

## NMR structural characterization of oligo-N-substituted glycine lead compounds from a combinatorial library

Erin K. Bradley,\* Janice M. Kerr, Lutz S. Richter, Gianine M. Figliozzi, Dane A. Goff, Ronald N. Zuckermann, David C. Spellmeyer & Jeffrey M. Blaney  
Small Molecule Drug Discovery, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, U.S.A.

Received 15 May 1997; Accepted 1 July 1997

**Key words:** combinatorial library, hydrophobic collapse, NMR, n-substituted glycines, structure

### Summary

Synthesis and screening of combinatorial libraries for pharmaceutical lead discovery is a rapidly expanding field. Oligo-N-substituted glycines (NSGs) were one of the earliest sources of molecular diversity in combinatorial libraries. In one of the first demonstrations of the power of combinatorial chemistry, two NSG trimers, CHIR-2279 and CHIR-4531, were identified as nM ligands for two 7-transmembrane G-protein-coupled receptors. The NMR characterization of these two lead compounds was undertaken to verify covalent connectivity and to determine solution conformations, if any. The sequential chemical shift assignments were performed using a new strategy for assigning  $^1\text{H}$  and  $^{13}\text{C}$  resonances of NSGs. The conformational preferences were then determined in both an aqueous co-solvent system and an organic solvent to probe the effects of hydrophobic collapse. NSGs are expected to be more flexible than peptides due to the tertiary amide, with both cis and trans amide bond conformations being accessible. Solution NMR studies indicate that although CHIR-2279 and CHIR-4531 have identical backbones and termini, and very similar side chains, they do not display the same solution conformational characteristics.

### Introduction

Synthesis and screening of combinatorial libraries for pharmaceutical lead discovery is a rapidly expanding field. Oligo-N-substituted glycines (NSGs) were one of the earliest sources of molecular diversity in combinatorial libraries [1,2]. In one of the first demonstrations of the power of combinatorial chemistry, an NSG library of over 4400 compounds was screened against various 7-transmembrane G-protein-coupled receptor targets. Two NSG trimers, CHIR-2279 and CHIR-4531 (Figure 1), were identified as nM ligands for two distinct receptors [3]. The first, CHIR-2279, binds to the  $\alpha 1$ -adrenergic receptor with a  $K_i$  of 5 nM, and CHIR-4531 binds to the  $\mu$ -opioid receptor with a  $K_i$  of 6 nM.

Due to the novelty of NSGs, there is little experimental information on their conformational preferences. However, since the amide nitrogen is substituted, NSGs are expected to be more flexible than peptides, with both cis and trans amide bond geometries being easily accessible. The NMR characterization of these two lead compounds was undertaken to verify covalent connectivity and to determine the solution conformations. Although NSGs are similar to peptides, the chemical nature of NSGs required the development of a new strategy [4] for assigning  $^{13}\text{C}$  and  $^1\text{H}$  resonances to confirm the chemical connectivity of these oligomeric molecules. This method employs the HMBC [5] and HMQC [6] pulse sequences and is robust enough to make the chemical shift assignments of four main chain geometries in solution simultaneously.

Solution NMR was used to determine conformational preferences in both an aqueous co-solvent system and an organic solvent. Wiley and Rich [7] have

\* To whom correspondence should be addressed. Present address: Combi Chem. Inc., 1804 Embarcadero Road, Suite 201, Palo Alto, CA 94303, U.S.A.

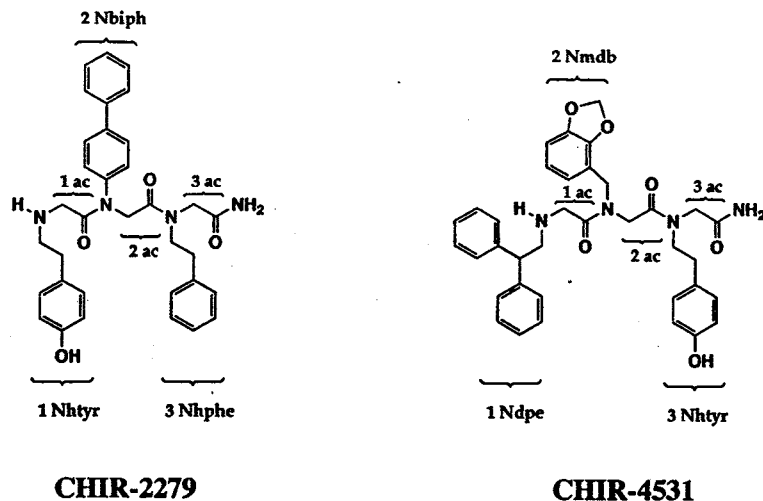


Figure 1. High-affinity ligands for the  $\alpha 1$ -adrenergic receptor (CHIR-2279) and the  $\mu$ -opioid receptor (CHIR-4531) discovered with a combinatorial NSG library [3]. Amine side chain and main chain segments are indicated using the nomenclature adopted for NSG NMR assignments [4]. In order to accommodate the chemistry and avoid confusion in naming of the two carbons attached to the main chain amide nitrogen, the polymer chain is divided into submonomer units. As a result, two pieces (submonomers) are associated with each residue (monomer) number. For example, 1 Nhtyr and 1 ac make up residue 1 of CHIR-2279.

