala	Gln	Ser	Phe	Arg
pD = 3.18	~ 0.1 min^{-1}	0.017	0.0151	0.0232	0.0131	0.0155
pD = 2.94		0.0274	0.0189	0.0172	0.0172	0.0172
pD = 3.01		0.0218	0.0171	0.0248	0.0121	0.0169
AASLQR	Ala	Ala	Ser	Leu	Gln	Arg
pD = 3.00	~ 0.1 min^{-1}	0.0219	0.0201	0.00587	0.0109	0.0106
pD = 2.98		0.0155	0.0141	0.00753	0.0127	0.0155
pD = 2.96		0.0221	0.0207	0.00587	0.0116	0.0113
ALAQSR	Ala	Leu	†Ala	Gln	Ser	Arg
pD = 3.04	~ 0.1 min^{-1}	0.00703	0.0145	0.0141	0.022	0.0179
pD = 2.95		0.00749	0.0169	0.0144	0.0232	0.0239
ALSAFR	Ala	Leu	Ser	Ala	Phe	Arg
pD = 3.01	~ 0.1 min^{-1}	0.00705	0.0126	0.0108	0.01	0.0138
pD = 2.97		0.00637	0.0133	0.0132	0.012	0.0129
pD = 3.01		0.00523	0.012	0.0134	0.00841	0.0158
AQFASR	Ala	Gln	Phe	Ala	Ser	Arg
pD = 3.01	~ 0.1 min^{-1}	0.0177	0.00912	0.00952	0.0189	0.0155
pD = 2.98		0.0186	0.00973	0.0101	0.0219	0.016
pD = 3.00		0.0199	0.0102	0.0111	0.0197	0.0181
AQSFLR	Ala	Gln	Ser	Phe	Leu	Arg
pD = 3.07	~ 0.1 min^{-1}	0.0148	0.0149	0.0104	0.00492	0.0104
pD = 3.01		0.0162	0.0169	0.00925	0.00506	0.0129
ASFLAR	Ala	Ser	Phe	Leu	†Ala	Arg
pD = 3.11	~ 0.1 min^{-1}	0.0197	0.0104	0.00436	0.0138	0.0122
pD = 2.95		0.0215	0.00961	0.00444	0.0153	0.0133
pD = 2.96		0.0207	0.00966	0.00465	0.0177	0.0143
ASLQFR	Ala	Ser	Leu	Gln	Phe	Arg
pD = 3.03	~ 0.1 min^{-1}	-	0.00686	-	0.00926	-
pD = 3.00			0.00604		0.00842	
pD = 3.01			0.00643		0.00882	

Table 16. Amide exchange rate constants for SO_3Hel_1-peptide conjugates. All exchange rate constants are in min^{-1}.

SO_3Hel_1-Peptide	Exchange Rate Constants by Amide Position					
	1st	2nd	3rd	4th	5th	6th
AAQSFR	Ala	ǂAla	Gln	Ser	Phe	Arg
pD = 3.09	0.0126	0.0114	0.0128	0.017	0.0125	0.0157
pD = 3.03	0.0132	0.0112	0.0124	0.0194	0.0114	0.0138
AASLQR	Ala	Ala	Ser	Leu	Gln	Arg
pD = 3.04	0.0133	0.00812	0.0121	0.00406	0.0103	0.00987
pD = 3.02	0.0121	0.00905	0.0124	0.00396	0.0103	0.0105
pD = 3.00	0.0129	0.00872	0.0124	0.00414	0.0109	0.0107
ALAQSR	Ala	Leu	ǂAla	Gln	Ser	Arg
pD = 3.05	0.00677	0.00129	0.00935	0.00867	0.0131	0.0102
pD = 3.01	0.00734	0.00122	0.00991	0.00899	0.0124	0.0124
ALSAFR	Ala	Leu	Ser	Ala	Phe	Arg
pD = 3.01	0.00863	0.00174	0.00705	0.0111	0.00876	0.0105
pD = 3.01	0.00924	0.00189	0.00959	0.00958	0.00932	0.0114
pD = 3.01	0.00841	0.00175	0.00679	0.0114	0.00872	0.0106
AQFASR	Ala	Gln	Phe	Ala	Ser	Arg
pD = 3.04	0.00854	0.00362	0.0055	0.0095	0.021	0.0132
pD = 3.02	0.00838	0.00382	0.00576	0.00914	0.0223	0.0132
AQSFLR	Ala	Gln	Ser	Phe	Leu	Arg
pD = 3.04	0.0109	0.00478	0.011	0.00672	0.00435	0.0091
pD = 3.07	0.0113	0.0052	0.0123	0.00712	0.00411	0.00927
pD = 3.02	0.00979	0.00482	0.0107	0.00674	0.00403	0.00827
ASFLAR	Ala	Ser	Phe	Leu	ǂAla	Arg
pD = 3.04	0.0106	0.0063	0.0034	0.00263	0.0132	0.0091
pD = 3.03	0.0106	0.00701	0.00472	0.00257	0.012	0.00966
ASLQFR	Ala	Ser	Leu	Gln	Phe	Arg
pD = 3.03	0.00823	0.00823	0.00392	0.0083	0.00715	0.0115
pD = 3.00	0.00733	0.00827	0.00389	0.00877	0.00943	0.0105
pD = 3.01	0.00844	0.00844	0.00441	0.00817	0.00682	0.0123

