Short Course:

Solution Structure Determination
In Organic Chemistry and Chemical Biology

Samuel H. Gellman
University of Wisconsin – Madison, USA

August 2012
University of Jyväskylä, Finland

Short Course:

Solution Structure Determination
In Organic Chemistry and Chemical Biology

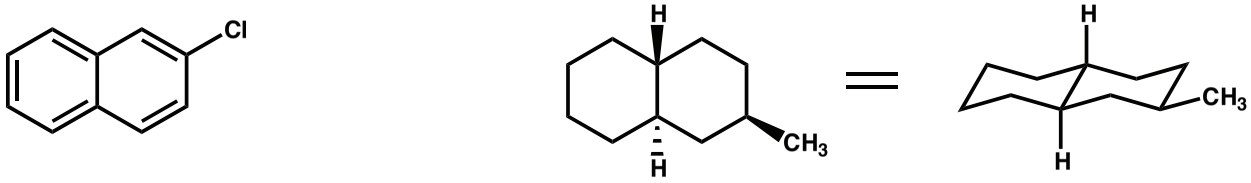
Samuel H. Gellman
University of Wisconsin – Madison, USA

August 2012
University of Jyväskylä, Finland

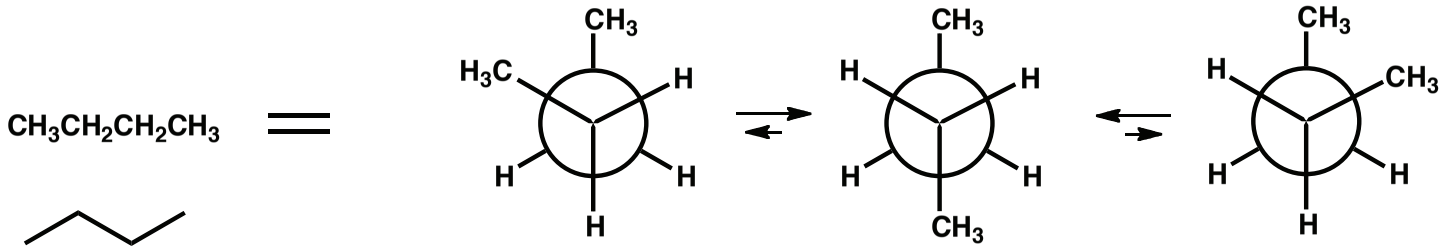
I. Introduction & structure mediated by hydrogen bonding

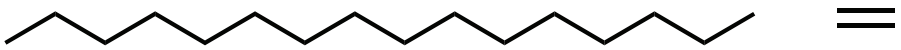
What is “solution structure”?

For many organic molecules, the solution structure is ‘obvious’ without any measurements:



For many other organic molecules, there is no single “solution structure”; instead, many conformations will be populated:





What is “solution structure”?

The phrase "solution structure determination" implies answering the question, "what is the three-dimensional shape of a given molecule in solution"?

For many molecules, this question suggests an overly simplistic starting premise, that there is just one three-dimensional shape in solution.

Many molecules are conformationally flexible, which is to say, they equilibrate among multiple conformations in solution. It is often these cases for which the question of "solution structure" is most interesting, particularly for the organic chemist or chemical biologist.

If a molecule is inherently rigid, then we can probably make a pretty good guess at the shape in solution from molecular mechanics calculations or physical models or crystal structures.

What is “solution structure”?

When flexible molecules equilibrate among different conformations in solution, then we have an opportunity to address a set of interrelated questions.

1. Why does the molecule adopt particular shapes in solution?

2. What intramolecular forces determine the shape preferences?

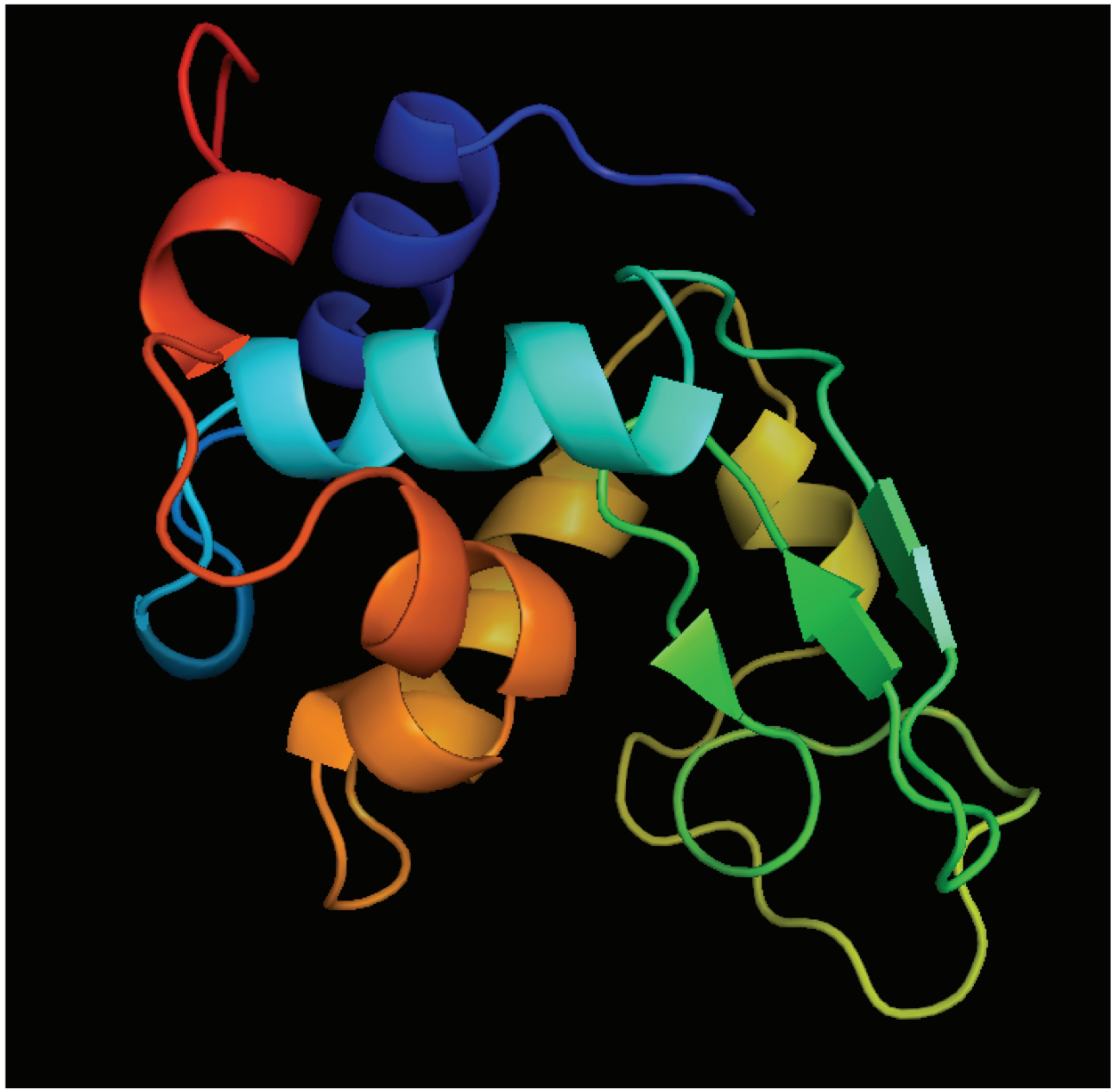
3. What shapes are preferred, and by how much (in terms of free energy)?

4. How does the environment (solvent, etc.) affect shape preferences?

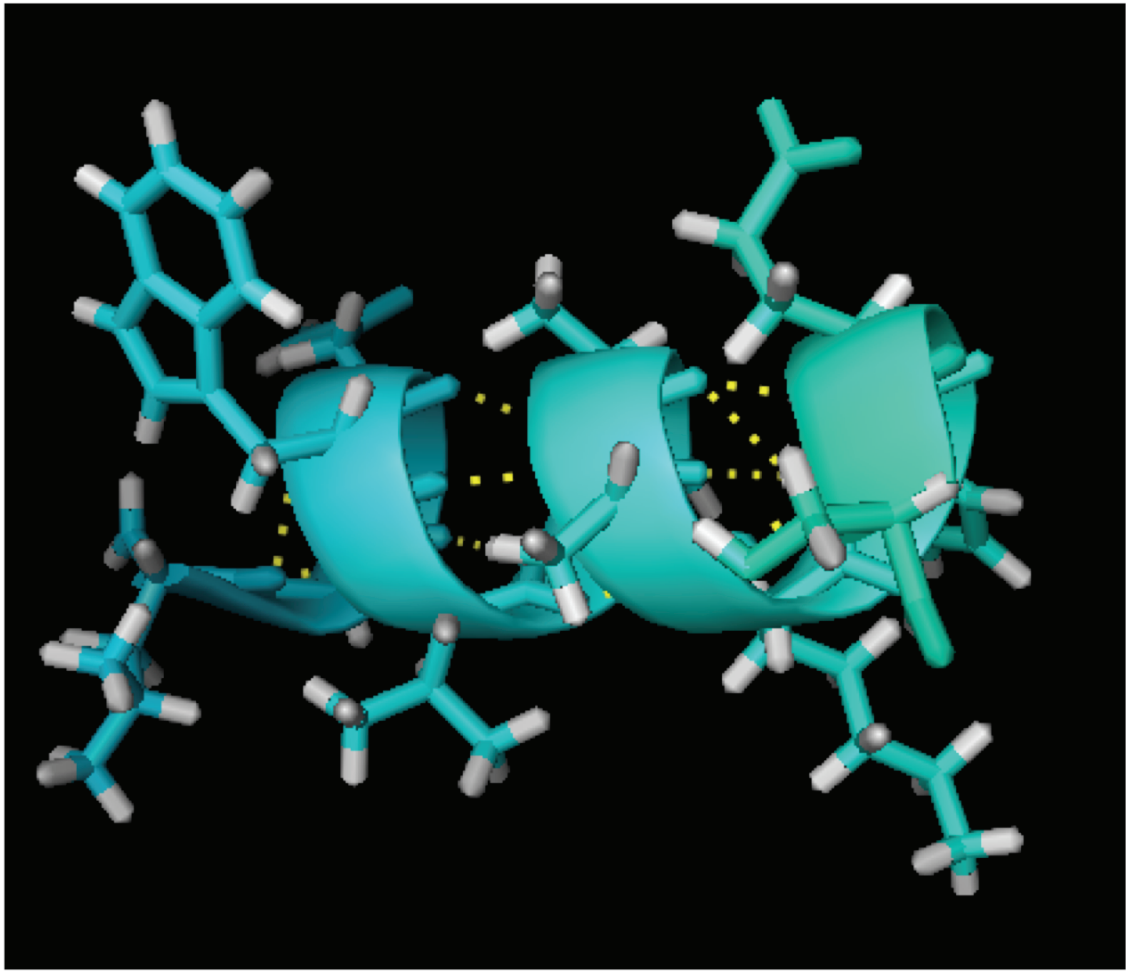
Most of the examples discussed in this short course will involve not only strategies for determining molecular shapes in solution, but also strategies for understanding the origins of those shapes. Often the questions are interwoven.

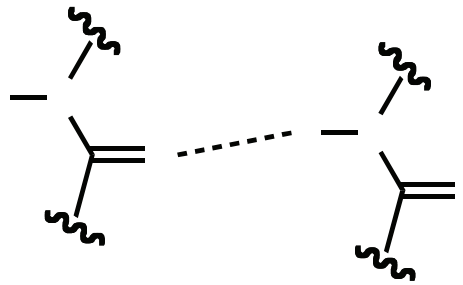
Hydrogen Bond-Mediated Structure

Crystal Structure of Lysozyme (PDB 3A8Z)

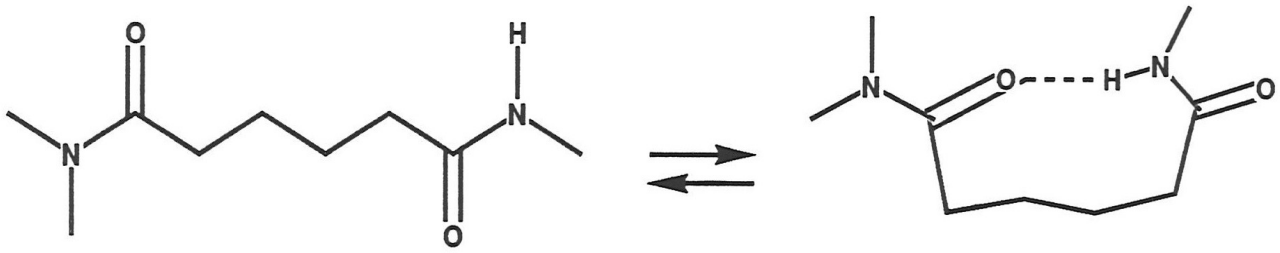


Crystal Structure of Lysozyme (PDB 3A8Z):
Close-up View of One α-Helix and H-bonds



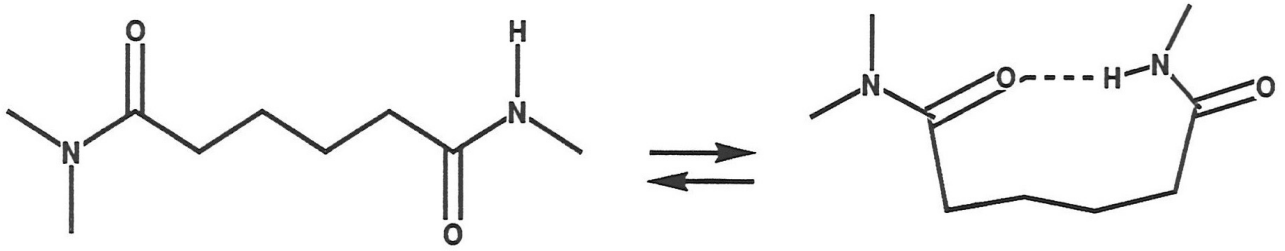


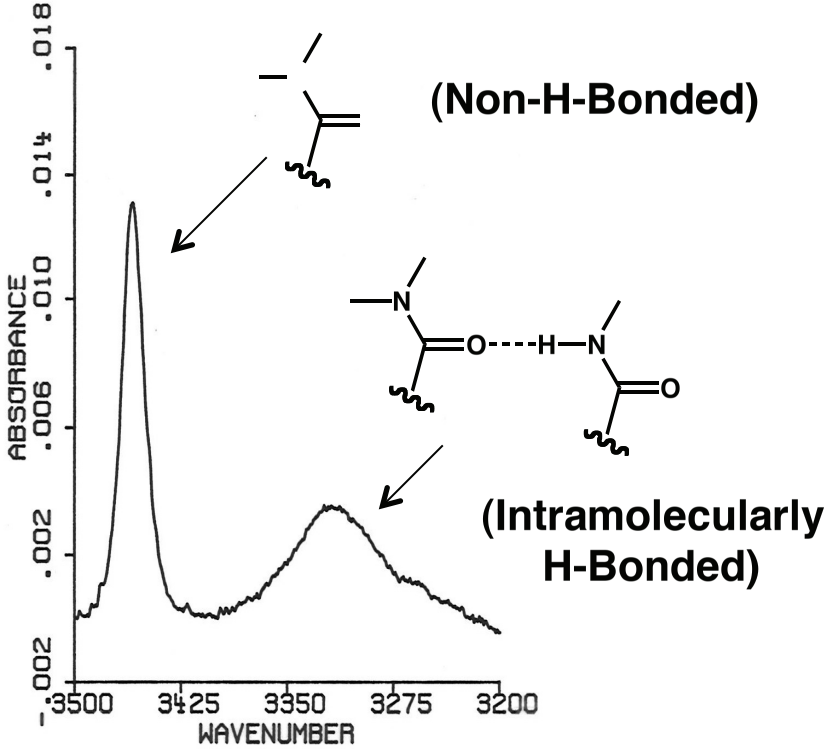
Hydrogen Bond-Mediated Folding in Simple Systems



Gellman, Dado, Liang, Adams ***J. Am. Chem. Soc.*** **113:1164 (1991)**

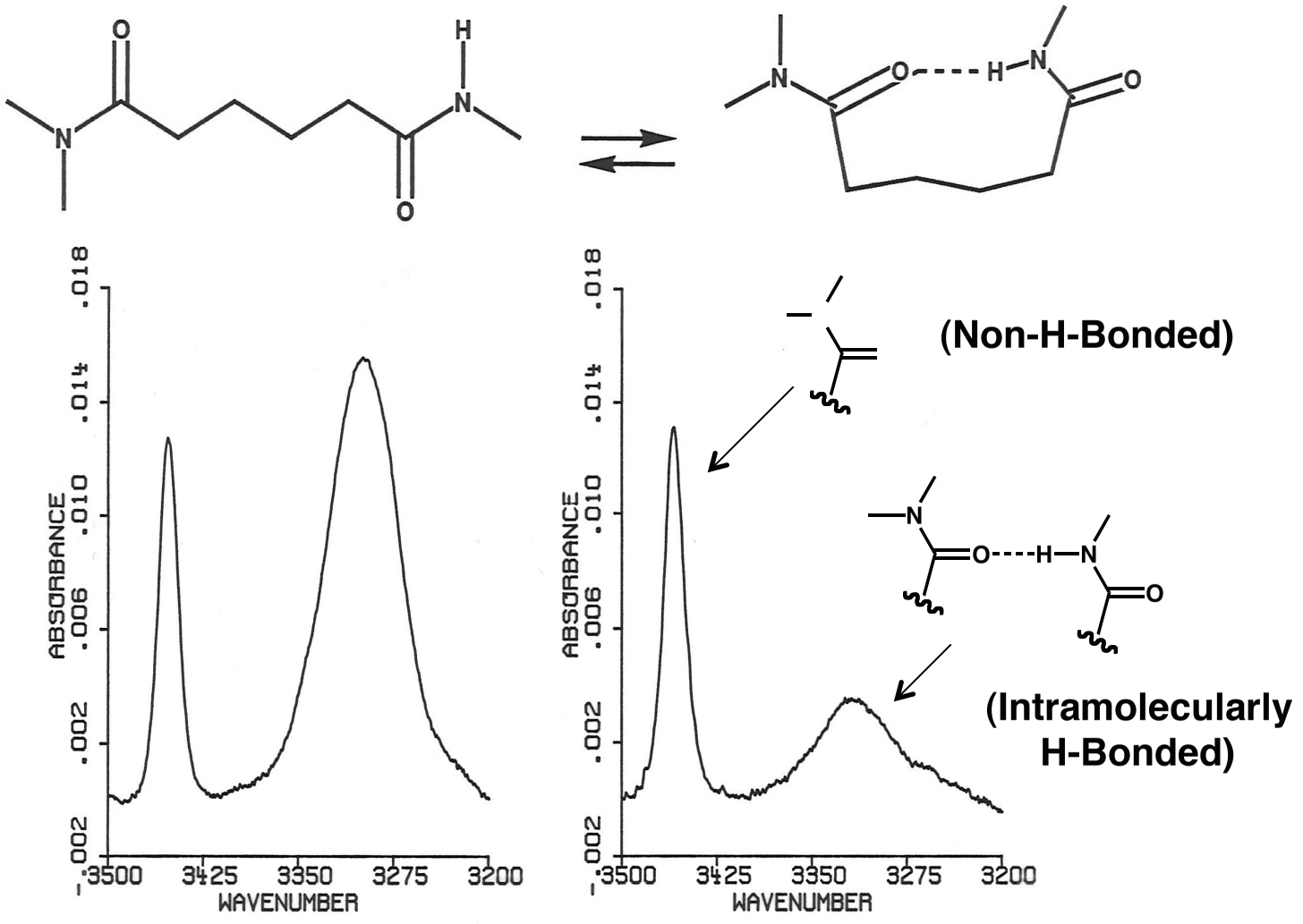
Infrared (IR) Spectroscopy (N-H Stretch region; 1 mM, CH_2Cl_2)