suggested that structures determined in organic solvent or by in vacuo simulations are not good predictors of ligand conformation, either in aqueous solution or bound to the receptor. They propose that flexible ligands undergo 'hydrophobic collapse' when dissolved in water, minimizing their exposed hydrophobic surfaces by packing their hydrophobic groups together. As a result, intramolecular hydrophobic interactions would have a strong impact in aqueous solutions and not in organic solvents. Comparison of the NMR data from the two solvent conditions was a useful probe for the effects of 'hydrophobic collapse'.

CHIR-2279 and CHIR-4531 are good candidates for hydrophobic collapse, having hydrophobic side chains attached to a flexible backbone. This hydrophobic clustering or collapse could result in a limited number of solution conformations. If so, NMR techniques would be useful in characterizing their solution structures under aqueous conditions and providing insight for modeling and further optimization of these lead compounds. This is one reason why we pursue the structure of NSGs in at least partially aqueous solutions.

### Materials and Methods

CHIR-2279 and CHIR-4531 were synthesized using the submonomer method as previously reported [2,3].

The  $^{13}\text{C}$  labeling at the main chain CA position was achieved by using bromoacetic acid with a  $^{13}\text{C}$  label at the methylene position. Compounds were purified with reverse-phase HPLC using a water/acetonitrile gradient with 0.1% TFA buffer on a Vydac  $\text{C}_{18}$  column. The NSGs were characterized as described previously [3], and with the NMR experiments described here. They were lyophilized from HPLC buffer and then re-lyophilized from deuterated acetic acid ( $d_4$ ) before being dissolved in NMR solvents.

Neither CHIR-2279 nor CHIR-4531 was soluble in water alone at high enough concentrations for NMR. However, a previous study by Atkinson and Pelton [8] has shown for at least one small peptide that small amounts of acetonitrile can be used as a co-solvent without affecting the solution conformation. We have followed their example and added acetonitrile as a solubilizing agent. NMR was performed at  $\sim 10$  mM NSG in the aqueous co-solvent system of 25% acetonitrile ( $d_3$ ) and 75%  $\text{D}_2\text{O}$ , pH 3, at  $10^\circ\text{C}$ . NMR experiments were also performed with the same parameters on samples of  $\sim 10$  mM NSG in DMSO at  $30^\circ\text{C}$ . Spectra were acquired on a Varian Unity 300. Standard 1D proton spectra were collected as 8K data points and zero-filled to 16K. The 1D spectra were also taken at  $10\ \mu\text{M}$ ,  $100\ \mu\text{M}$  and  $1\ \text{mM}$  to verify that spectra (chemical shift and populations) were concentration independent (data not shown). The HMQC [6] experiments were performed with  $J = 140$  Hz, and the HMBC [5] exper-

iments with  $J = 9$  Hz. The heteronuclear experiments were collected with spectral widths of 4000 Hz in the  $^1\text{H}$  dimension, and 16 000 and 18 000 Hz in the  $^{13}\text{C}$  dimension for the HMQC and the HMBC, respectively. The ROESY data were collected using the TROESY pulse sequence of Hwang and Shaka [9] with a 250 ms mixing time. The spin-lock pulse was  $180^\circ(x) 180^\circ(-x)$ , with a field strength of 8000 Hz. For the ROESY spectra, a total of 512  $t_1$  increments were collected, with 64 transients of 2048 complex data points in  $t_2$  for each. Spectral widths of 4000 Hz were used in both dimensions. Data processing was carried out with the NMR Pack software [10] on an IBM RS-6000/580 running AIX3.2.5. The ROESY spectra were zero-filled to give final real data matrices of 1024 points in  $\omega_1$  and 2048 points in  $\omega_2$ . Heteronuclear spectra were zero-filled to give a final real data matrix of 1024 points in  $\omega_2$  ( $^1\text{H}$ ) and 2048 points in  $\omega_1$  ( $^{13}\text{C}$ ). The HMQC was processed as phase sensitive in both dimensions, and the HMBC was processed as phase sensitive in  $\omega_2$  and absolute value in  $\omega_1$ .