Appendix 2. Mathematica® Program for Fitting Protection Factor Data According to Equations 5-4 and 5-5

```
<<Statistics`

(*Define variables; the lower case letters a, q, s, f, l, and r will
stand for the s values of the various residues; b is the template
constant for SO3Hel1*)

a=.; q=.; s=.; f=.; l=.; r=.; b=.; ga=.; gq=.; gs=.; gf=.; gl=.; gr=.;
gb=.;lnc1=.;lnca=.; lncs=.;n=.; totPF=.;param=.;sst=.; sse=.;

(*These correspond to the sequences of the SO3Hel1-peptides*)

seq={{a,a,q,s,f,r},{a,a,s,l,q,r},{a,l,a,q,s,r},{a,l,s,a,f,r},{a,q,f,a,s
,r},{a,q,s,f,l,r},{a,s,f,l,a,r},{a,s,l,q,f,r}}

(*Equilibrium constants for all the states, from the full coil with no
hydrogen bonds (1) to the state with only one hydrogen bond (b), to the
various N-terminally initiated helices*)

helixfactors=Table[Join[{1,b},Table[b*Product[seq[[j, ii]], {ii, 1,
i}], {i, 1, Length[seq[[j]]]}]], {j, 1, Length[seq]}];

(*The sums of the above equilibrium constants*)

statesums=Table[Sum[helixfactors[[i,j]], {j, 1, Length[helixfactors]}],
{i, 1, Length[helixfactors]}]

(*Sums of equilibrium constants for the states in whinch the 2nd - 6th
amides are not hydrogen bonded*)

nonhbondedstatesums=Table[Table[Sum[helixfactors[[i,j]], {j, 1, ii}],
{ii, 2, Length[helixfactors[[i]]]-2}], {i, 1, Length[helixfactors]}]

(*Expressions for the ln PFs of the 2nd - 6th amides*)

calculatedlnPF=Table[Table[Log[statesums[[i]]]-
Log[nonhbondedstatesums[[i,j]]], {j, 1,
Length[nonhbondedstatesums[[i]]]}],{i, 1, Length[statesums]}]

(*experimental protection factors for the 2nd - 6th protection factors;
lnc1, lnca, and lncs are correction factors for the 2nd amide, Ala and
Ser*)

obslnPF={{0.650+lnc1+lnca, 0.297, 0.167+lncs, 0.157,
0.115},{0.820+lnc1+lnca, 0.383+lncs, 0.453, 0.109, 0.172},{1.755+lnc1,
0.486+lnca, 0.479, 0.572+lncs, 0.609},{1.237+lnc1, 0.492+lncs,
0.152+lnca, 0.116, 0.265},{1.616+lnc1, 0.542, 0.092+lnca, -0.072+lncs,
0.212},{1.145+lnc1, 0.338+lncs, 0.358, 0.182, 0.267},{1.132+lnc1+lncs,
0.903, 0.545, 0.209+lnca, 0.345},{n, 0.459, n, 0.134, n}};
```

```
(*Creating SSQlnPF*)

leastsq=0;
Do[Do[If [obslnPF[[i,j]]=!=n, leastsq=leastsq+(obslnPF[[i,j]]-
calculatedlnPF[[i,j]])^2, leastsq=leastsq+0], {j, 1,
Length[obslnPF[[i]]]}], {i, 1, Length[obslnPF]}]

(*The parameters are set so that s=Exp[ln s] and b = 0.20*)

a=Exp[ga]; b=0.20;q=Exp[gq];s=Exp[gs];f=Exp[gf];l=Exp[gl];r=Exp[gr];

(*Naive model with b=0.2 and no adjustment of PFs*)

lnc1=0.0;lnca=0.0; lncs=0.0;

(*This command minimizes SSQlnPF from the stated starting point*)

FindMinimum[leastsq, {ga, 0.0},{gq, 0.0}, {gs, 0.0}, {gf, 0.0}, {gl,
0.0}, {gr, 0.0}, MaxIterations -> 300]

(*Model with PF1 adjusted*)

lnc1=.;
FindMinimum[leastsq, {ga, 0.0}, {gq, 0.0}, {gs, 0.0}, {gf, 0.0}, {gl,
0.0}, {gr, 0.0},{lnc1, 0.0}, MaxIterations -> 300]

(*Model with b=0.2, PF1 and PFs of Ala residues adjusted*)

lnca=.;
FindMinimum[leastsq, {ga, 0.0}, {gq, 0.0}, {gs, 0.0}, {gf, 0.0}, {gl,
0.0}, {gr, 0.0},{lnc1, 0.0}, {lnca, 0.0},  MaxIterations -> 300]

(*Model with b=0.2, PF1, Ala and Ser PFs adjusted*)

lncs=.;
FindMinimum[leastsq, {ga, 0.0}, {gq, 0.0}, {gs, 0.0}, {gf, 0.0}, {gl,
0.0}, {gr, 0.0}, {lnc1, 0.0}, {lnca, 0.0}, {lncs, 0.0},  MaxIterations
-> 300]
```