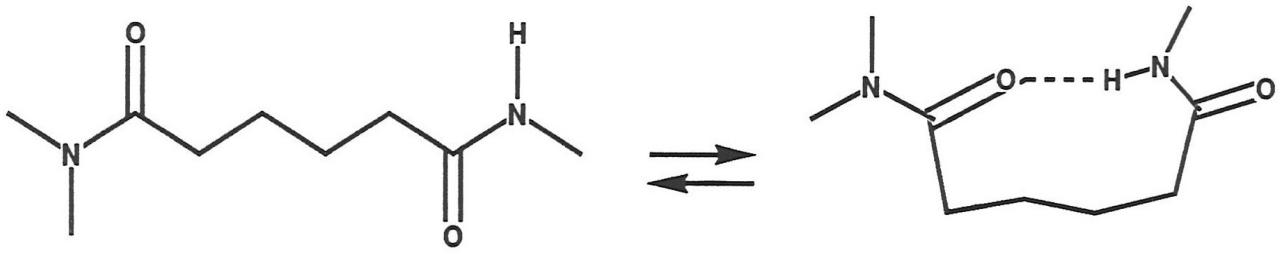


297 K

Infrared (IR) Spectroscopy (N-H Stretch region; 1 mM, CH_2Cl_2)



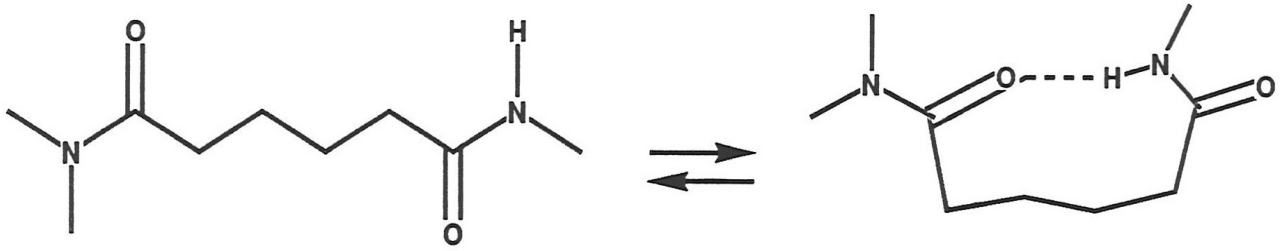
Infrared (IR) Spectroscopy (N-H Stretch region; 1 mM, CH_2Cl_2)



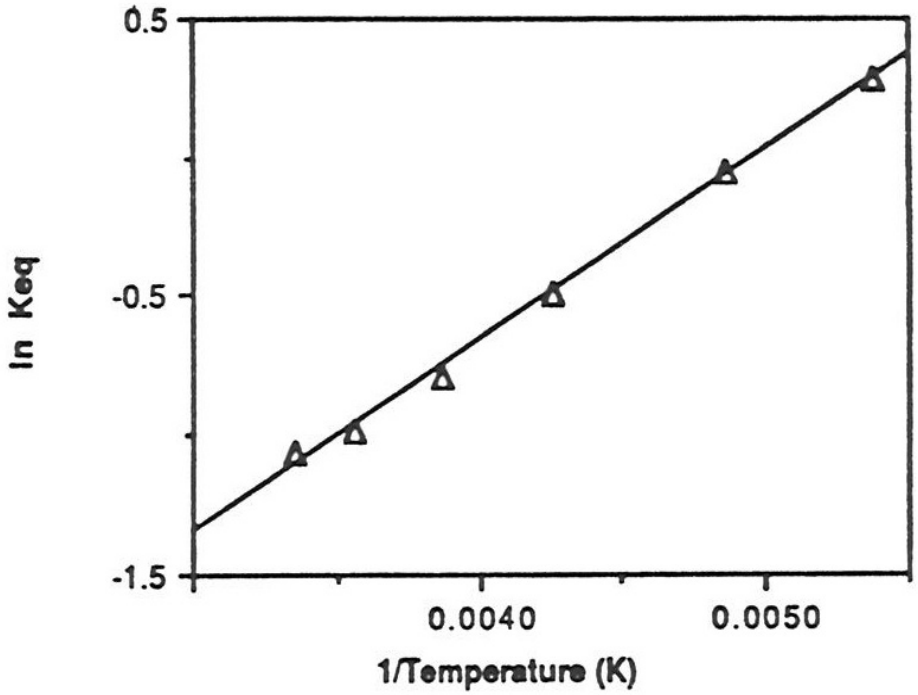
<u>Key Considerations</u>:

- **Experiments must be conducted at sufficient dilution to avoid *intermolecular* H-bonding.**
- **Distinct signals for non-H-bonded and internally H-bonded states; interchange is slow on the IR time scale.**
- **Population analysis (two-state; "unfolded" vs. "folded") is straightforward (extinction coefficient for non-H-bonded N-H stretch band).**

van't Hoff Analysis



Non-HB **HB**



$$K_{eq} = \frac{[HB]}{[non\text{-}HB]}$$

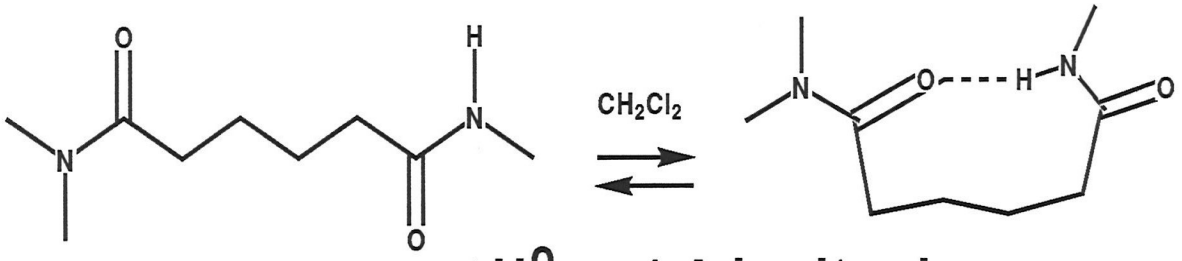
$$\ln K_{eq} = (-\Delta H^{o}/RT) + (\Delta S^{o}/R)$$

ΔH^{o} = -1.4 kcal/mol

ΔS^{o} = -6.8 e.u.

(CH_2Cl_2)

What do we learn?



ΔH^o = -1.4 kcal/mol

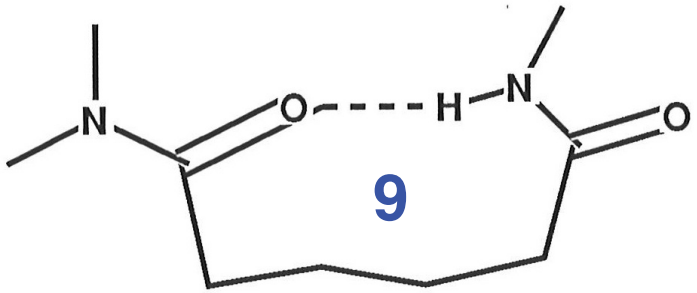
ΔS^o = -6.8 e.u.

ΔH^o: Competition between hydrogen bond energy in CH_2Cl_2 and internal strain in folded conformation(s).

ΔS^o: Cf. Page & Jencks *PNAS* 68:1678 (1971)
ca. 4.5 e.u. per single bond rotation,
for covalent ring closure.

⇨ Hydrogen bond formation is not as entropically costly as covalent bond formation.

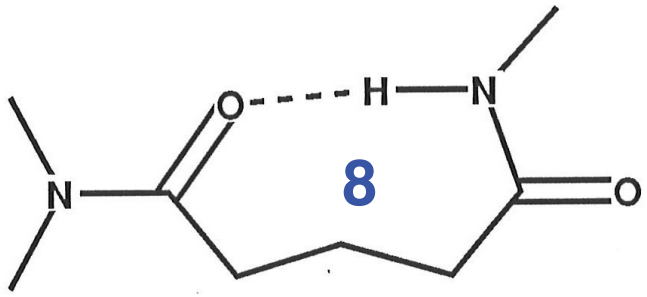
Ring Size Effects (CH_2Cl_2)



ΔH^o = -1.4 kcal/mol

ΔS^o = -6.8 e.u.

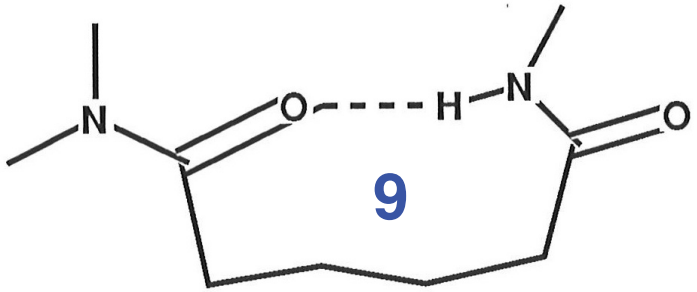
vs.



ΔH^o = -0.4 kcal/mol

ΔS^o = -3.3 e.u.

Ring Size Effects (CH_2Cl_2)

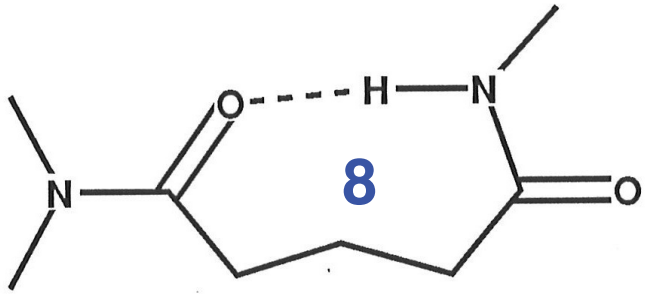


9

ΔH^o = -1.4 kcal/mol

ΔS^o = -6.8 e.u.

vs.



8

ΔH^o = -0.4 kcal/mol

ΔS^o = -3.3 e.u.

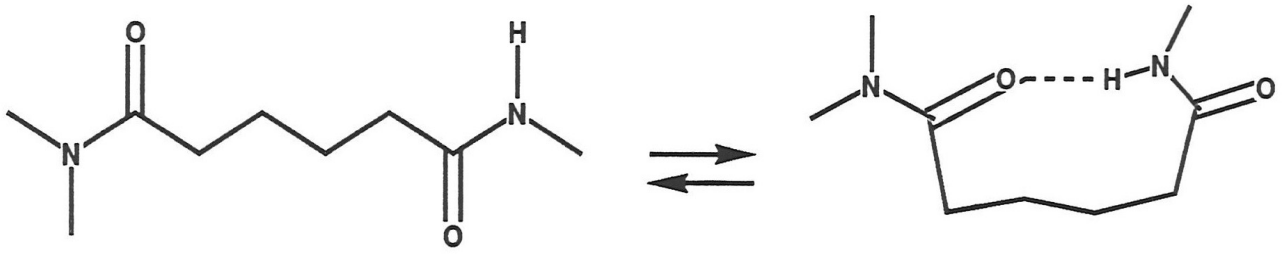
The impact of an H-bond on folding depends on context.

The effect of a single H-bond is limited.

Population Analysis by ^{1}H NMR (δNH)

Folding equilibrium is *rapid* on the NMR time scale.

δNH reflects a *weighted average* of contributing states.



δNH ~ 5.5 ppm (CD_2Cl_2)

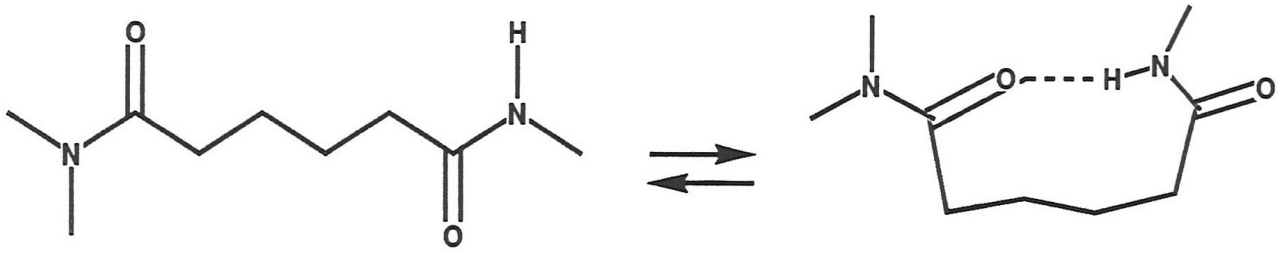
[Non-H-bonded limit]

δNH ~ 8.2 ppm (CD_2Cl_2)

[Fully H-bonded limit]

Population Analysis by ^{1}H NMR (δNH)

- **Folding equilibrium is *rapid* on the NMR time scale.**
- **δNH reflects a *weighted average* of contributing states.**



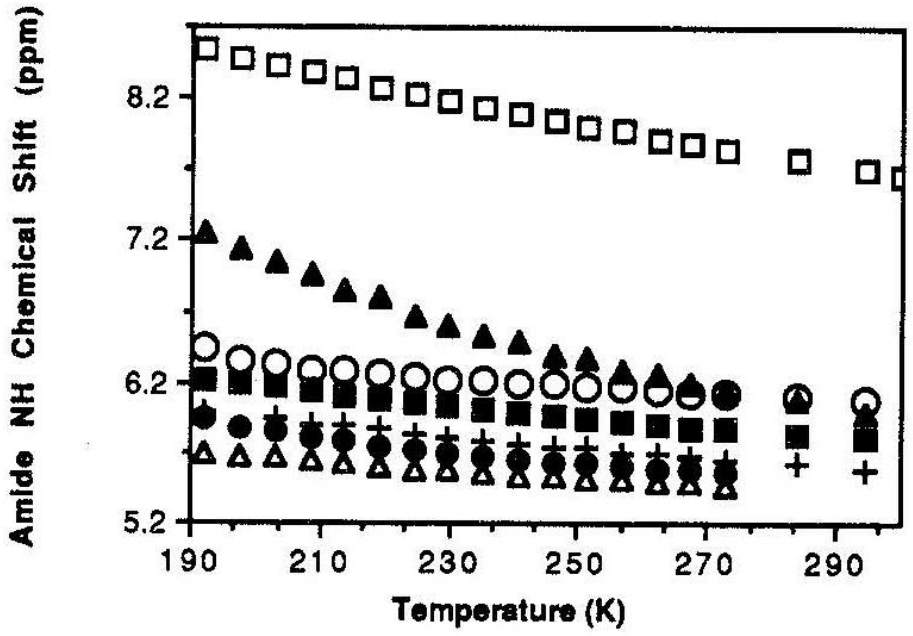
$$K_{eq} = \frac{\delta_{obs} - \delta_n}{\delta_b - \delta_{obs}}$$

δ_{obs} = δNH observed for diamide

δ_n = δNH for 'pure' non-H-bonded state

δ_b = δNH for 'pure' H-bonded state

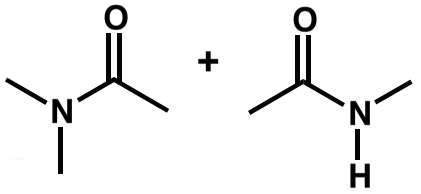
Population Analysis by ^{1}H NMR (δNH; CD_2Cl_2)



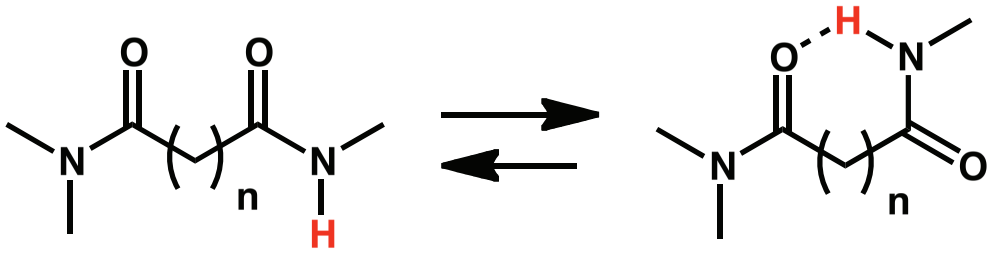
n = 1

n = 4 ▲

n = 2, 3, 5 & 6

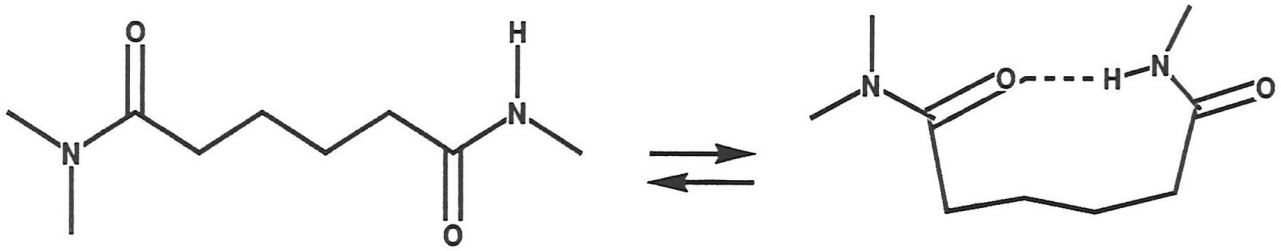


(control: δ_n) △



n = 1-6

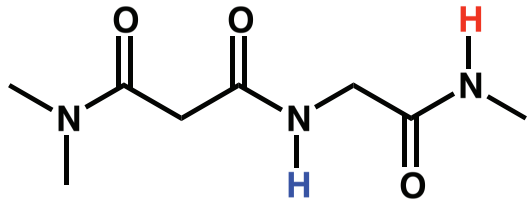
van't Hoff Analysis: IR vs. NMR (CH_2Cl_2)