Although NSGs are similar to peptides, a new strategy [4] had to be employed for assigning  $^{13}\text{C}$  and  $^1\text{H}$  resonances. For peptide sequential NMR assignments, spin systems consisting of  $\text{H}^N$ ,  $\text{H}^\alpha$ ,  $\text{H}^{\beta 1}$ ,  $\text{H}^{\beta 2}$ , etc., are identified using homonuclear COSY [11,12] or TOCSY [13] experiments. These homonuclear experiments are adequate for peptide systems in which the main chain protons and the side chain protons are within three bonds of another proton. Thus, for most amino acid monomers, the main chain and part of the side chain are one spin system. With NSGs, the side chain branch point is shifted from the main chain  $\alpha$ -carbon to the main chain nitrogen, replacing the amide proton with the first carbon of the side chain. As a result, each monomer unit has two spin systems, the side chain branching from the main chain nitrogen, 'Nxxx', and the main chain glycine unit, 'ac' (Figure 1). The assignment method and the nomenclature used here are described briefly in Figures 1 and 2, and have been published in detail elsewhere [4]. Chemical shift assignments for CHIR-2279 and CHIR-4531 in 25%  $\text{CH}_3\text{CN}/75\%$   $\text{D}_2\text{O}$  at 10  $^\circ\text{C}$  and for CHIR-2279 in DMSO at 30  $^\circ\text{C}$  are listed in Tables 1, 2 and 3, respectively.

Using the sequential connections determined with the HMBC spectra (Figure 3), it was possible to determine the geometry about each of the amide bonds with the ROESY spectra (Figures 4, 5 and 6). This can be clearly seen with the tracing of the sequential connections of CHIR-2279 in a ROESY spectrum taken in

Table 1.  $^{13}\text{C}$  and  $^1\text{H}$  chemical shift<sup>a</sup> assignments of CHIR-2279 in 25%  $\text{CH}_3\text{CN}/75\%$   $\text{D}_2\text{O}$  at 10  $^\circ\text{C}$

Submonomer		trans-trans		trans-cis	
		$^{13}\text{C}$	$^1\text{H}$	$^{13}\text{C}$	$^1\text{H}$
1 Nhtyr	A	47.2	3.08	47.2	3.13
	B	29.5	2.82	29.5	2.85
	C	126.5			
	D	129.0	7.05	128.9	7.07
	E	114.0	6.79	114.0	6.79
	F	154.0		154.0	
1 ac	A	46.7	3.58	46.8	3.72
	CO	164.7		164.8	
2 Nbiph	A	137.5			
	B	127.1	7.15	127.1	7.37
	C	127.4	7.70	127.4	7.75
	D	140.6			
	E	138.3			
	F	125.8	7.75	125.9	7.71
	G	128.2	7.55	128.2	7.53
	H	127.2	7.48	127.2	7.51
2 ac	A	50.2	3.89	50.3	4.48
	CO	167.9		167.6	
3 Nhphe	A	48.6	3.59	48.6	3.58
	B	32.1	2.79	31.9	2.78
	C	137.0			
	D	127.7	7.04	127.7	7.16
	E	127.7	7.00	127.5	7.23
	F	125.6	6.74	125.4	7.22
3 ac	A	47.2	4.07	48.4	3.96
	CO	172.0		171.0	

<sup>a</sup> Units of ppm  $^{13}\text{C}$  referenced to 117.7 ppm  $\text{CH}_3\text{CN}$ ;  $^1\text{H}$  referenced to 0.0 ppm TSP.

25%  $\text{CH}_3\text{CN}/75\%$   $\text{D}_2\text{O}$  (Figure 4). CHIR-2279 has a distinct preference at each of its amide bonds in the aqueous co-solvent. It is a little more complicated in the CHIR-4531 spectrum, which has four main chain geometries populated under partially aqueous conditions (Figure 5). The main chain geometries are noted in the assignment tables and figure legends. This same procedure was followed in DMSO as well.

## Results

Although CHIR-2279 and CHIR-4531 have identical backbones and termini, and similar side chains, they do not display the same solution conformational char-