Appendix 3. Fitting of Simulated Protection Factor Data

A3.1 Case of No Context Dependent s Values or Added Error

This case is included as a positive control, to show that the method returns the correct parameters in the ideal situation. Protection factors for the 2nd through the 6th positions of the SO_3Hel_1-peptide conjugates where peptide = AAQSFR, AASLQR, ALAQSR, ALSAFR, AQFASR, AQSFLR, ASFLAR, and ASLQFR were calculated using equation 5-2, with $B_{SO3} = 0.2$, $s_A = 1.07$, $s_F = 1.09$, $s_L = 1.14$, $s_Q = 0.98$, $s_R = 1.03$, and $s_S = 0.76$. The synthesized protection factors were converted to their natural logarithms and then fitted as per section 5.4.2a. Note that the protection factors for the template-peptide junction and C-terminal amides were not used here because, as will be detailed in section 5.4.4, experimental protection factors could not be obtained for these amides. Thus, forty synthesized protection factors were produced to yield forty terms of the form (ln synthesized PF – ln calculated PF)^2 in the expression for SSQ_{lnPF}. The parameters that were found to minimize SSQ_{lnPF} (when starting with all the ln s values set to 0) are listed in table 17 below.

Table 17. Parameters at SSQ_{lnPF} minimum in the case of no context dependent s values or added error.

SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
7.6×10^{-6}	1.07	1.09	1.14	0.98	1.01	0.76

The very small SSQ_{lnPF} shows that the synthesized data can be fit almost exactly, and the s values that result from the fit reproduce the s values that were input almost perfectly. This shows that, at least in an ideal case, the fitting procedure performs as it should.