IR	**NMR**
ΔH^o = -1.4 kcal/mol	**ΔH^o = -1.5 kcal/mol**
ΔS^o = -6.8 e.u.	**ΔS^o = -8.0 e.u.**

Gellman, Dado, Liang, Adams *J. Am. Chem. Soc.* 113:1164 (1991)

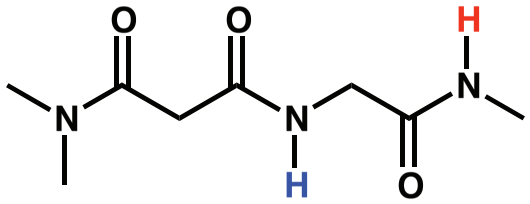
A Slightly More Complex System…



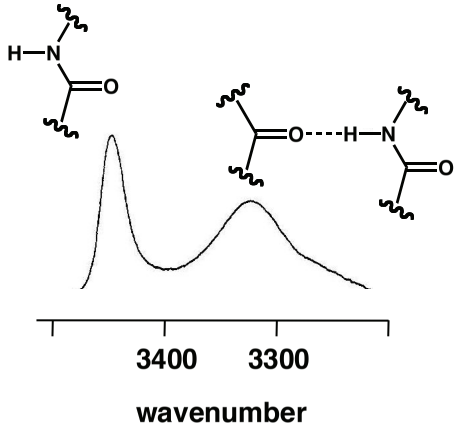
Gellman, Adams, Dado, *J. Am. Chem. Soc.* 112:460 (1990)

Dado & Gellman, *J. Am. Chem. Soc.* 115:4228 (1993)

A Slightly More Complex System…



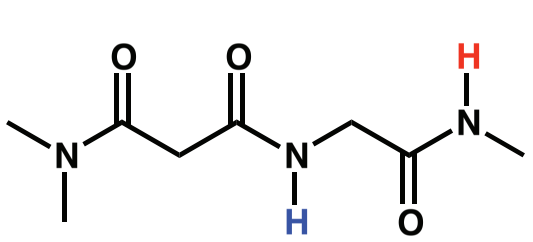
(1 mM, CH_2Cl_2)



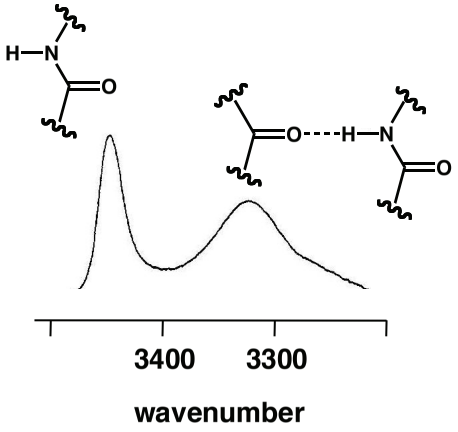
IR (N-H stretch)

Problem: Band Assignment

A Slightly More Complex System…



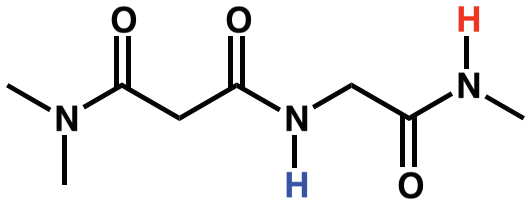
(1 mM, CH_2Cl_2)



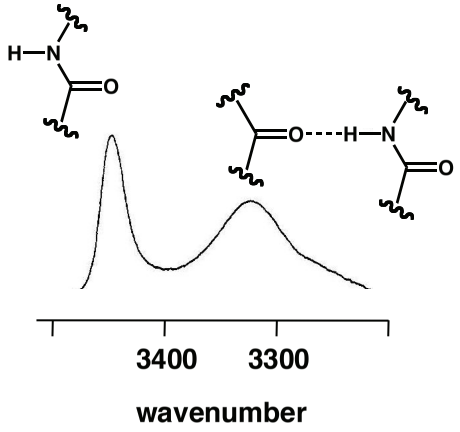
IR (N-H stretch)

Problem: Band Assignment

A Slightly More Complex System…

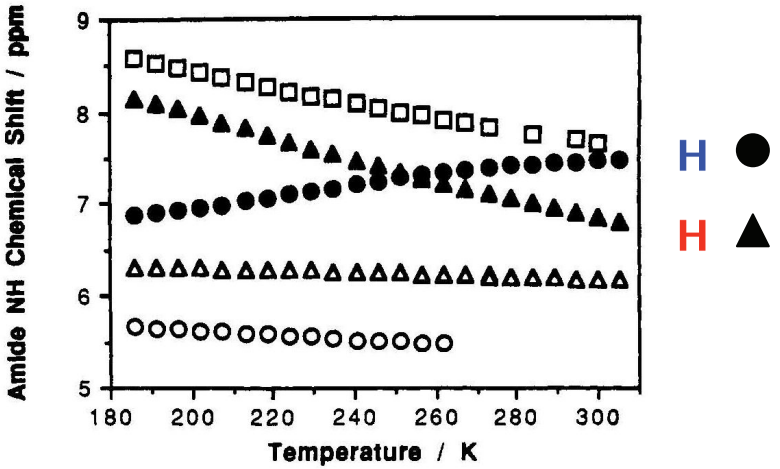


(1 mM, CH_2Cl_2)



IR (N-H stretch)

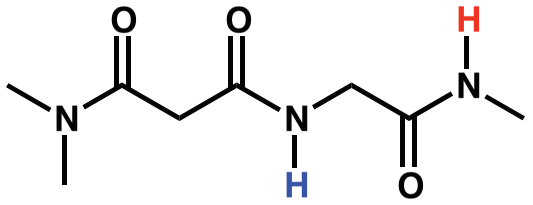
Problem: Band Assignment



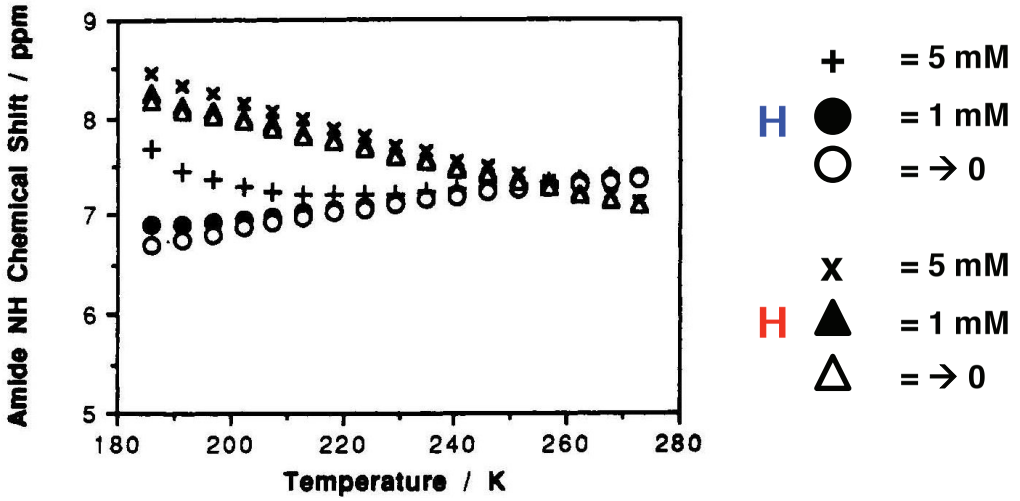
NMR (δNH)

Problem: Population Averaging

Control: Avoid Self-Association

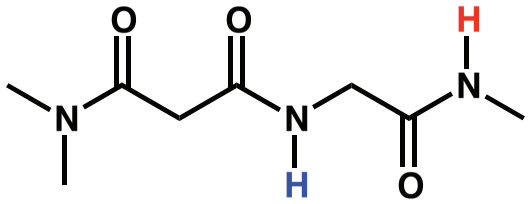


(CD_2Cl_2)

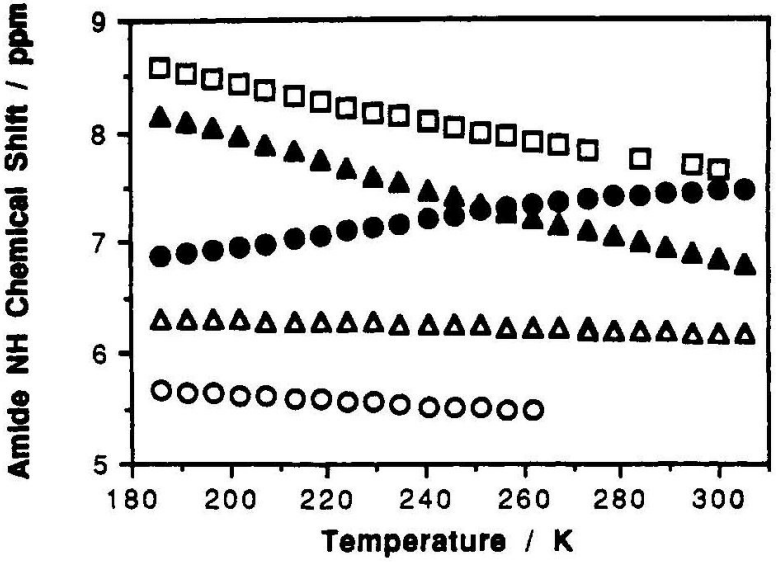


NMR (δNH)

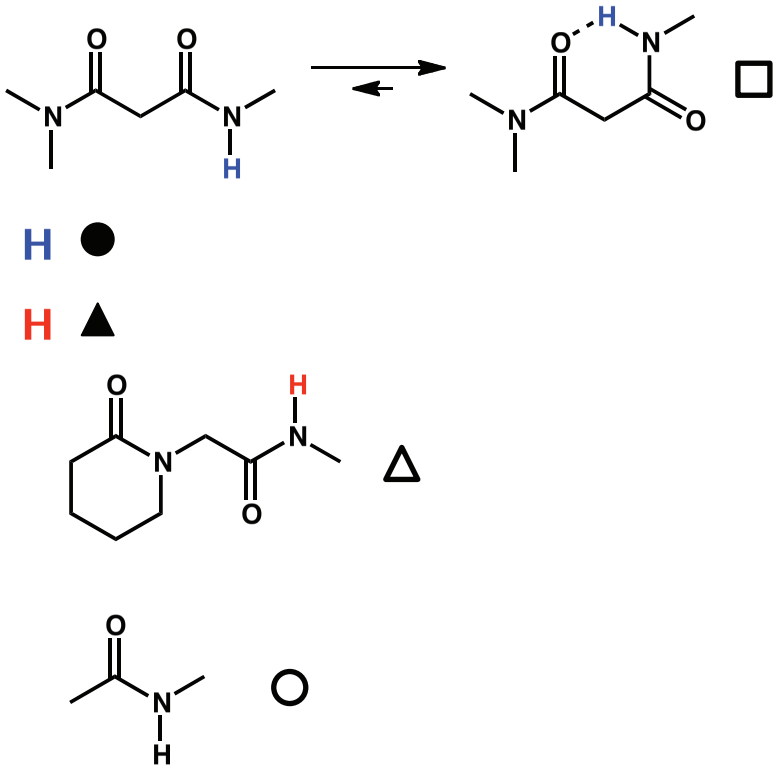
Reference Compounds



(1 mM, CD_2Cl_2)

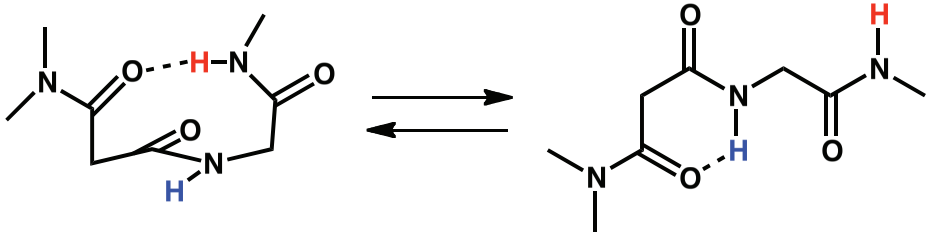


NMR (δNH)



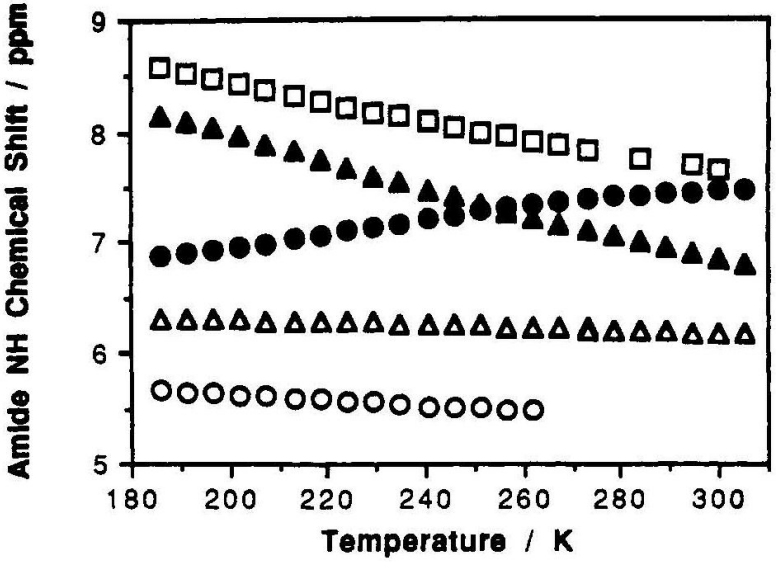
What is the “Solution Structure”?

Hypothesis:

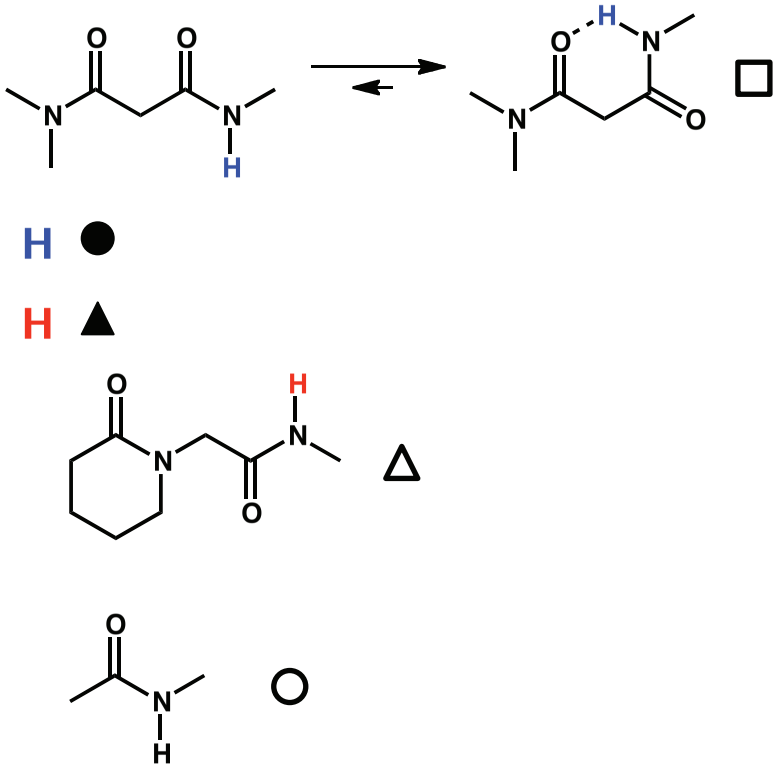


Favored at low temperature

Favored at high temperature

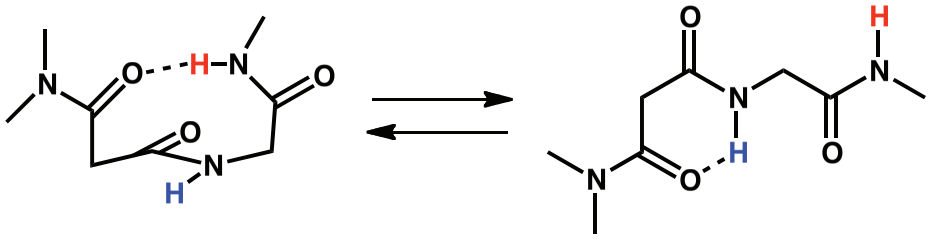


NMR (δNH)



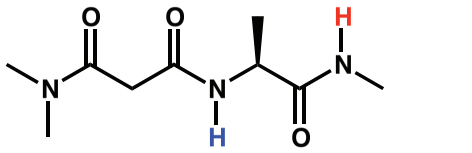
Test Hypothesis: ‘Steric Labeling’ in IR

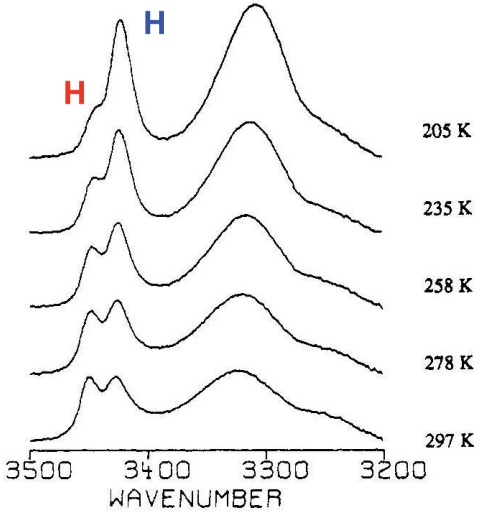
Hypothesis:

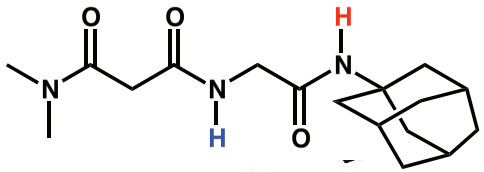


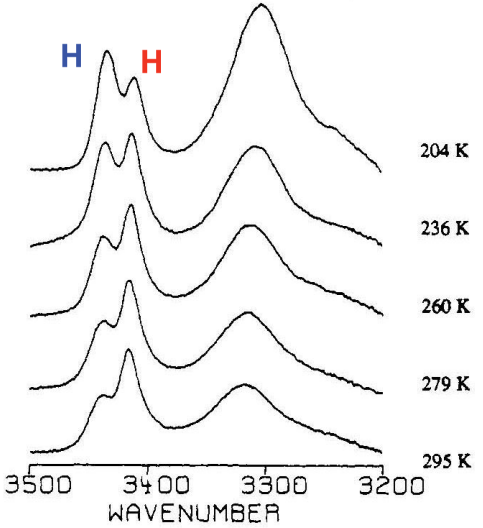
Favored at low temperature

Favored at high temperature



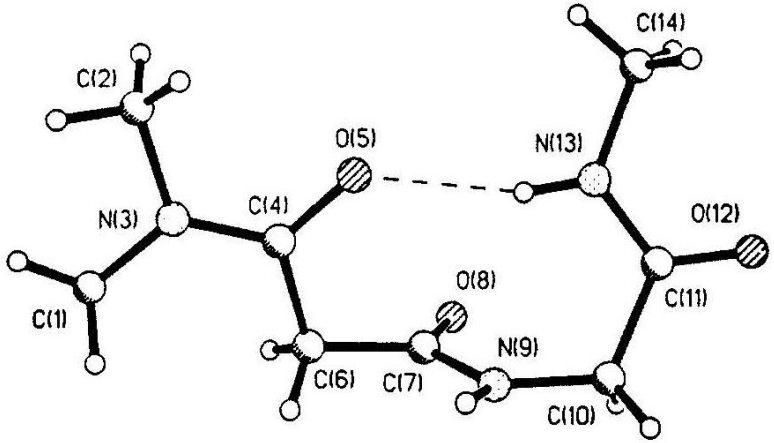


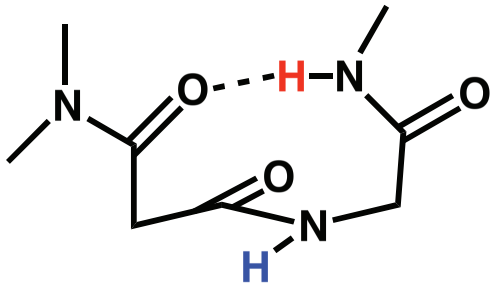




Adjacent branch point shifts N-H stretch to lower wavenumber:
Boussard, Marraud, Aubry *Biopolymers* 181297 (1979)

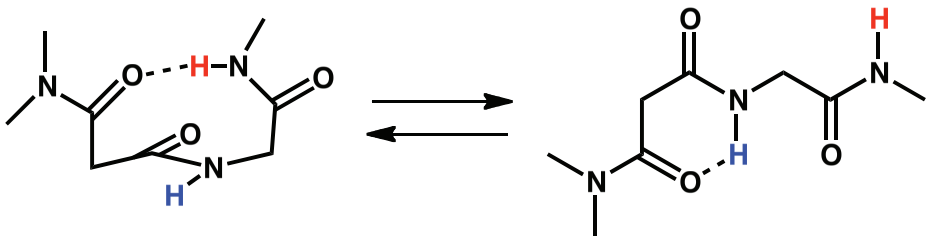
Support from Crystallography





Dado, Desper, Gellman *J. Am. Chem. Soc.* 112:8630 (1990)

What do we learn?



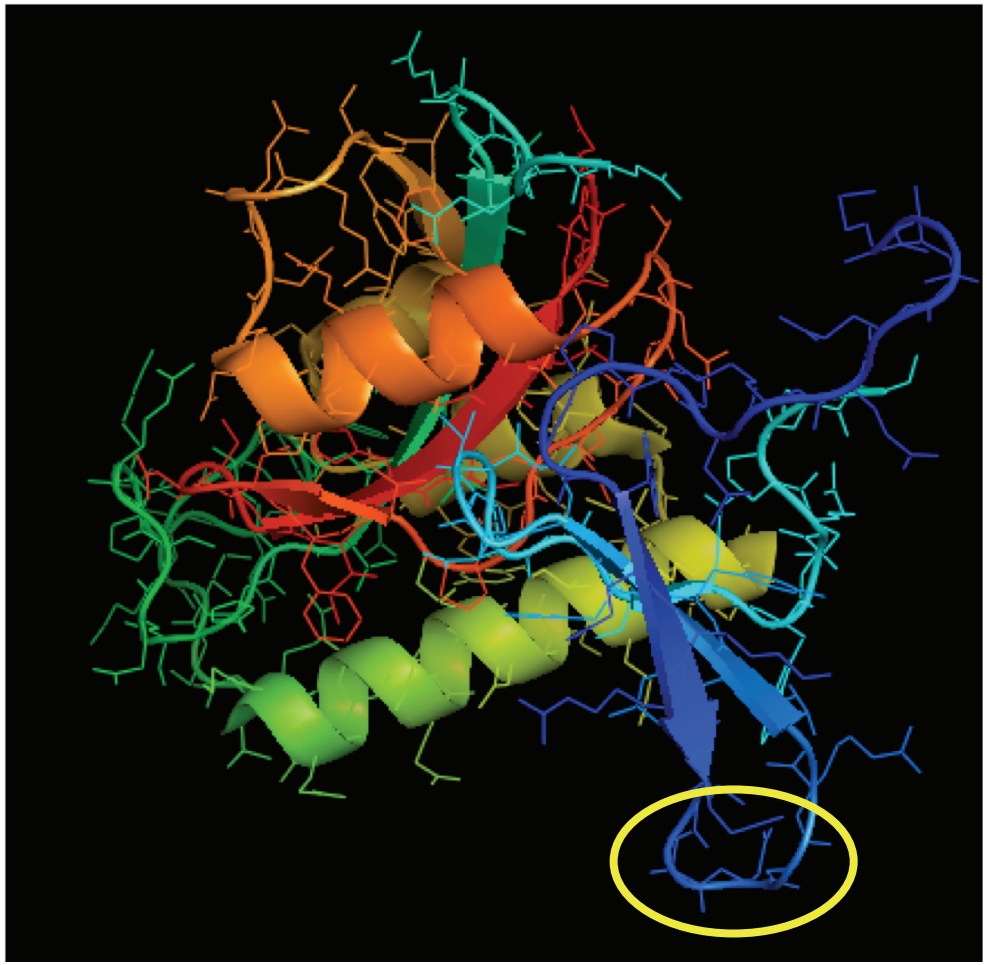
Favored at low temperature

Favored at high temperature

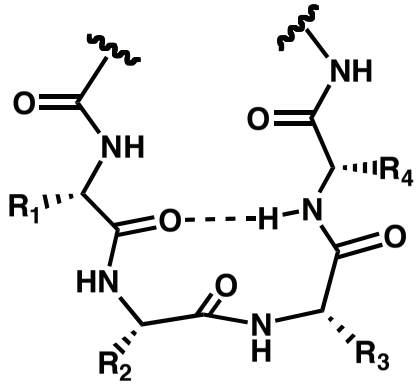
<u>Delicate thermodynamic balance</u>:

- **9-Ring H-bonded state is favored by ΔH; superior H-bond geometry?**
- **6-Ring H-bonded state is favored by ΔS; less ordering in small ring?**

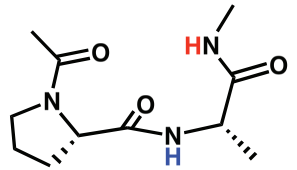
β-Turns: Common Local Structures in Proteins



Pin WW Domain-Rotamase (PDB 2F21)



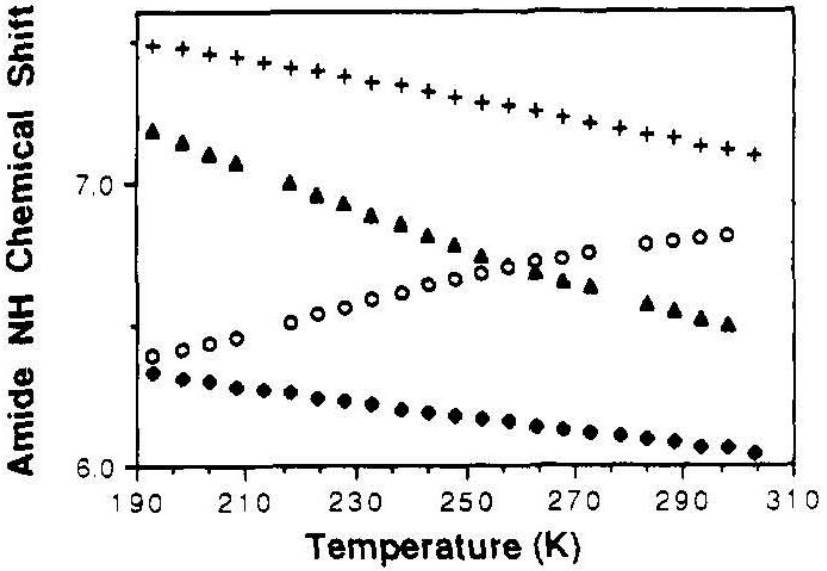
Minimal β-Turn Peptide (Ac-Pro-Ala-NHMe)



Liang, Rito, Gellman *J. Am. Chem. Soc.* 114:4440 (1992)

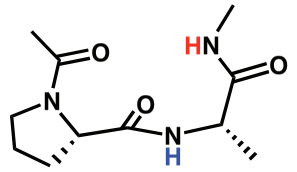
Minimal β-Turn Peptide (Ac-Pro-Ala-NHMe)
Variable Temperature ^{1}H NMR

(1 mM, CD_2Cl_2)



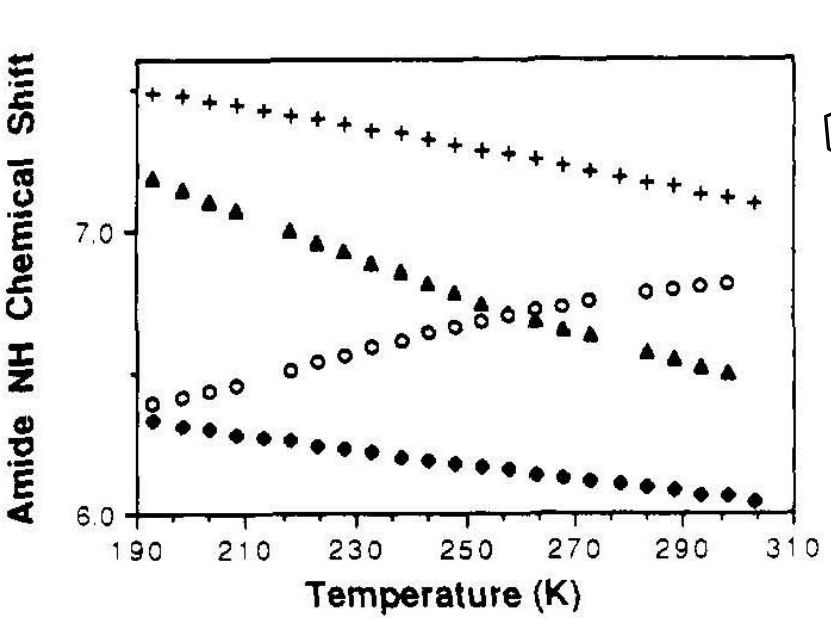
H ○

H ▲



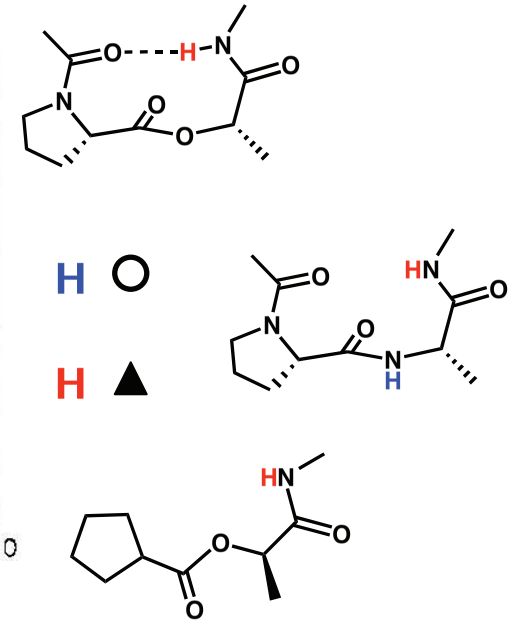
Minimal β-Turn Peptide (Ac-Pro-Ala-NHMe) Reference Compounds

(1 mM, CD_2Cl_2)



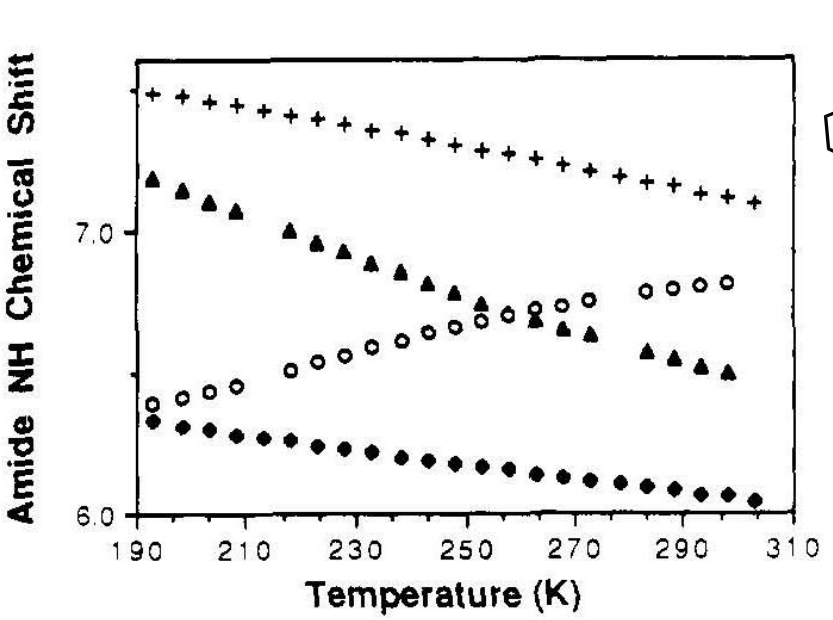
H ○

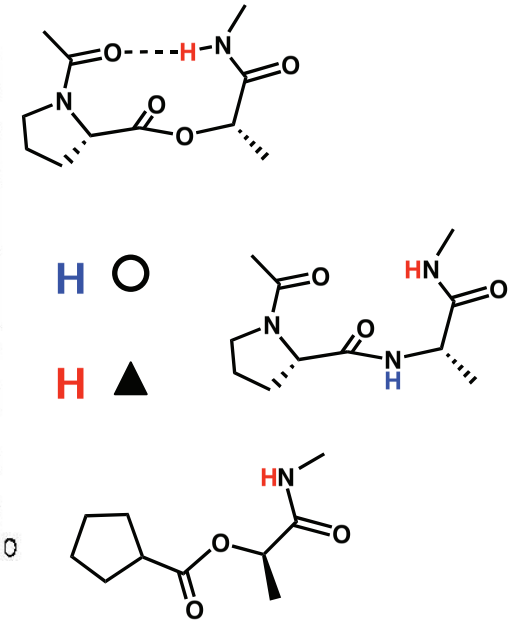
H ▲

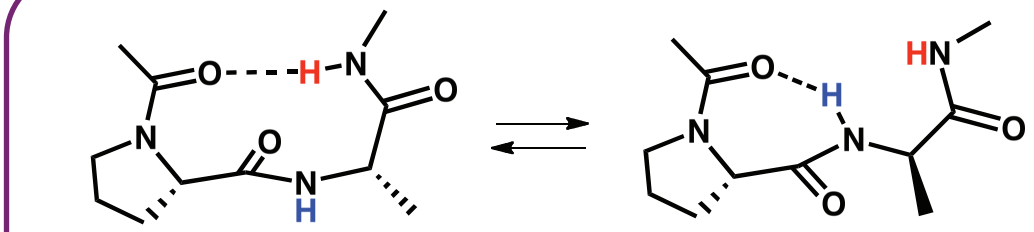


Minimal β-Turn Peptide (Ac-Pro-Ala-NHMe) Conformational Hypothesis

(1 mM, CD_2Cl_2)





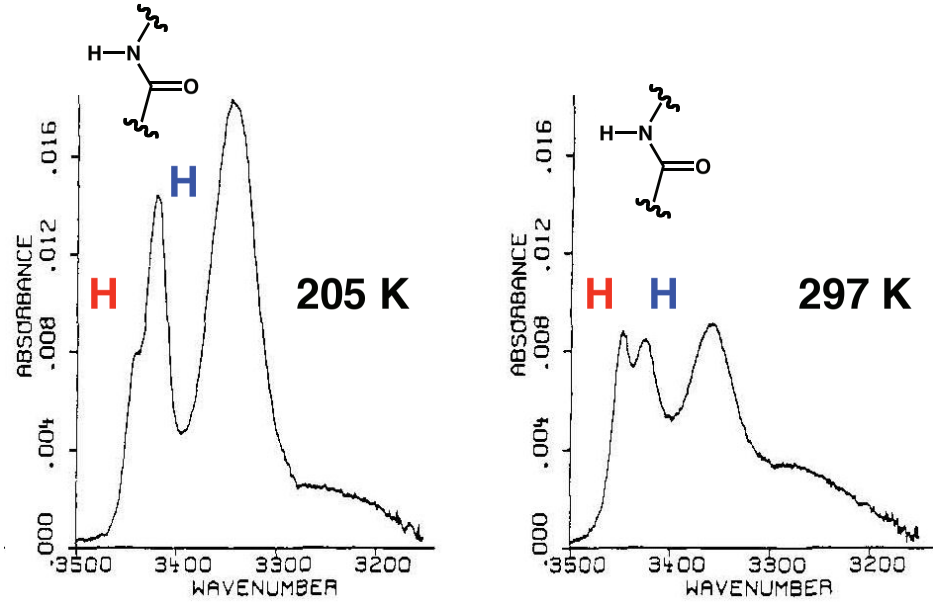


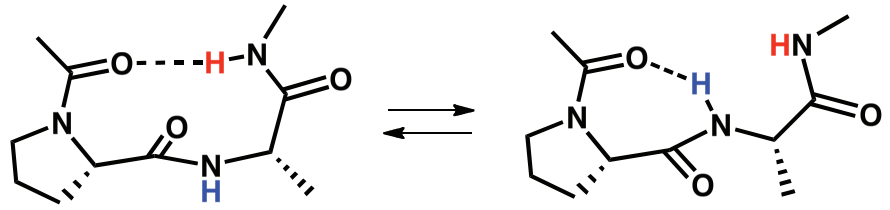
Favored at low temperature

Favored at high temperature

Minimal β-Turn Peptide (Ac-Pro-Ala-NHMe) Support from Variable-Temperature IR (Recall 'Steric Labelling')

(1 mM, CH_2Cl_2)

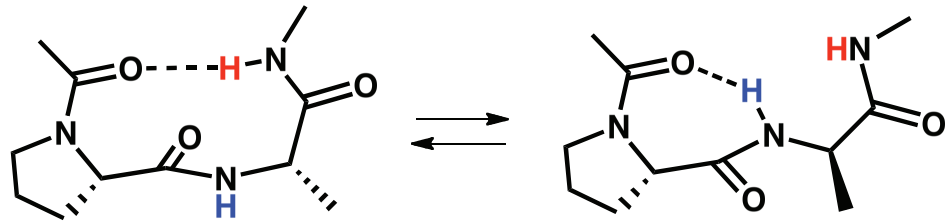




Favored at low temperature

Favored at high temperature

What do we learn?