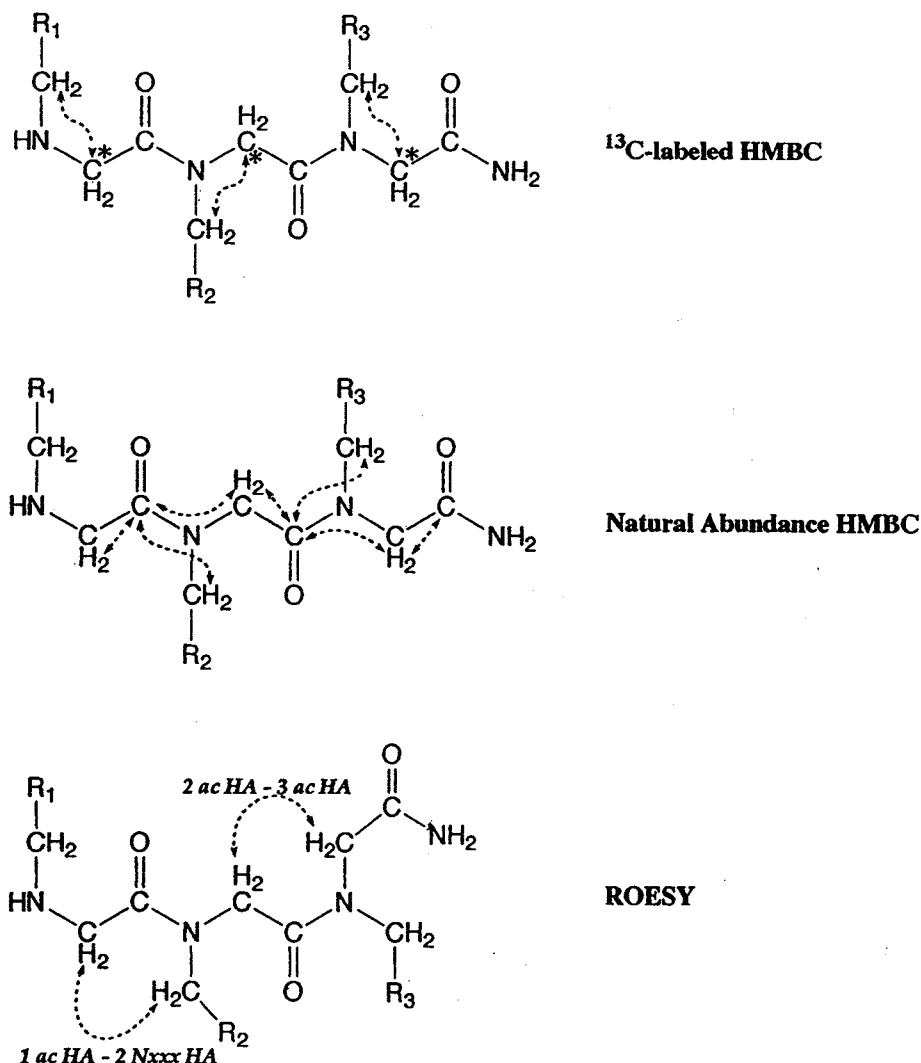
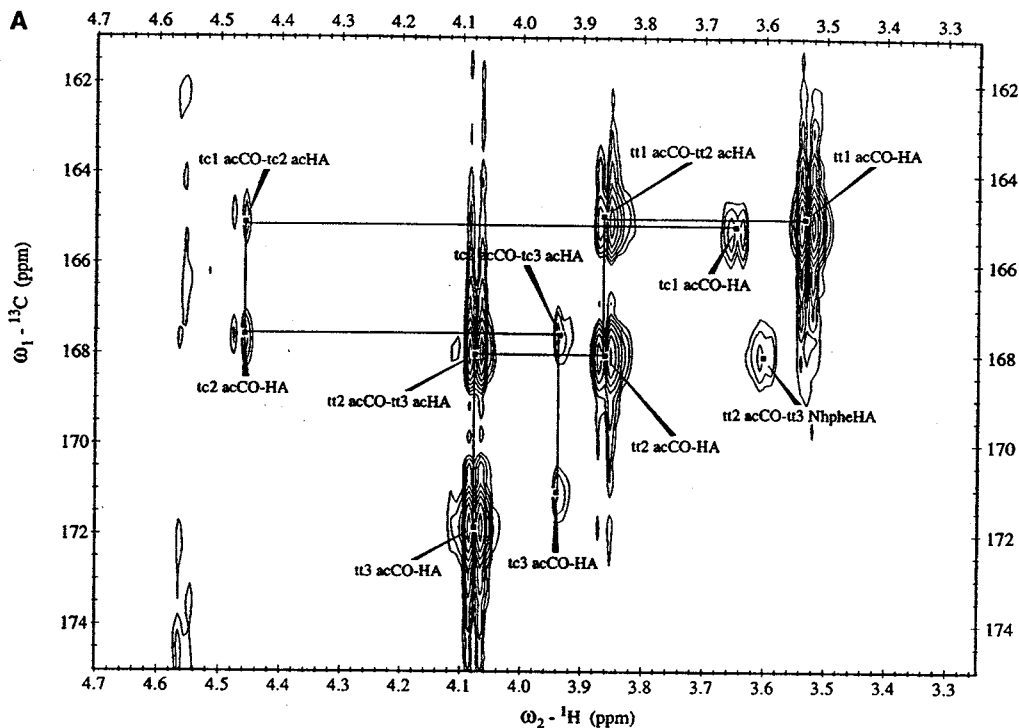