A3.2 Case of No Context Dependencies, but with Added Error

The consequences of measurement error in the protection factors can be simulated by the following Monte Carlo-type method. Measurement error is artificially added to the protection factors from the preceding section by replacing the natural log of the protection factor of the j^{th} amide in the i^{th} peptide, ln PF_{ij}, with a number randomly generated from a normal distribution centered at the value of ln PF_{ij} and with standard deviation 0.1. This has the effect of adding about 10% noise to the protection factors, which is roughly the amount of error in the real, experimentally determined protection factors. The set of noise adjusted, synthesized ln PF_{ij} are used to build the SSQ_{lnPF} expression that is then minimized to determine the best-fit parameters. This process is repeated several times, from the generation of error laden protection factors to the fitting, to yield a number of sets of best-fit parameters that should represent the range of likely results. Ten sets of parameters were so obtained in this case, and the results for SSQ_{lnPF} and each s value are presented in table 18:

Table 18. Parameters at SSQ_{lnPF} minimum and their ranges over repeated trials in the case of no context dependent s values but with 10% added error.

	SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
Median	0.36	1.09	1.14	1.08	1.02	0.64	0.85
Max	0.44	1.15	1.40	1.70	1.10	5.33	1.26
Min	0.22	1.00	0.79	0.87	0.77	0.01	0.42

The median value for SSQ_{lnPF}, which corresponds to a root-mean-square deviation (RMSD) of 0.095 per protection factor, is about what one would expect given that an error of ±0.10 was introduced into the ln PFs. The range in SSQ_{lnPF} shows that the goodness of fit can vary, despite the same random error having been introduced in each trial, depending on how the error is distributed. This is expected, since SSQ_{lnPF} is proportional to a χ^2 random variable. The median values found for the parameters are, with the exception of s_R, good approximations to the input parameters. However, their ranges indicate that the error introduced into ln PF can cause the parameters to fluctuate substantially. The least susceptible is s_A, while the most susceptible is s_R, which varies over more than a 500-fold range probably because the arginine is the C-terminal residue in every peptide and therefore in every peptide s_R only contributes to one term in the sum of equilibrium constants. The results of this simulation show that in most cases accurate s values can be recovered from fitting protection factor data with 10% error, but it is likely that at least one of the set of optimal parameters will be significantly inaccurate, and under no circumstances can the s_R produced by the fitting be trusted.

A3.3 Several Different Levels of s Value Context Dependence with Added Error

The consequences of increasingly common s value context dependencies in addition to measurement error can be simulated by a Monte Carlo-type method similar to that used above. The difference comes only at the beginning, where s values different from those used previously are introduced into the computation of the synthesized protection factors for a few peptides. For all other peptides the s values are not altered. Error is then added, the resulting set of protection factors is fit, and the process is repeated to obtain ten sets of optimal parameters as above.

A3.3.1 Two Context Dependent s Values

Two sources of context dependency are introduced here. First, when Ser precedes Arg, s_S is reduced by a factor of 1.5 (corresponding to a coil-stabilizing interaction in which Ser and Arg can only interact when neither one is helical). Thus, an s_S of 0.51 is used for $SO_3Hel_1ALAQSR\text{-}NH_2$ and $SO_3Hel_1AQFASR\text{-}NH_2$ while for all other peptides $s_S = 0.76$. Second, when Leu is the second residue, s_L is increased by a factor of 1.5 (corresponding to a position dependence of s_L). Thus, an s_L of 1.71 is used for $SO_3Hel_1ALAQSR\text{-}NH_2$ and $SO_3Hel_1ALSAFR\text{-}NH_2$ while for all other peptides $s_L = 1.14$. The results of ten repetitions of the error introduction and fitting cycle with these two context dependencies are summarized in table 19.

Table 19. Parameters at SSQ_{lnPF} minimum and their ranges over repeated trials in the case of two context dependent s values and 10% added error.

	SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
Median	0.38	1.21	1.06	1.44	0.79	0.89	0.70
Max	0.61	1.27	1.19	1.75	1.02	1.99	0.85
Min	0.25	1.05	0.96	1.05	0.65	0.01	0.48

The median value of SSQ_{lnPF} is only slightly larger with the two context dependencies and added error than it is with just added error, indicating that it would be difficult by this method to distinguish mild cases of s value context dependence from complete context independence. The effect on the parameters is more substantial. The median value of s_L is significantly higher (1.44 vs. 1.08) and that of s_S significantly lower (0.70 vs. 0.85) than in section 5.3.7b; this is expected given that a higher s_L and a lower s_S were introduced in two cases each. In addition, the median of s_Q is much lower (0.79 vs. 1.02) than previously; this, however, is not expected as no interactions involving Gln were introduced. This is probably a manifestation of some unanticipated correlation between s_Q and s_L or s_S or both. The other parameters do not change as notably as these three. The ranges of the parameters are also comparable to those found in previous section.

A3.3.2 Four Context Dependent s Values

Two more sources of s value context dependence are introduced here in addition to the two of the preceding section. First, when Phe follows Ser, s_F is increased by a factor of 1.5 (corresponding to a helix-stabilizing interaction between Phe and Ser). Thus, an s_F of

1.64 is used for $SO_3Hel_1AAQSFR\text{-}NH_2$ and $SO_3Hel_1ASFLAR\text{-}NH_2$ while for all other peptides s_F = 1.09. Second, when Gln precedes Phe, s_Q is reduced by a factor of 1.5 (corresponding to a coil-stabilizing interaction in which Gln and Phe can only interact when neither one is helical). Thus, an s_Q of 0.65 is used for $SO_3Hel_1AQFASR\text{-}NH_2$ and $SO_3Hel_1ASLQFR\text{-}NH_2$ while for all other peptides s_Q = 0.98. The results of ten repetitions of the error introduction and fitting cycle with these four context dependencies are summarized in table 20.

Table 20. Parameters at SSQ_{lnPF} minimum and their ranges over repeated trials in the case of four context dependent s values and 10% added error.

	SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
Median	0.59	1.17	1.08	1.38	0.69	1.50	0.68
Max	1.03	1.25	1.33	1.75	0.86	2.89	1.15
Min	0.44	1.07	1.00	1.07	0.66	1.00	0.43

The median value of SSQ_{lnPF} is now significantly larger than in either of the previous two cases, indicating that s value context dependence at this level is easily distinguishable from s value context independence. The median values of most of the parameters change little upon the addition of these latest context dependencies. One would have expected s_F to increase and s_Q to decrease, and s_Q does indeed decrease slightly but the change in s_F is hardly noticeable. This could be because Phe is always associated with the decrease in s_Q, since s_Q only decreases when Gln follows Phe. Thus, its association with the decrease in s_Q could balance the increase in s_F. The parameter that changes by far the most is s_R, which increases by 70%. Apparently, the helix-stabilizing context dependencies are

expressing themselves partly in s_R. The parameters' ranges do not vary remarkably from before.

A3.3.3 Six Context Dependent s Values

Two more sources of s value context dependence are introduced here in addition to the four of the preceding sections. First, when Arg follows Phe, s_R is increased by a factor of 1.5 (corresponding to a helix-stabilizing interaction between Arg and Phe). Thus, an s_R of 1.55 is used for $SO_3Hel_1AAQSFR\text{-}NH_2$ and $SO_3Hel_1ALSAFR\text{-}NH_2$ while for all other peptides s_R = 1.01. Second, when Leu precedes Gln, s_L is reduced by a factor of 1.5 (corresponding to a coil-stabilizing interaction in which Leu and Gln can only interact when neither one is helical). Thus, an s_L of 0.76 is used for $SO_3Hel_1AASLQR\text{-}NH_2$ and $SO_3Hel_1ASLQFR\text{-}NH_2$ while for all other peptides s_L does not change. The results of ten repetitions of the error introduction and fitting cycle with these six context dependencies are summarized in table 21.

Table 21. Parameters at SSQ_{lnPF} minimum and their ranges over repeated trials in the case of six context dependent s values and 10% added error.

	SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S
Median	0.85	1.18	1.15	1.29	0.68	1.90	0.65
Max	1.24	1.35	1.53	1.53	0.74	3.46	0.94
Min	0.60	1.15	1.01	1.13	0.53	0.68	0.34

The median value of SSQ_{lnPF} again increases between the preceding case and this one, and the medians and ranges of the parameters again stay roughly the same for the most part. The median of s_L decreases slightly, that of s_F increases slightly, and that of s_R increases more substantially. The changes in s_L and s_R are to be expected based on the context dependencies introduced; that in s_F is probably due to the association of Phe with the context dependence of s_R.

Appendix 4. Exchange Rate Constants in Peptides in which SO_3ProPro replaces $HHel_1$, Ac, or SO_3Hel_1

The only notable difference between $HHel_1$-peptide conjugates and SO_3Hel_1-peptide conjugates is that the former bears a functional group with a positive charge at its N-terminus while the latter bears a functional group with a negative charge at its N-terminus. This could be enough to affect the k_{int} of the amides closest to the templates. Exchange rate constants, after all, are known to be influenced by effects more subtle than the presence of nearby charges; for example, the pH minimum for the exchange of proline amides is different by 0.75 pH units depending on whether the proline is in the cis or trans states.[21] The possibility that the negative charge in SO_3Hel_1-peptide conjugates affects k_{int} values was tested using two peptides in which SO_3ProPro replaces SO_3Hel_1: SO_3ProPro-AQSFLR-NH_2 (mass spec.: M^- = 992.43 observed, 992.46 calculated) and SO_3ProPro-ASFL†AR-NH_2 (mass spec.: M^- = 936.20 observed, 936.44 calculated) These were prepared using as the H-ProPro-peptides by standard solid-phase peptide synthesis methods and then sulfamated as described in chapter 4. The SO_3ProPro motif is very similar to SO_3Hel_1, except that it is missing the thiomethylene bridge between the proline rings. The lack of this constraint should keep the diproline unit from adopting the analog of SO_3Hel_1's helix-initiating structure. The k_{int} values for the amides in these peptides ($^{SO3PP}k_{int}$) were measured in the same way as for the $HHel_1$-peptide conjugates, and their averages are reported in table 22. The corresponding $^{HHel}k_{int}$ and $^{Ac}k_{int}$ values are included for comparison, and in figure 10 the average differences ln $^{SO3PP}k_{int}$ – ln $^{HHel}k_{int}$ are plotted as a function of amide position. Unlike in the $HHel_1$-peptide conjugates or the

acetylated peptides, the exchange reactions of the first amides are slow enough to be accurately measured (the $^{SO3PP}k_{int}$ is roughly half of the estimated $^{HHel}k_{int}$ of 0.1 min^{-1}). Similarly, the k_{int} of the second amide tends to be somewhat smaller (~ 20% on average) in the SO_3ProPro peptides than in the $HHel_1$-peptide conjugates. All of the other amides in the SO_3ProPro-peptides have exchange rate constants that are roughly comparable to those of their counterparts in the Ac- and the $HHel_1$-peptide conjugates.

Figure 10. Average difference between ln $^{Ac}k_{int}$ and ln $^{HHel}k_{int}$ (both measured in 200 mM phosphate, pD 3.0 ± 0.15, 5 °C) as a function of amide position. Datum at position 1 is approximate assuming ln $^{HHel}k_{int}$ = 0.1 min^{-1}.

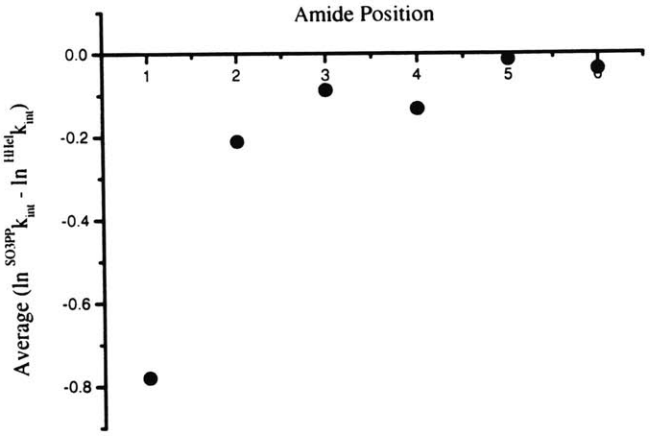


Table 22. Comparison of exchange rate constants determined at pD 3.0 ± 0.15 and 5 °C in SO_3ProPro-peptides, Ac-peptides, and $HHel_1$-peptide conjugates.