Favored at low temperature

Favored at high temperature

<u>Delicate thermodynamic balance</u>:

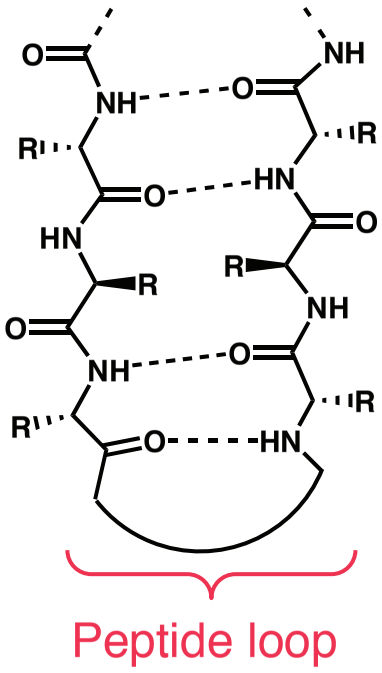
- **β-Turn conformation is favored by ΔH, by ~1.4 kcal/mol (CH_2Cl_2).**
- **β-Turn conformation is disfavored by ΔS, relative to less ordered states.**

Apply Small Peptide Folding Tools to Important Questions

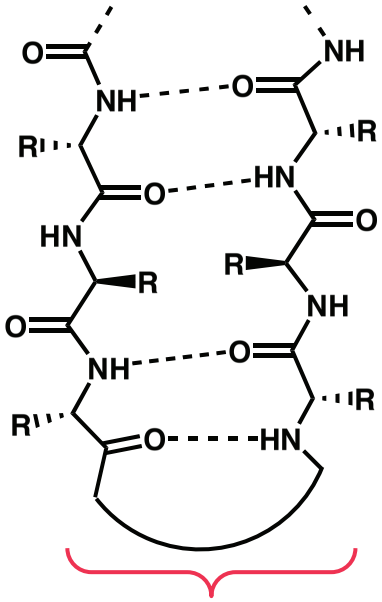
1. **Relationship between β-turn stereochemistry and antiparallel β-sheet nucleation?**

2. **Secondary H-bonding interactions in peptides and related molecules?**

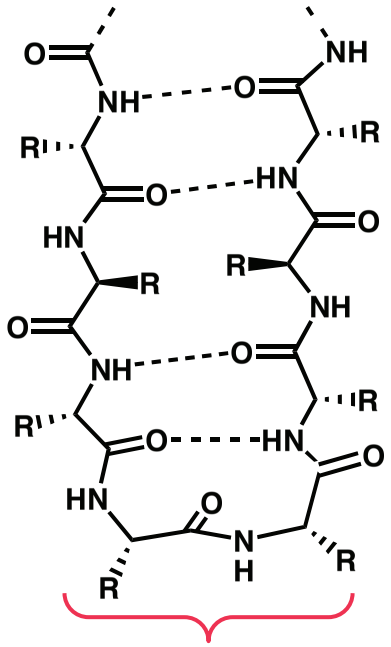
β-Hairpins: Prototype for Minimum-Sized Autonomously Folding β-Sheets (Antiparallel)



β-Hairpins: Prototype for Minimum-Sized Autonomously Folding β-Sheets (Antiparallel)

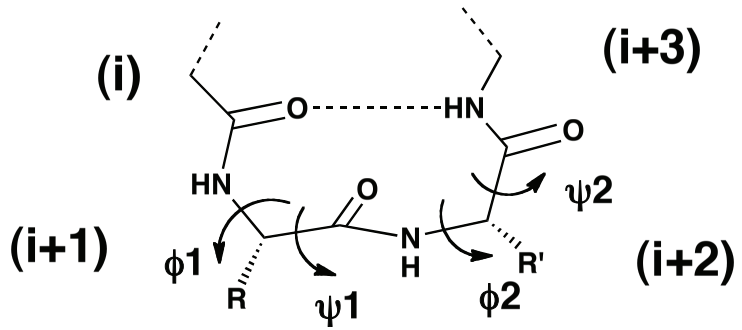


Peptide loop



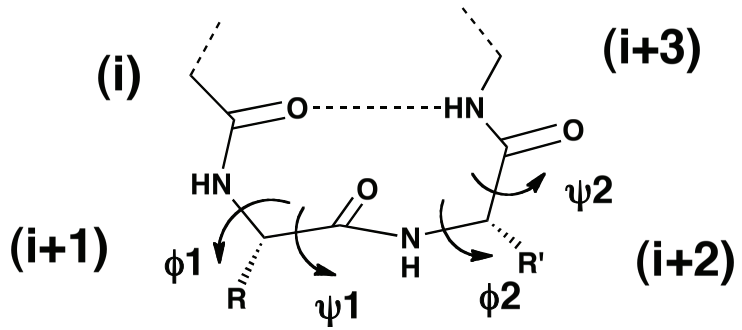
Most common small loop
= β-turn (two residues)

β-Turn Definitions



Four-Residue Unit; reversal of polypeptide backbone direction.

β-Turn Definitions

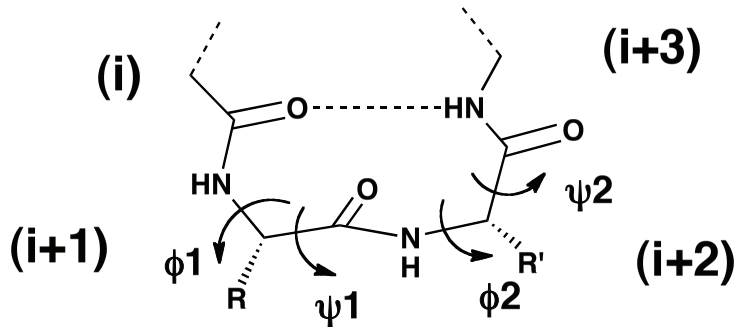


Common	φ1	ψ1	φ2	ψ2
Type I	-60	-30	-90	0
Type II	-60	120	80	0
etc.				

See: Hutchinson and Thornton, *Protein Science* 1994, *3*, 2207.

Note: 10-membered ring H-bond not obligatory.

β-Turn Definitions

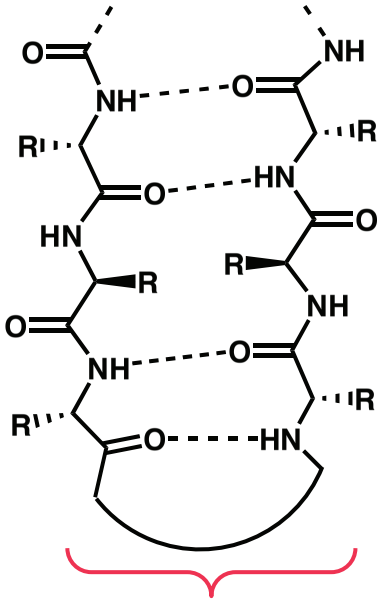


Common	φ1	ψ1	φ2	ψ2	
Type I	-60	-30	-90	0	
Type II	-60	120	80	0	
etc.					
"Mirror Image"	**φ1**	**ψ1**	**φ2**	**ψ2**	
Type I'	60	30	90	0	(Rare)
Type II'	60	-120	-80	0	
etc.					

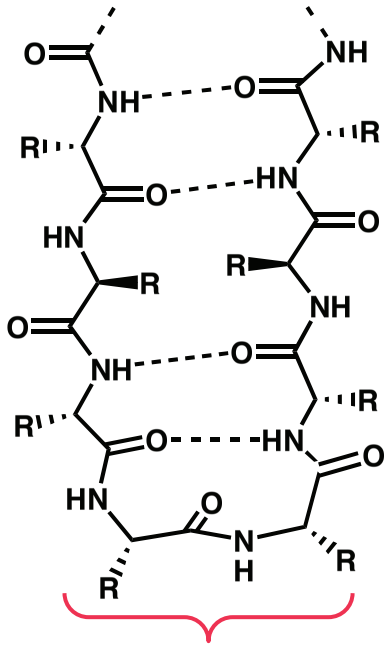
See: Hutchinson and Thornton, *Protein Science* 1994, *3*, 2207.

Note: 10-membered ring H-bond not obligatory.

β-Hairpins: Prototype for Minimum-Sized Autonomously Folding β-Sheets (Antiparallel)



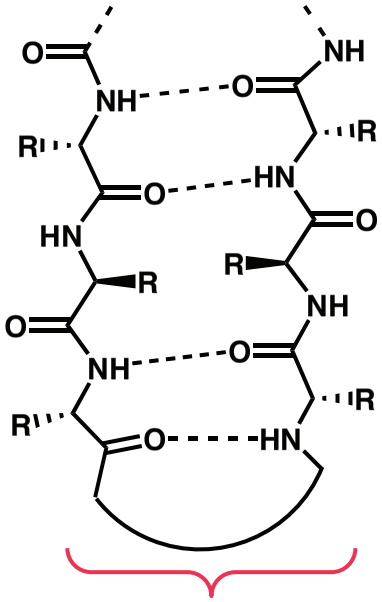
Peptide loop



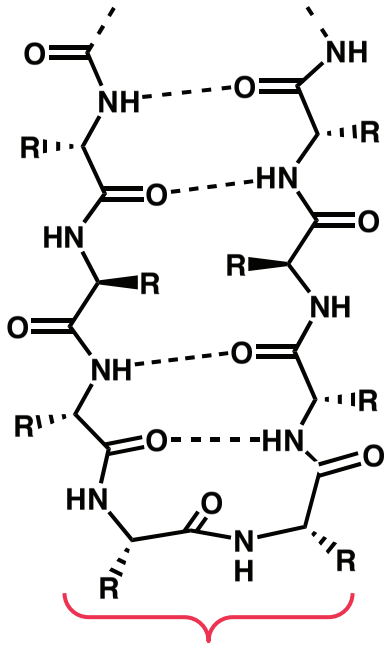
Most common small loop
= β-turn (two residues)

How do we specify the size and position of the loop in an autonomously foldling β-hairpin? (Two residues ideal.)

β-Hairpins: Prototype for Minimum-Sized Autonomously Folding β-Sheets (Antiparallel)



Peptide loop

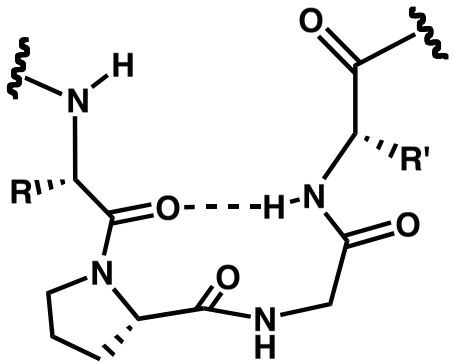


Most common small loop
= β-turn (two residues)

How do we specify the size and position of the loop in an autonomously foldling β-hairpin? (Two residues ideal.)

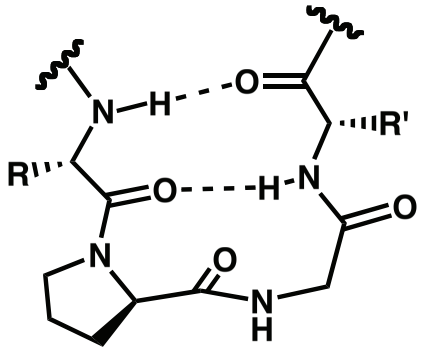
Importance of Type I' and Type II' ("mirror image") β-turns:
Wilmot & Thornton *J. Mol. Biol.* 1988, *203*, 221.

Hypothesis: Use of a D-residue in the Loop to Promote β-Hairpin Formation by L-residues in the Strands



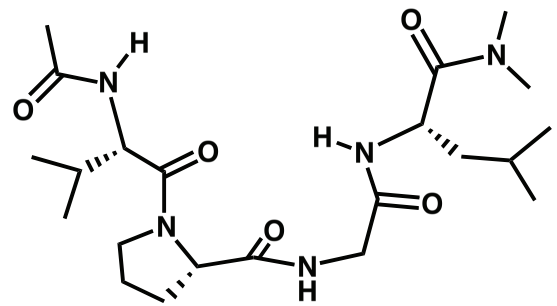
L-Pro

vs.



D-Pro

Minimal β-Hairpin Design

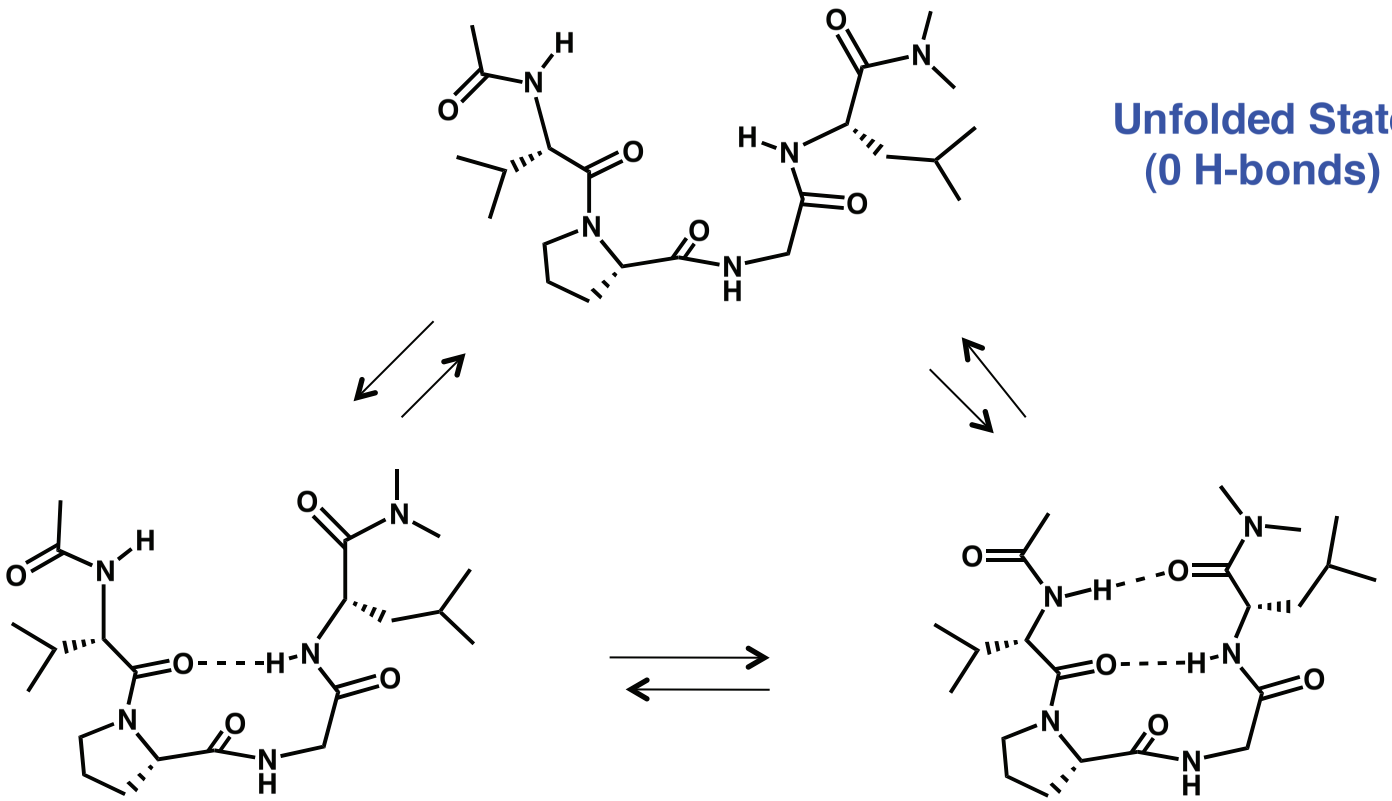


Tetrapeptide:

Ac-L-Val-L-Pro-Gly-L-Leu-NMe_2

Haque, Little, Gellman ***J. Am. Chem. Soc.*** **118:6975 (1996)**

Minimal β-Hairpin Design

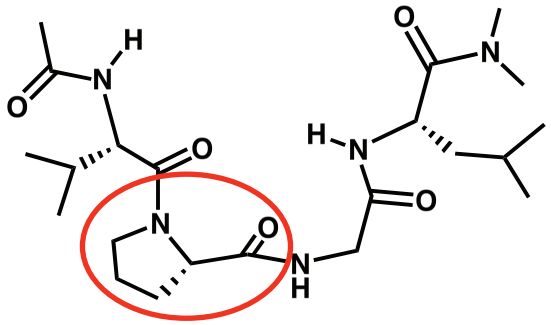


Unfolded State
(0 H-bonds)

β-Turn State
(1 H-bond; 'inner')

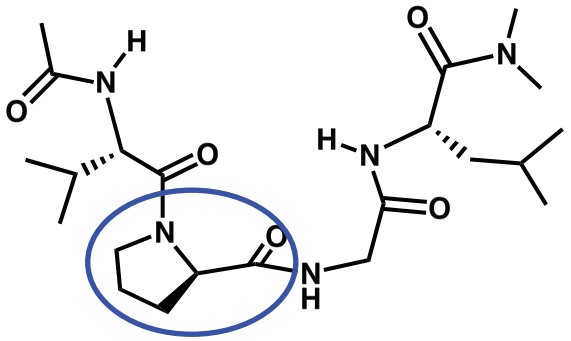
β-Hairpin State
(2 H-bonds)

Minimal β-Hairpin Design



Ac-L-Val-L-Pro-Gly-L-Leu-NMe$_2$

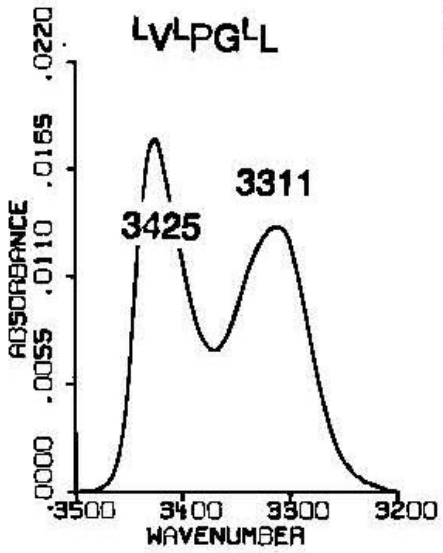
vs.