Figure 2. (Top) Intramonomer connections (e.g. 1 Nxxx to 1 ac) are established using the heteronuclear three-bond coupling of Nxxx HA to ac CA. The dashed arrows indicate cross peaks which connect side chain spin systems to the main chain spin system in the HMBC [5] of an NSG <sup>13</sup>C labeled at each ac CA position (designated by \*). (Middle) Intermonomer sequential connections are established using heteronuclear two- or three-bond couplings of main chain carbonyl carbons (ac CO) to side chain and main chain protons (Nxxx HA and ac HA). The dashed arrows indicate cross peaks found in a natural abundance HMBC that are used to trace sequential connectivities. (Bottom) The amide bond geometries are established using ROESY cross peaks. The distances between the main chain protons (ac HA) differ, depending on whether the intervening amide bond places these protons cis or trans to each other. The dashed arrows indicate cross peaks found in a TROESY [9]. The trans geometry at the first amide will result in a 1 ac HA to 2 Nxxx HA cross peak, while the cis geometry at the second amide will result in a 2 ac HA to 3 ac HA cross peak.

acteristics under either the aqueous co-solvent system or the organic solvent used here. The HMBC ac CO-HA regions of the two compounds are clearly different under an aqueous co-solvent system (Figures 3A and B). The assigned HMBC spectrum indicates that CHIR-2279 exists as one major species which can be easily observed in Figure 3A and a minor species coming in just above the noise. These species differ in

the main chain geometry. In contrast, CHIR-4531 has a much more complex HMBC spectrum (Figure 3B). This spectrum along with ROESY (Figure 5) demonstrate that the CHIR-4531 backbone does have access to all four possible main chain geometries. The presence of multiple peaks means that the cis and trans states are in slow exchange on the NMR time scale.

With the chemical shift assignments completed and



**Figure 3.** The carbonyl carbon/ $\alpha$ -proton region of natural abundance HMBC [5] spectra. The assignments and sequential connections are indicated for both major and minor conformations. The peak assignments are shown with the main chain amide bond geometries indicated for each species, with *tt* denoting the trans-trans main chain geometry, *tc* the trans-cis, *ct* the cis-trans, and *cc* the cis-cis. For CHIR-2279 in 25%  $\text{CH}_3\text{CN}/75\% \text{D}_2\text{O}$  at 10 °C (A) the HMBC clearly shows a major and a minor main chain geometry, *tt* and *tc*, respectively. In addition to the two main chain cross peaks at the 2 ac CO carbon frequency, there is also a cross peak from the side chain 3 Nhphe HA (3.60 ppm). The 2 Nbiph side chain (Figure 1) does not have an HA, and thus does not have the corresponding cross peak at the 1 ac CO carbon frequency. The HMBC of CHIR-4531 in 25%  $\text{CH}_3\text{CN}/75\% \text{D}_2\text{O}$  at 10 °C (B) has many more cross peaks, reflecting the presence of four main chain geometries, two major and two minor. The side chain connectivities are also present with 2 Nmdb HA to 1 ac CO cross peaks apparent from 4.5 to 4.2 ppm and the 3 Nhtyr HA to 2 ac CO cross peaks around 3.5 ppm. The HMBC of CHIR-2279 in DMSO at 30 °C (C) has two main chain conformers of roughly equal intensity and very little chemical dispersion between them.

the main chain geometries determined, it was possible to use the  $^1\text{H}$  spectra to estimate the populations occupying the different main chain amide geometries (Figures 7A and B). CHIR-2279 has a distinct amide bond geometry preference at both of the main chain amide bonds, with one major species, trans-trans ( $\sim 85\%$ ), and one minor species, trans-cis ( $\sim 15\%$ ). This is somewhat unexpected, since the energy difference between cis and trans isomers is low for N-substituted amide bonds. The unequal population distribution of the main chain amide geometries is indicative of conformational preferences.

CHIR-4531 behaves as expected, clearly occupying both cis and trans amide bond geometries at each of the two amide bonds. This results in four main chain species. Figure 7B is a portion of the 1D spectrum used to make population estimates for each of the con-

formers of CHIR-4531. There are two major components, trans-trans and cis-trans ( $\sim 40\%$  each), and two minor ones, cis-cis and trans-cis ( $\sim 10\%$  each). The differences between CHIR-2279 and CHIR-4531 in the number of main chain geometries and the occupancies of those states is not a result of aggregation. The 1D spectra were also taken at 10  $\mu\text{M}$ , 100  $\mu\text{M}$  and 1 mM, and verify that the number of peaks, their chemical shifts and normalized integrations were concentration independent.

The fact that CHIR-2279 has a preference for one main chain geometry over the others under aqueous conditions suggests that there is a favored 'collapsed' conformation. However, there are no ROESY cross peaks observed between the hydrophobic side chains to indicate close packing. Thus, there is no evidence of a single collapsed conformation. The same is true