Peptide	N-terminus	Amide Exchange Rate Constants (min^{-1}) by Position					
		1st	2nd	3rd	4th	5th	6th
AQSFLR	SO_3ProPro	0.0494	0.0138	0.0141	0.0083	0.0051	0.0132
	Ac	> 0.1	0.0220	0.0171	0.0075	0.0048	0.0130
	$HHel_1$	~ 0.1	0.0155	0.0159	0.0098	0.0050	0.0116

ASFL†AR	SO_3ProPro	0.0426	0.0151	0.0093	0.0041	0.0147	0.0107
	Ac	> 0.1	0.0224	0.0099	0.0035	0.0147	0.0121
	$HHel_l$	~ 0.1	0.0206	0.0099	0.0045	0.0155	0.0132

These results indicate that the negative charge in SO_3ProPro-peptides depresses the exchange rate constants of the two most N-terminal amides, or that the positive charge in $HHel_l$ elevates them, or both. This could be interpreted in terms of the SO_3ProPro-peptides obtaining some structure in which the first two amides are hydrogen bonded to some extent to the sulfamate. In fact, it was proposed in appendix 2 of chapter 4 that a small amount of helicity might be induced by the similar SO_3ProAla motif. It is unlikely, however, that SO_3ProPro unit could assume the conformation necessary to accept hydrogen bonds from the first three amides since the adoption of a helical attitude by a proline precludes the preceding residue from doing the same. (For this reason prolines rarely occupy any of the internal positions in protein helices[70] although they are often the N-terminal residues.[66]) Moreover, AcProPro-peptides, which should be as helical relative to $AcHel_l$-peptide conjugates as SO_3ProPro-peptides are relative to SO_3Hel_l-peptide conjugates, have been shown to be structureless.[71,72] This is instead interpreted as an effect of the charge on k_{int} and this effect should, if anything, be stronger in peptides where SO_3Hel_l, in which the charge is constrained to be closer to the rest of the peptide, replaces the SO_3ProPro unit.

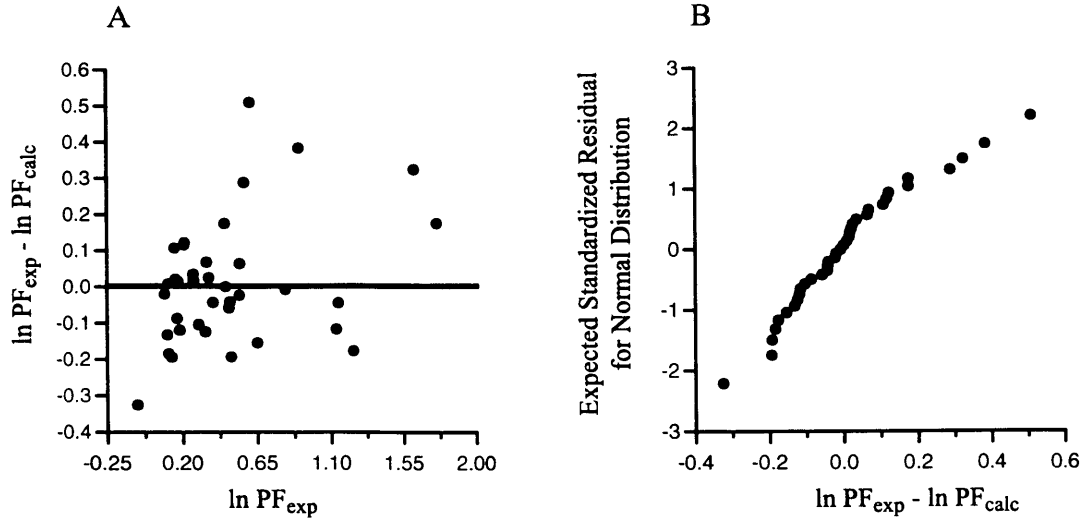
[70] Chou, P. Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, *47*, 45.
[71] Kemp, D. S.; Allen, T. J.; Oslick, S. L.; Boyd, J. G. *J. Am. Chem. Soc.* **1996**, *118*, 4240.
[72] Renold, P.; Kemp, D. S. Unpublished results.