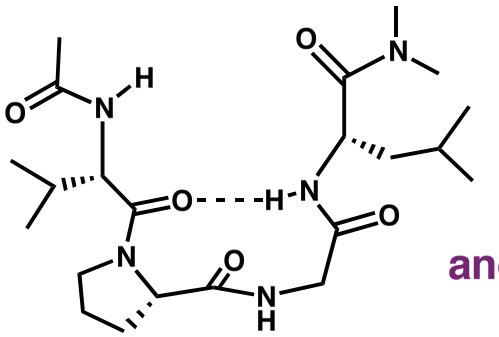


Ac-L-Val-D-Pro-Gly-L-Leu-NMe$_2$

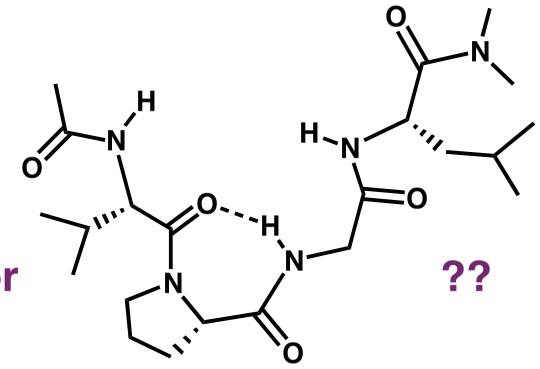
D-Pro will favor 'mirror image' β-turn.

Test Hypothesis: IR (N-H stretch; 1 mM CH_2Cl_2)



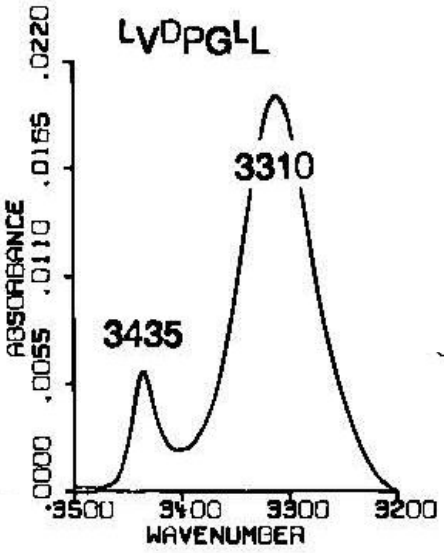


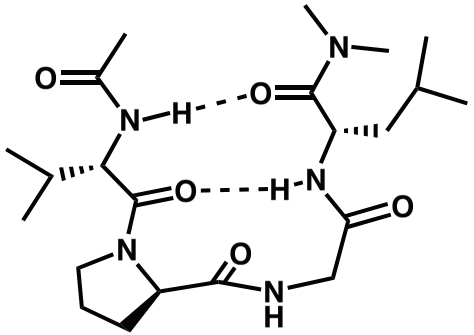
and/or



??

L-Pro

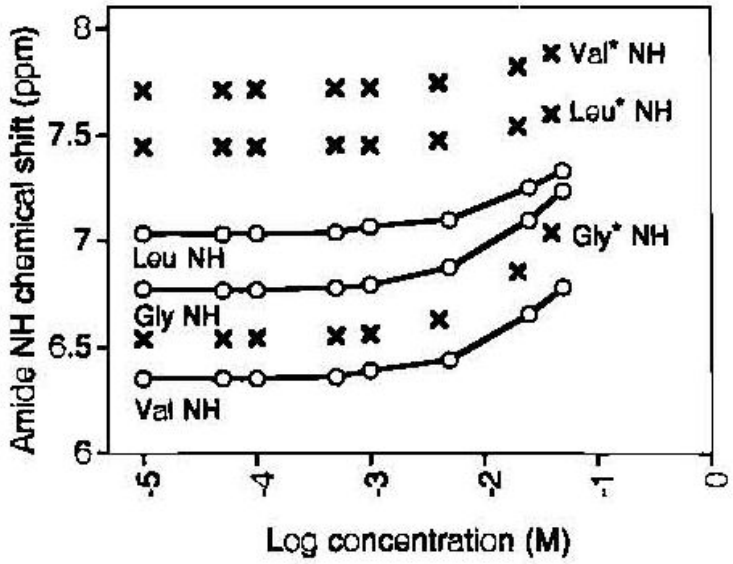




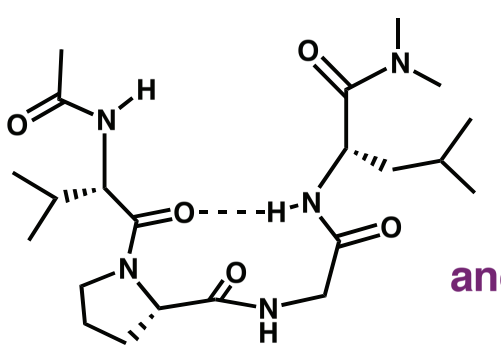
D-Pro

β-Hairpin State Dominant?

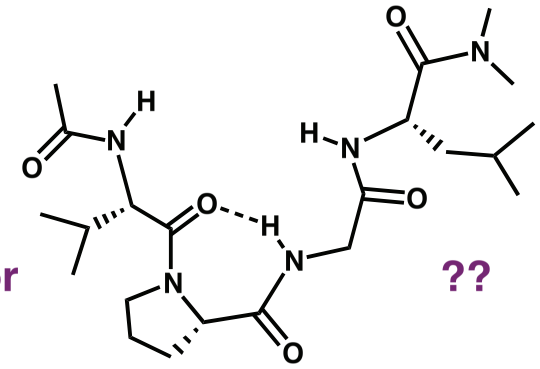
Test Hypothesis: ^{1}H NMR (CD_2Cl_2)



≤ 1 mM to avoid aggregation

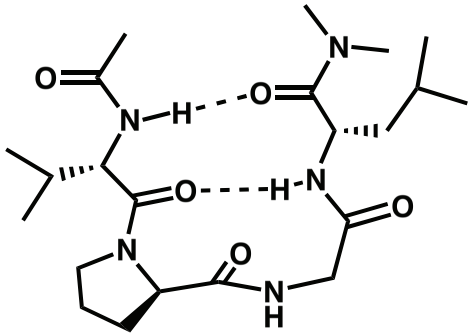


and/or



??

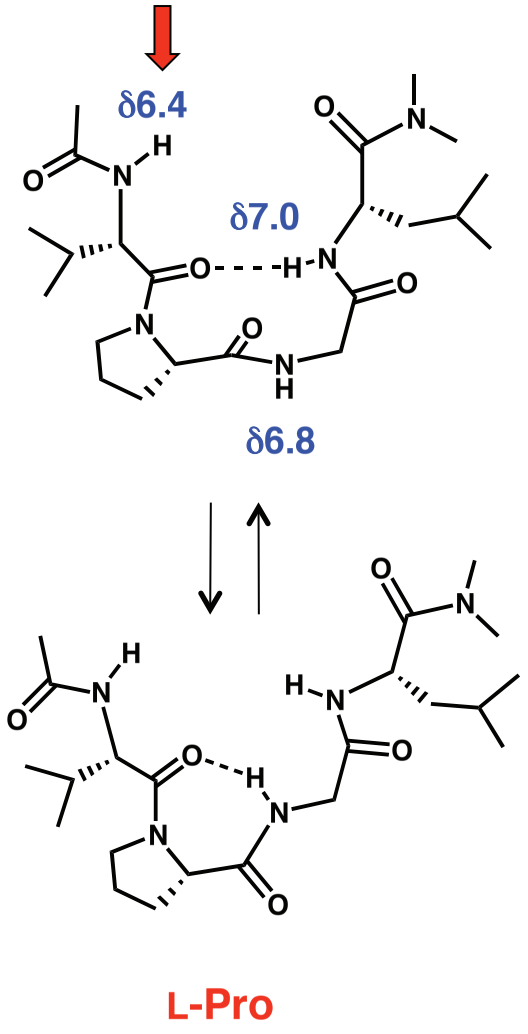
L-Pro



D-Pro

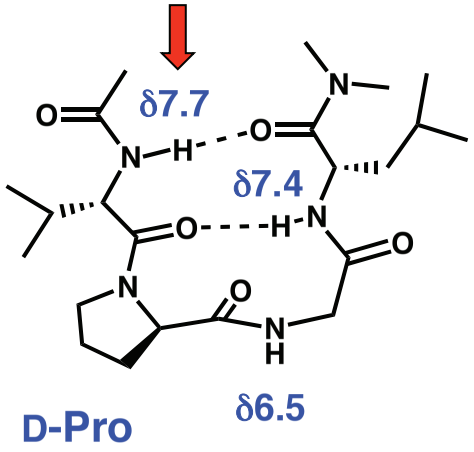
β-Hairpin State Dominant?

Test Hypothesis: ^{1}H NMR (CD_2Cl_2)



L-Pro

vs.

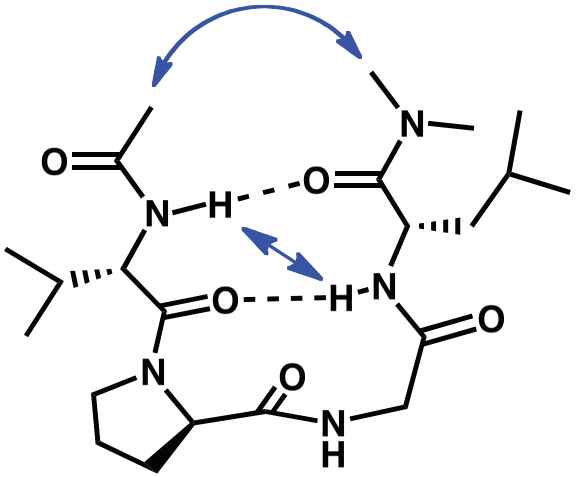


D-Pro

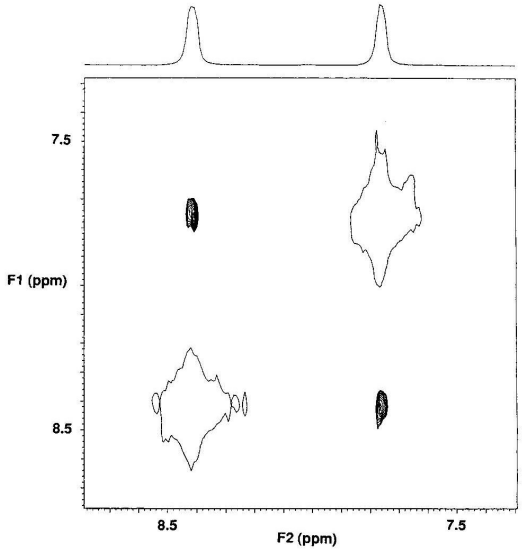
δNH data suggest a dramatic difference between D-Pro and L-Pro, with D-Pro strongly promoting β-hairpin formation.

Supports hypothesis that 'mirror image' β-turns stabilize β-hairpins.

Test Hypothesis: NOE Data

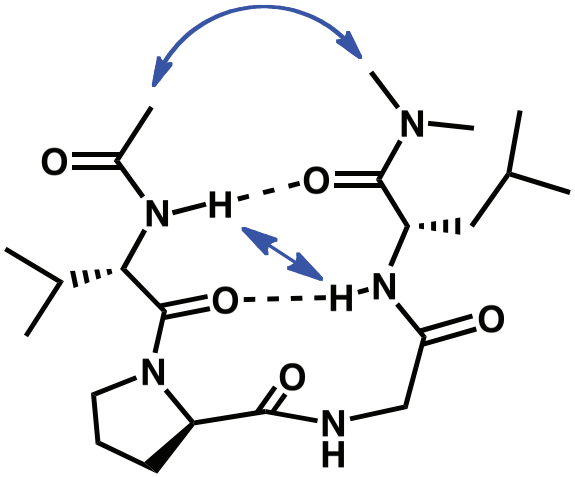


D-Proline

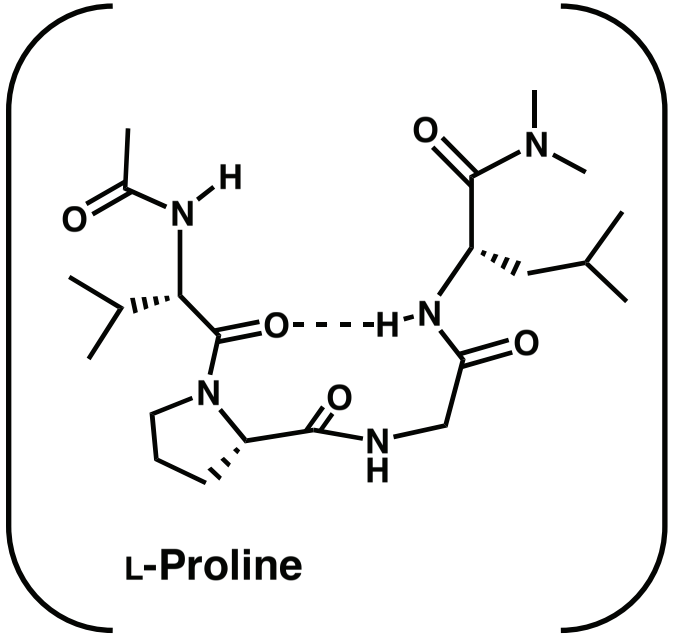


Amide NH region

Test Hypothesis: NOE Data

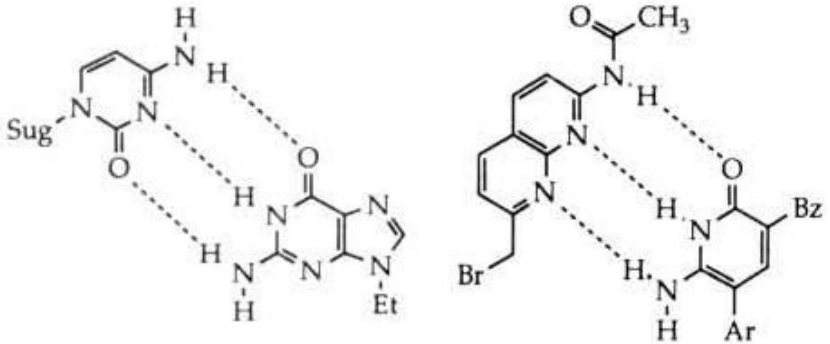


D-Proline



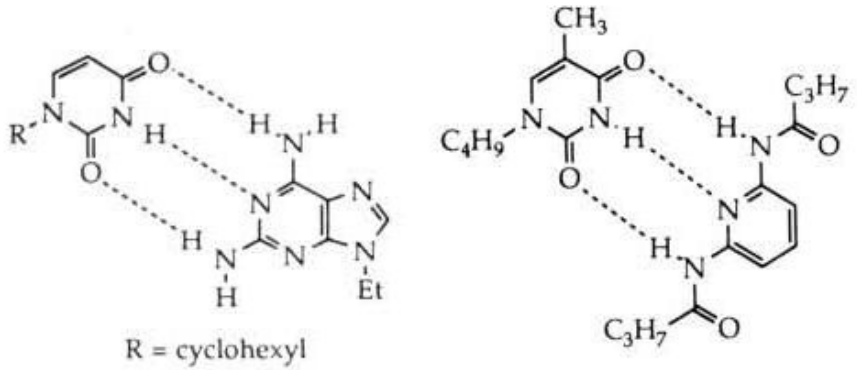
L-Proline

Secondary Interactions in Hydrogen Bond Arrays



K_a > 10,000 M^{-1}
(chloroform)

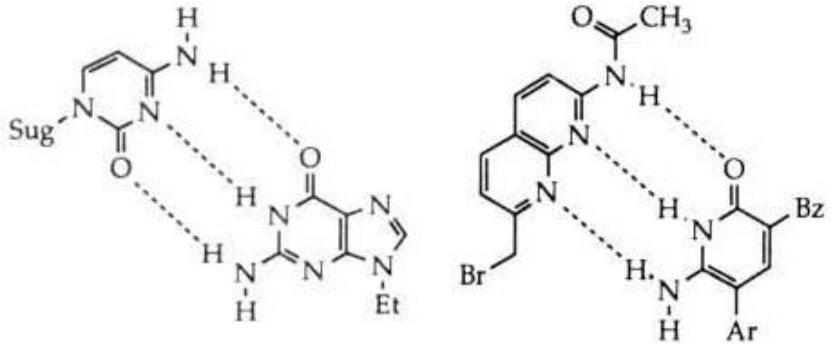
vs.