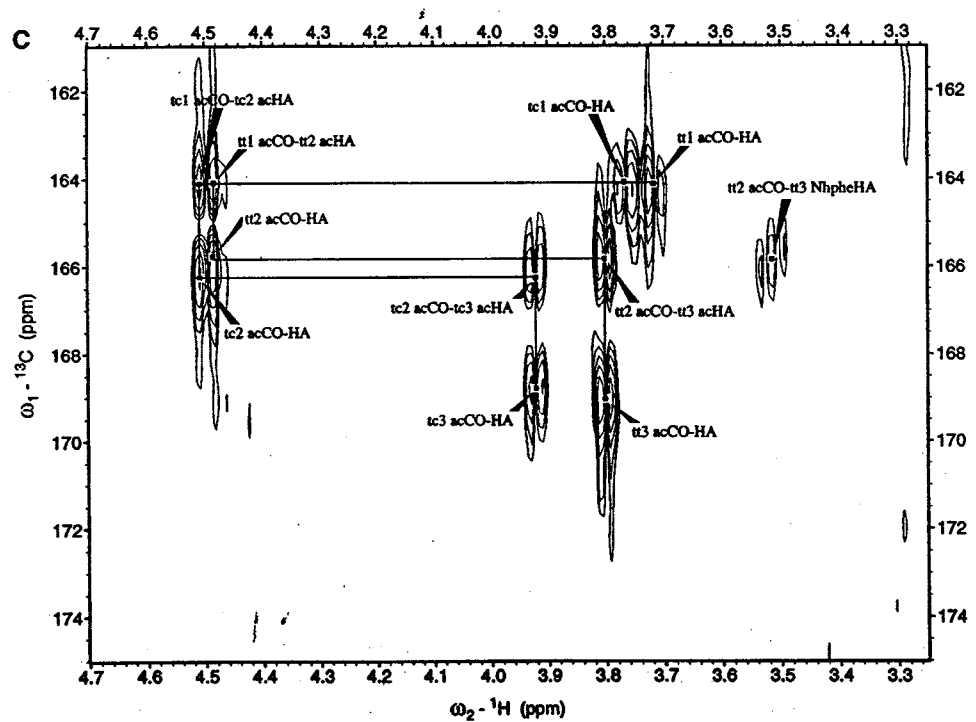
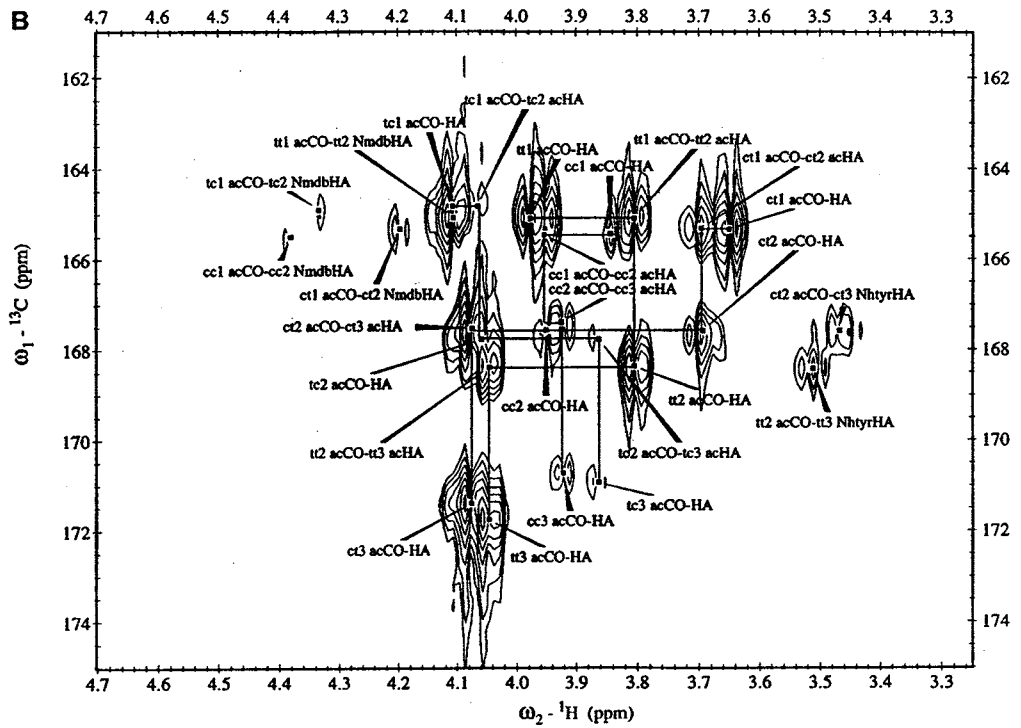


Figure 3. (Continued).