Appendix 5. Features of the SSQ_{lnPF} Minimum in the Fit with Corrected Protection Factors

A5.1 Residuals of the Least-Squares Fit

The residuals of the fit (observed ln PFs – calculated ln PFs) are plotted as a function of the observed ln PFs in figure 11A, and as a function of the expected standardized residuals figure 11B (also known as a probability plot). Figure 11A shows that the residuals do not vary systematically as a function of the observed protection factors, and the linearity of the plot in figure 11B shows that the residuals are approximately normally distributed. These two observations support the assertion that the SSQ_{lnPF} is proportional to a χ^2 statistic.

Figure 11. Panel A: Plot of residuals from the protection factor fit as a function of the experimental protection factors. **Panel B:** Plot of the residuals from the protection factor fit plotted as a function of the expected standardized residuals from a normal distribution (probability plot).



A5.2 Uniqueness of the Minimum

Assertions made regarding the physical nature of helix formation based on the minimum quoted above for SSQ_{lnPF} are only relevant if some evidence can be presented that this is the global minimum, and not merely a local minimum. Minima located by optimization algorithms are tied to the point from which the algorithm is started, so given a particular starting point an algorithm will unfailingly find the same minimum, but given a different starting point the algorithm might find a different minimum. Thus, in order to show that the minimum at $SSQ_{lnPF} = 1.06$ is the global minimum, fifty random starting points for the minimization were chosen where the starting coordinate for each of the parameters (the ln s values, ln c_{PF2}, and ln c_{Ala}) was a random number between –5 and 5 (corresponding to a range 0.007 to 148 for the s values). In 44 out of 50 trials, the minimization converged to the minimum with $SSQ_{lnPF} = 1.06$; in the other six trials, the minimization converged to a slightly different minimum with $SSQ_{lnPF} = 1.08$. Representative values of the parameters that are found at these two critical values are summarized in table 23. This shows that the SSQ_{lnPF} function is smooth in the region between -5 and +5 along each parameter axis. If the search is extended such that the starting coordinates along each axis are random numbers between –10 and +10 (corresponding to a range from 4.5×10^{-5} and 2.2×10^{4} for the s values), SSQ_{lnPF} is revealed to be more rugged. While the minima at 1.06 and 1.08 are still by far the most commonly located, the optimization algorithm discovers several new minima. However, most of these can be disregarded as the values of SSQ_{lnPF} at these points are generally

between 1.5 and 5.2. None of the minima are less than those found before and only one minimum at $SSQ_{lnPF} = 1.12$ is comparable to them. The parameters for this minimum are also listed in table 23 (note that the parameters have been transformed from their natural logarithms).

Table 23. Three minima in the SSQ_{lnPF} function.

SSQ_{lnPF}	s_A	s_F	s_L	s_Q	s_R	s_S	c_{PF2}	c_{Ala}
1.06	1.06	1.67	2.47	0.91	1.30	0.27	0.57	1.39
1.08	1.02	2.05	2.78	0.97	5×10^{-4}	0.36	0.57	1.38
1.12	1.09	1.15	2.08	0.82	4×10^{-3}	3×10^{-4}	0.55	1.39

These minima are similar; the differences in SSQ_{lnPF} among them are slight and, with the exception of s_S and s_R, the parameters are comparable across all three cases. Nevertheless, a single minimum must be selected for further analysis. As this is the case, the minimum with $SSQ_{lnPF} = 1.06$ is preferred because it has the lowest SSQ_{lnPF} (trivial though the differences may be), and the s values in all cases, even for s_R, are of the order of magnitude expected for helix propensities (more or less close to 1). All of the following analysis is performed for the fit at this minimum under the assumption, consistent with the observations recorded in this section, that it is the global minimum.