K_a ~ 100 M^{-1}
(chloroform)

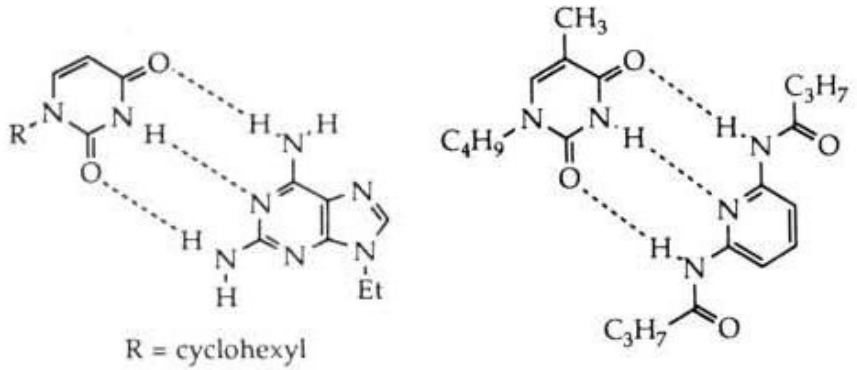
Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays



K_a > 10,000 M^{-1} (chloroform)

vs.

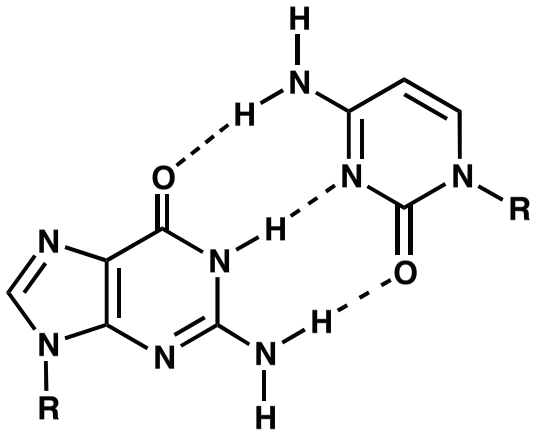


K_a ~ 100 M^{-1} (chloroform)

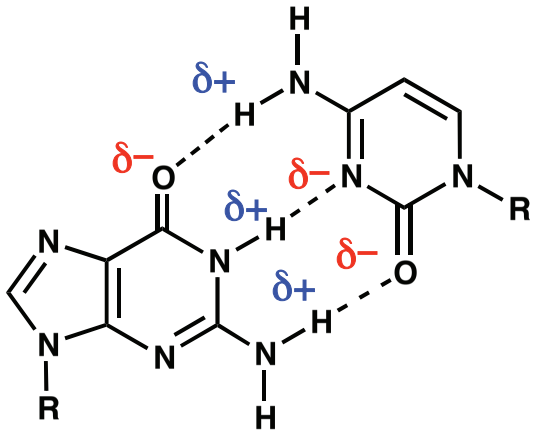
Why do pairs with equivalent H-bond numbers have diverse stability?

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interaction Hypothesis (G-C Base Pair)



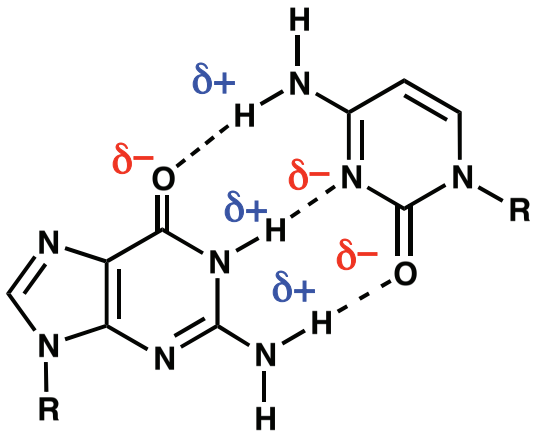
Secondary Interaction Hypothesis (G-C Base Pair)



δ+ = H-bond donor

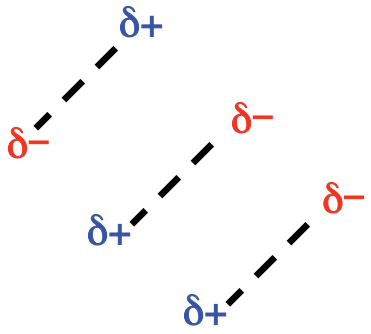
δ− = H-bond acceptor

Secondary Interaction Hypothesis (G-C Base Pair)



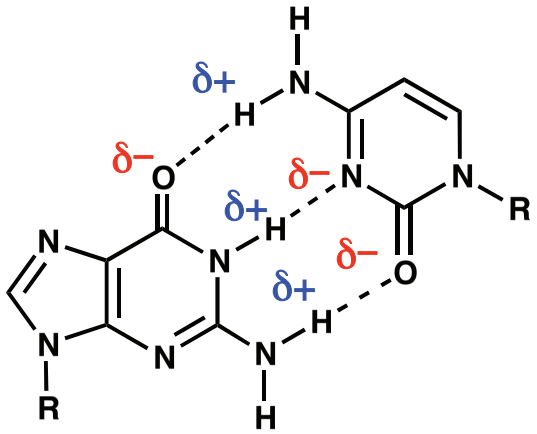
δ+ = H-bond donor

δ− = H-bond acceptor



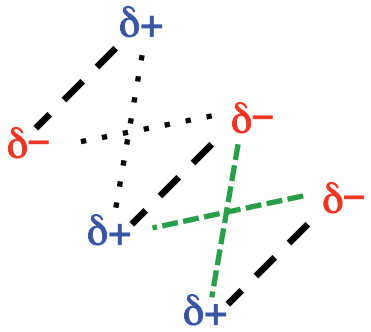
= H-bond

Secondary Interaction Hypothesis (G-C Base Pair)



δ+ = H-bond donor

δ− = H-bond acceptor



= H-bond

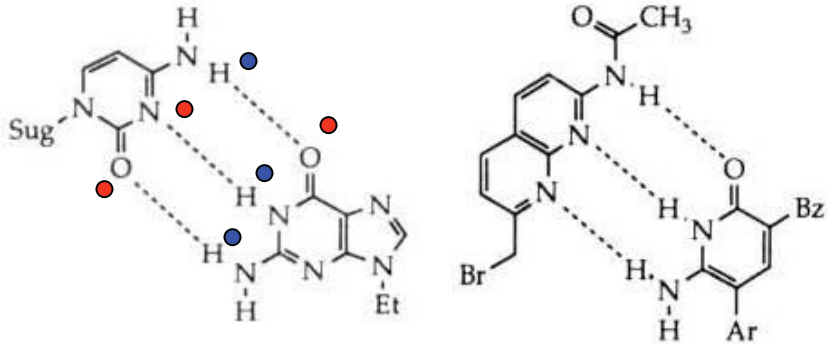
= Repulsive 2° interaction

= Attractive 2° interaction

Secondary Interactions in Hydrogen Bond Arrays

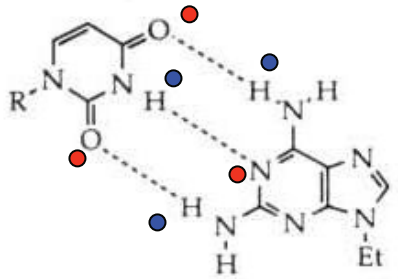
δ+ = ●

δ− = ●

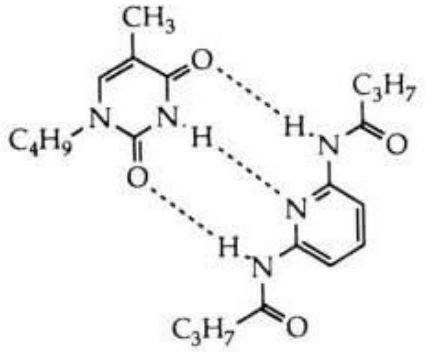


K_a > 10,000 M^{-1}
(chloroform)

vs.



R = cyclohexyl

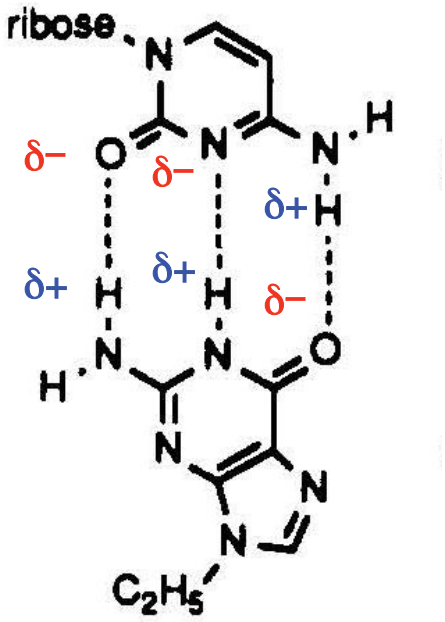


K_a ~ 100 M^{-1}
(chloroform)

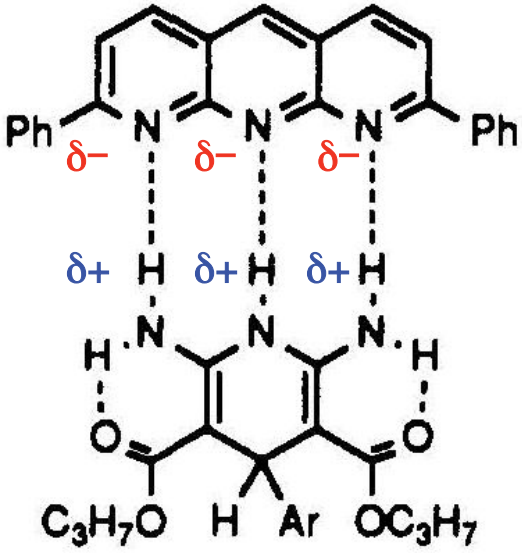
Why do pairs with equivalent H-bond numbers have diverse stability?

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays: Prediction Based on the Theory



vs.

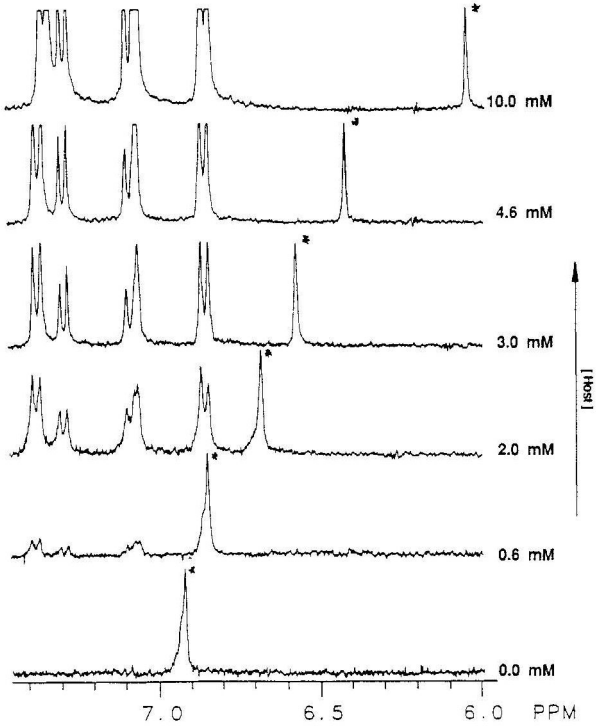


K_a = 1,000 M^{-1} (chloroform)

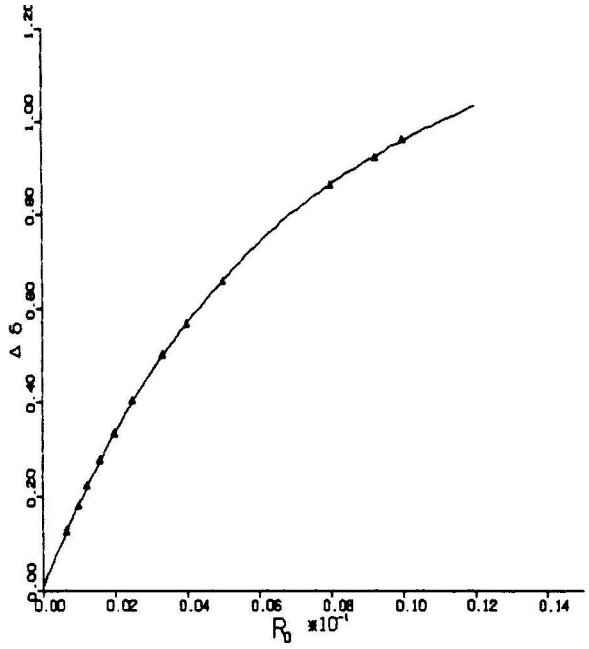
K_a > 10,000 M^{-1} (chloroform)

Murray, Zimmerman *J. Am. Chem. Soc.* 114:4010 (1992)

Methodology: NMR-Based Determination of K_a



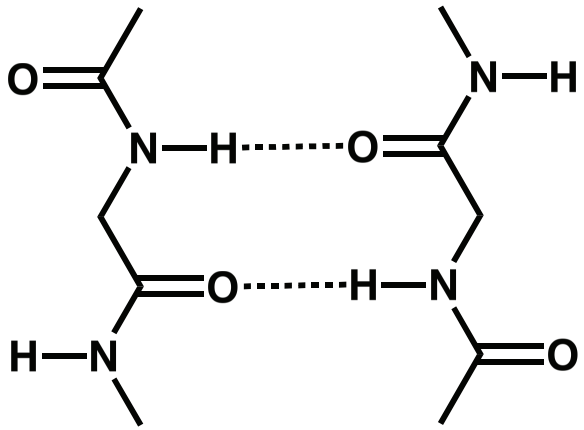
Δδ for guest proton



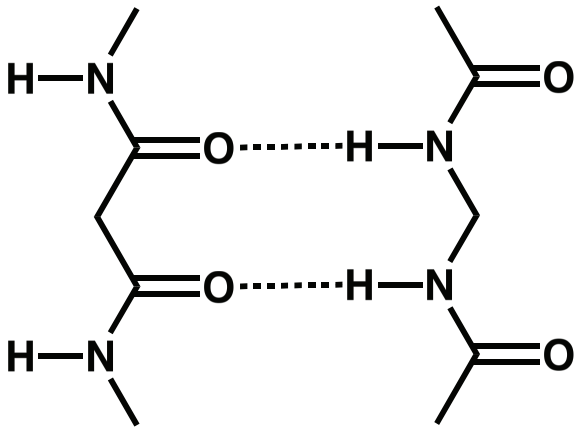
Concentration of host

Cowart, Sucholeiki, Bukownik, Wilcox *J. Am. Chem. Soc.* 110:6204 (1988)

Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics



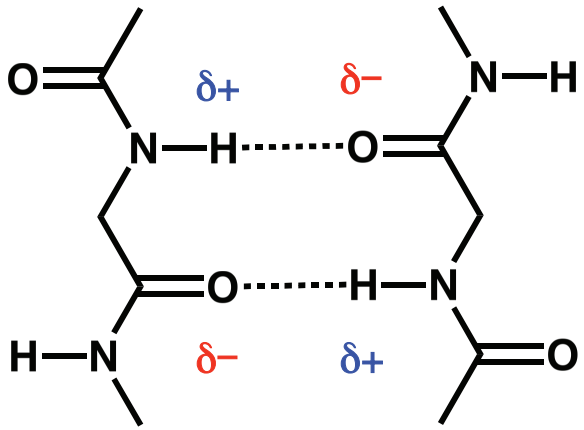
Antiparallel β-Sheet

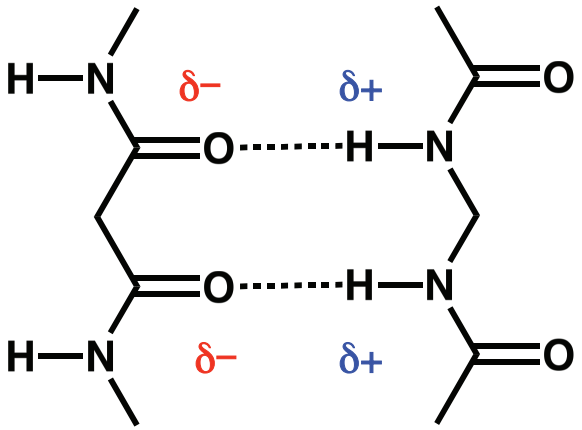


Peptide Mimics

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics



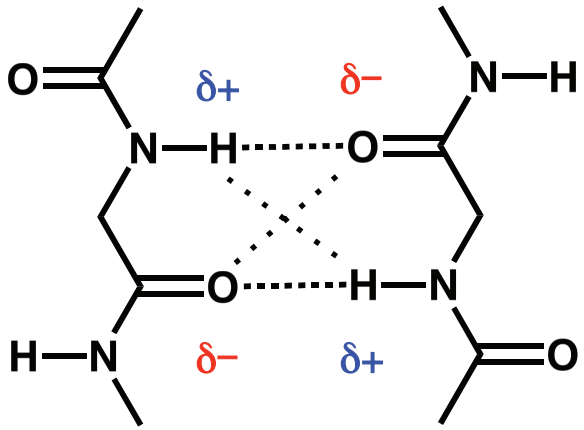


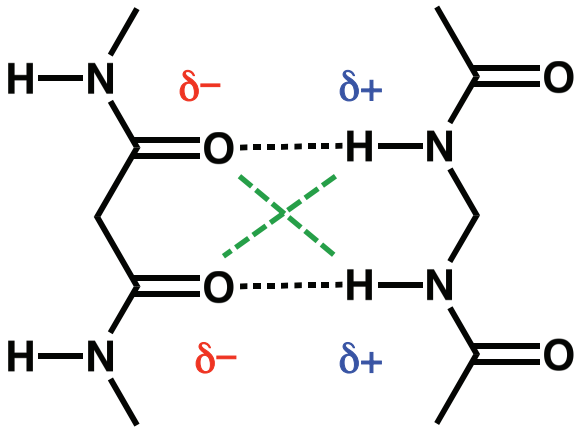
δ+ = H-bond donor

δ− = H-bond acceptor

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics





δ+ = H-bond donor

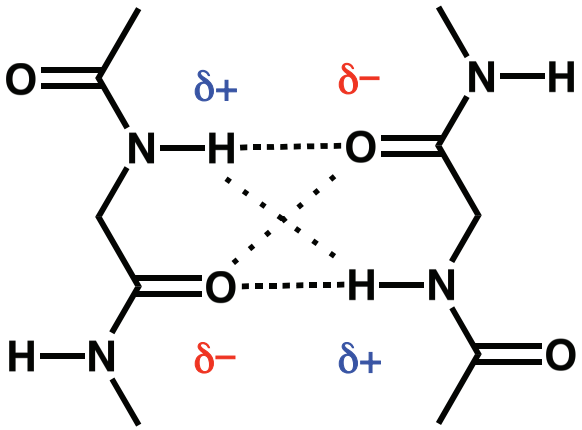
δ− = H-bond acceptor

. . . . . = Repulsive 2° interaction

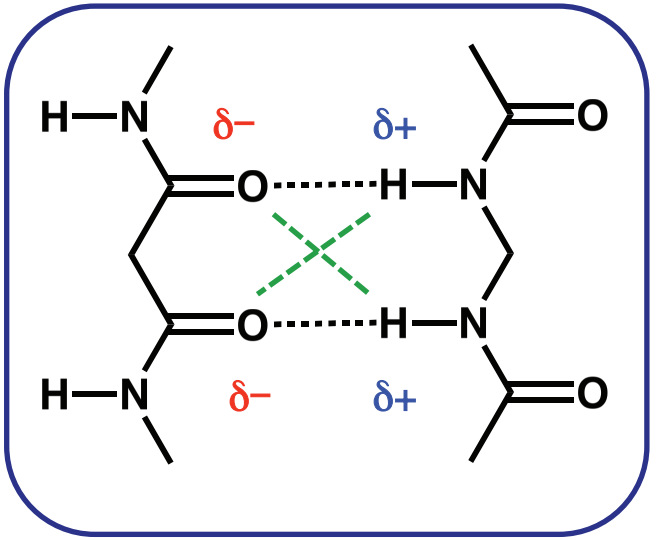
- - - - - - - = Attractive 2° interaction