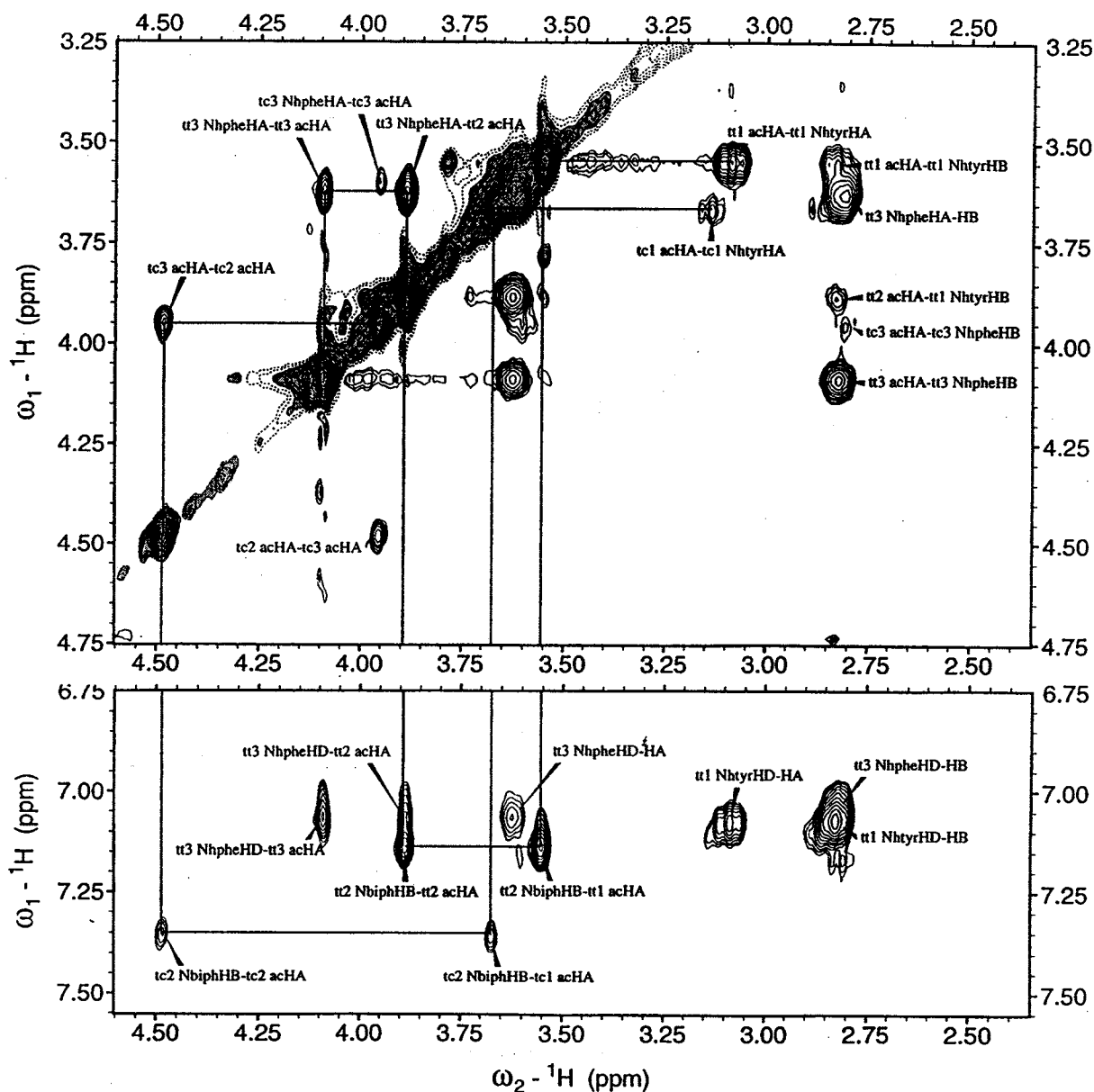


Figure 4. Portions of the TROESY [9] spectra of CHIR-2279 in 25%  $\text{CH}_3\text{CN}$  ( $d_3$ )/75%  $\text{D}_2\text{O}$  at 10 °C. These spectra contain the sequential connectivities of both the major and minor species, and were used in conjunction with main chain assignments to determine amide bond geometries. The assignments are shown, with tt denoting the trans-trans main chain geometry and tc the trans-cis. Note that the 2 Nbiph side chain has no HA (Figure 1), so the diagnostic cross peak for the trans amide bond is 1 ac HA – 2 Nbiph HB (lower portion of the spectra).

for CHIR-4531, with all the NMR distance constraints derived from ROESY spectra being of a local nature.

When the same NMR characterization was performed on CHIR-2279 and CHIR-4531 in DMSO at 30 °C, the results were dramatically different. It is immediately clear from the 1D spectra of CHIR-2279 in the main chain  $\alpha$ -proton region, between 4.5 and

3.5 ppm, that there are more peaks present in DMSO (Figure 7C) than in 25%  $\text{CH}_3\text{CN}$  (Figure 7A). The assigned HMBC spectrum of CHIR-2279 in DMSO (Figure 3C) has the same number of peaks, and shows that there are still only two main chain geometries. But the peaks now have the same intensity, and the chemical shift dispersion is greatly reduced. This suggests