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics



Predicted to be more stable



δ+ = H-bond donor

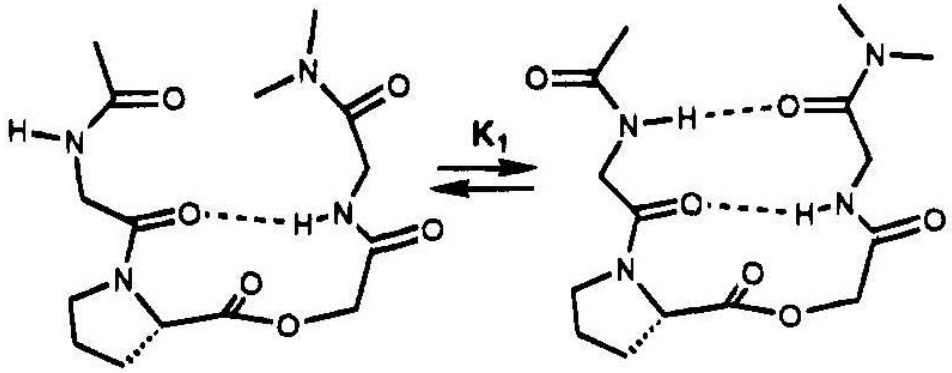
δ− = H-bond acceptor

. . . . . = Repulsive 2° interaction

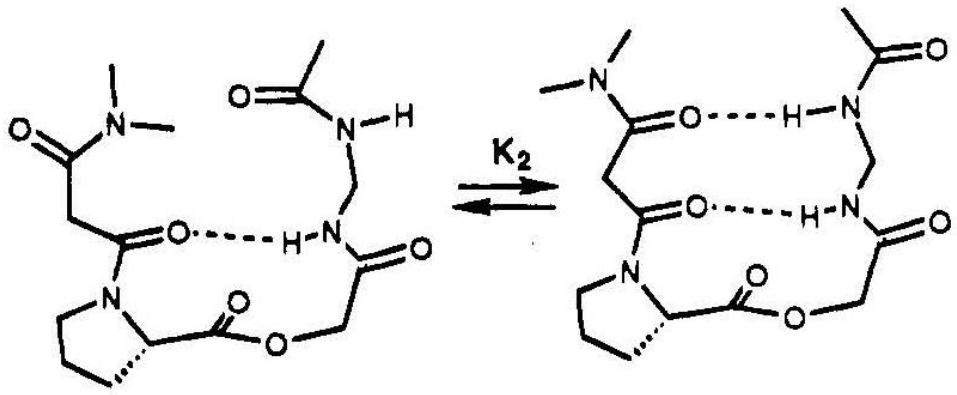
- - - - - - - = Attractive 2° interaction

Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics TEST

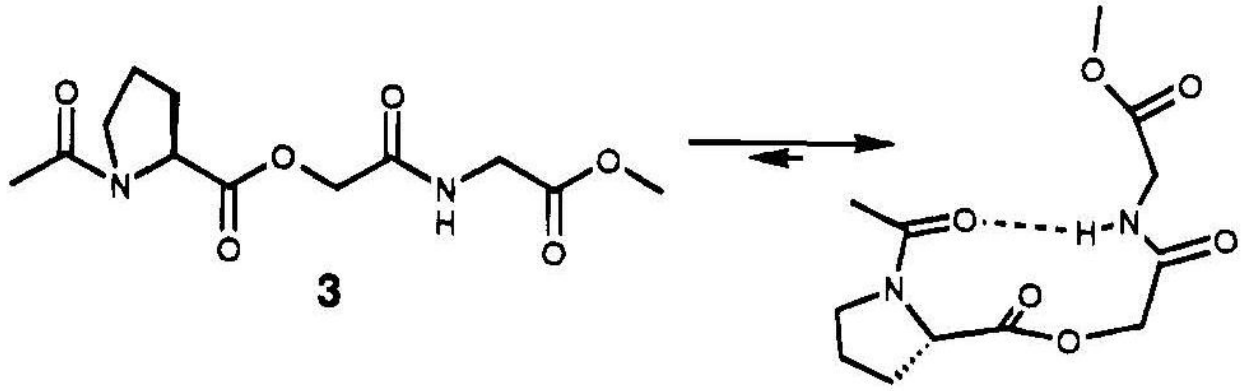


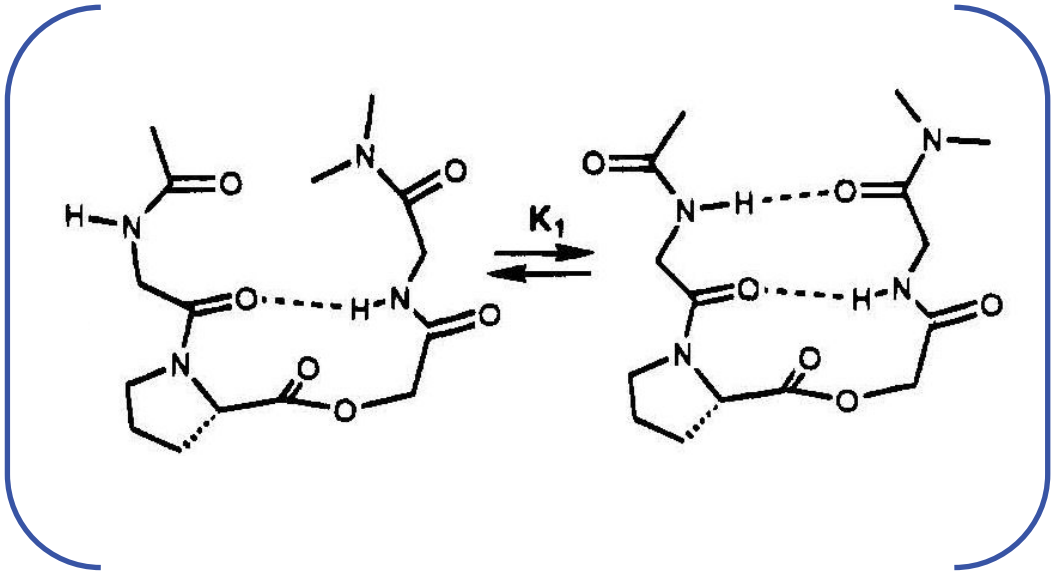
vs.



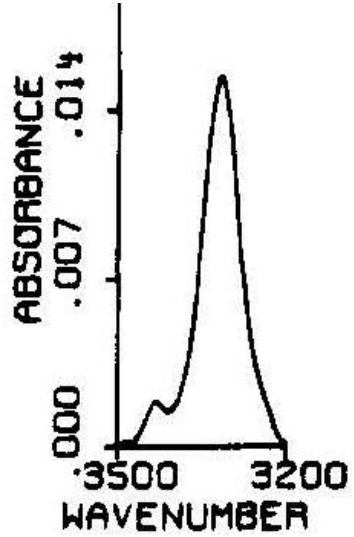
Gardner, Gellman *J. Am. Chem. Soc.* 117:10411 (1995)

Control: High Population of "Inner" H-bond (β-turn-like)

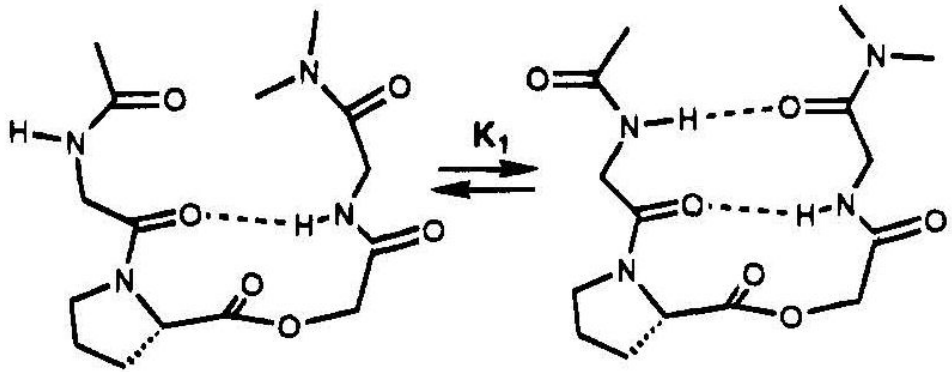




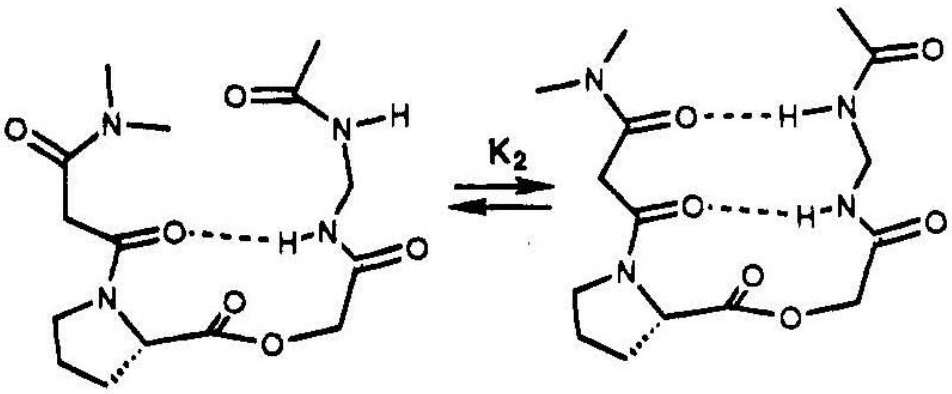
FT-IR, N-H Stretch Region
(1 mM, CH_2Cl_2)



Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics TEST

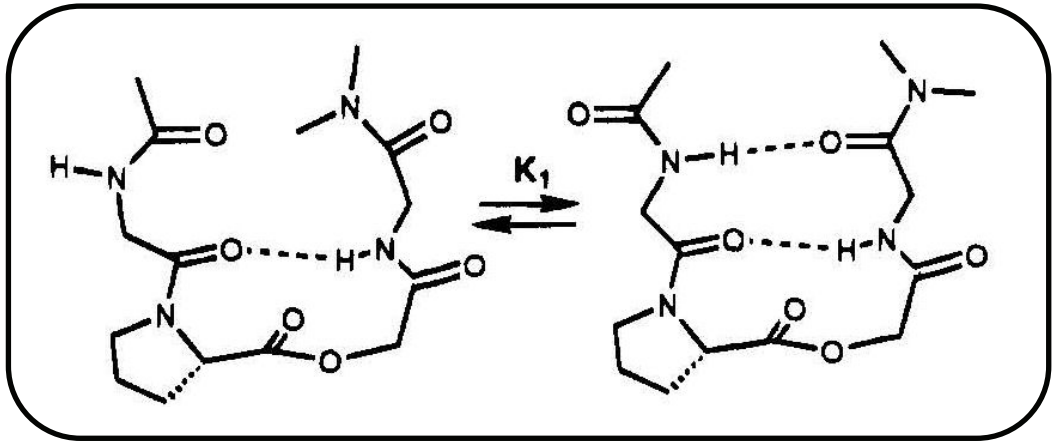


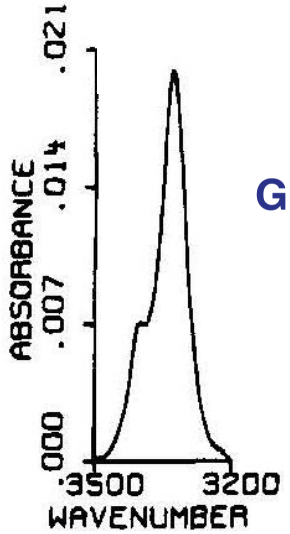
vs.



Jorgensen and Panata predict greater population of double H-bond state for this molecule

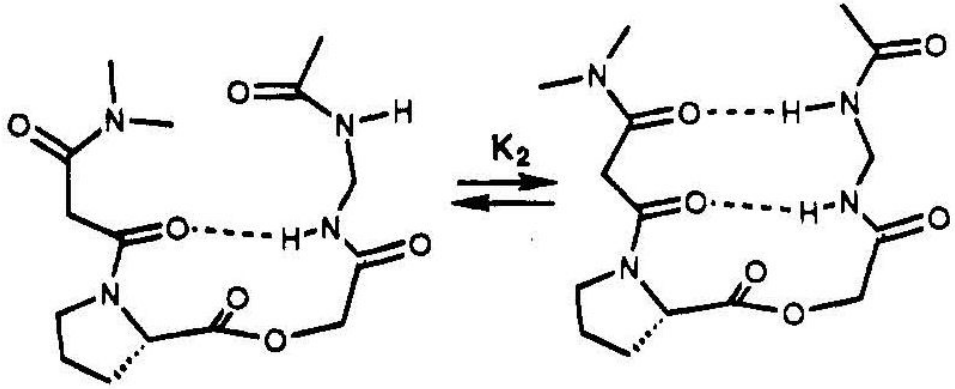
Secondary Interactions in Hydrogen Bond Arrays: Extension of Hypothesis to Peptides and Peptide Mimics TEST

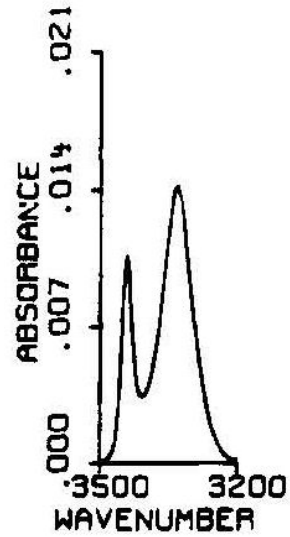




Greater double H-bonding observed

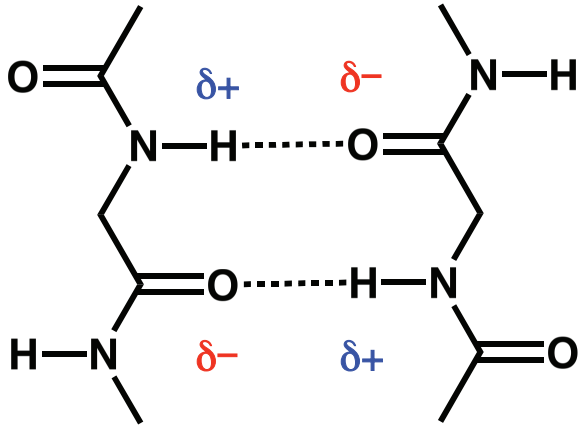
vs.

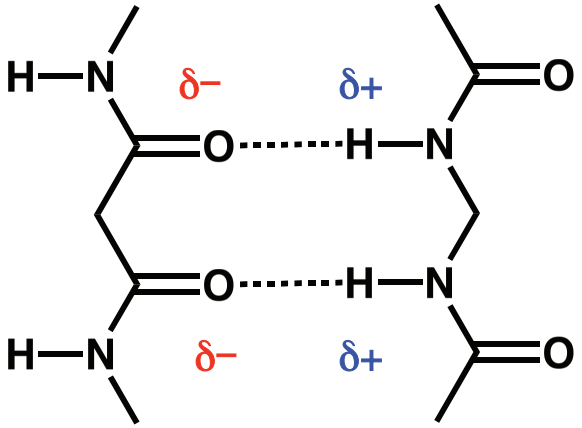




Predict more double H-bond

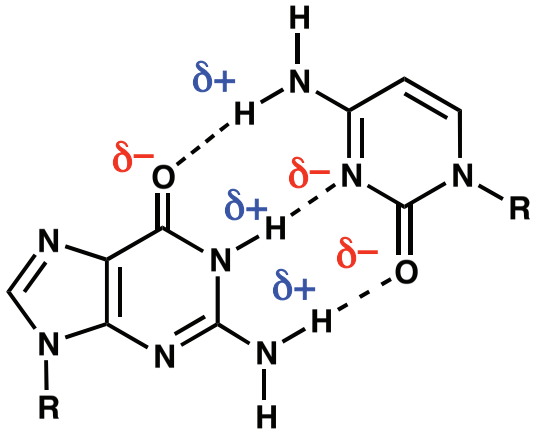
Why is the Prediction Incorrect?
Conformational Flexibility in Peptides





δ+ = H-bond donor

δ− = H-bond acceptor



Jorgensen, Pranata *J. Am. Chem. Soc.* 112:2008 (1990)

Conclusions

H-bonds are not dominant drivers of molecular folding processes, even in non-polar environments. Folded conformations that contain H-bonds reflect the operation of noncovalent networks that extend beyond the H-bonds themselves.

H-bonds are readily detected by spectroscopic methods, which is useful in terms of conformational analysis in solution.