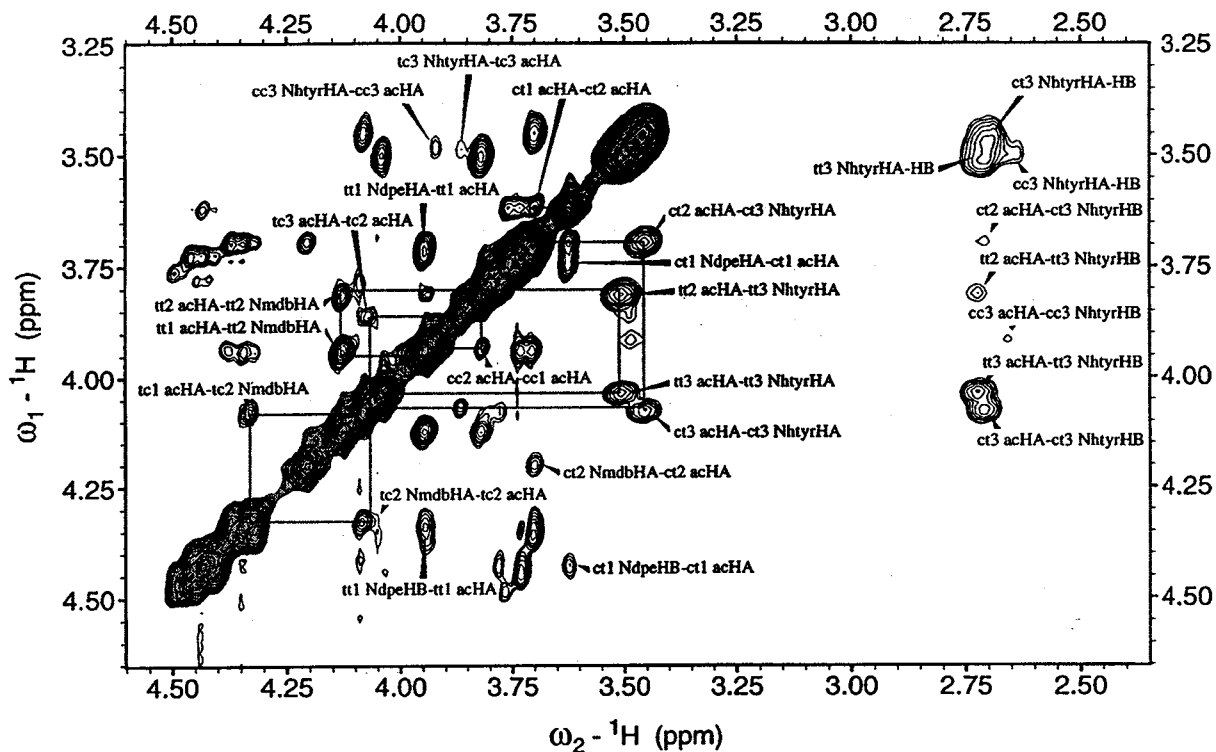


Figure 5. Portion of the TROESY [9] spectrum of CHIR-4531 in 25%  $\text{CH}_3\text{CN}$  ( $d_3$ )/75%  $\text{D}_2\text{O}$  at 10 °C. This spectrum contains the sequential connectivities of both the major and minor species, and was used in conjunction with main chain assignments to determine amide bond geometries. The assignments are shown, with tt denoting the trans-trans main chain geometry, tc the trans-cis, ct the cis-trans, and cc the cis-cis.

that there is little difference in the chemical environment of the nuclei in the two different main chain geometries. This is consistent with the unfolded states of peptides. For CHIR-4531 four conformers are still present but they have altered populations. In CHIR-2279 the first amide bond is aromatically substituted and occupies only the trans geometry. This is consistent with the work of Itai et al. [14,15], and will be discussed in more detail in the next section.

The 1D spectra (Figure 7) show that solvent has a dramatic effect on the populations. CHIR-2279 shifts from trans-trans being favored 85:15 over trans-cis in a partially aqueous environment, to an equal distribution in DMSO for the non-benzylic amide bond. CHIR-4531 shows similar behavior, with no main chain conformational preference at either amide bond. Each combination of amide bond geometries, trans-trans, trans-cis, cis-trans, and cis-cis, is occupied equally (~25% each). In DMSO, all the ROESY cross peaks are once again of a local nature, and there is no evidence of close packing. This was expected in DMSO,

where the hydrophobic effect is not a driving force, and both the polar main chain and the hydrophobic side chains would be fully solvated.

## Discussion

The NMR characterization of CHIR-2279 and CHIR-4531 was undertaken to verify covalent connectivity and to determine if there was any preferred solution conformation. The NMR characterization included the assignment of  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts and the determination of main chain amide geometry for each species present in solution. The work presented here demonstrates that the  $^1\text{H}$  and  $^{13}\text{C}$  assignment method using HMQC [6], HMBC [5], and TROESY [9] pulse sequences can be used on NSGs for up to four main chain geometries of varying populations, simultaneously in solution.

Once the assignments were complete, it was possible to check for conformational preferences. Since the

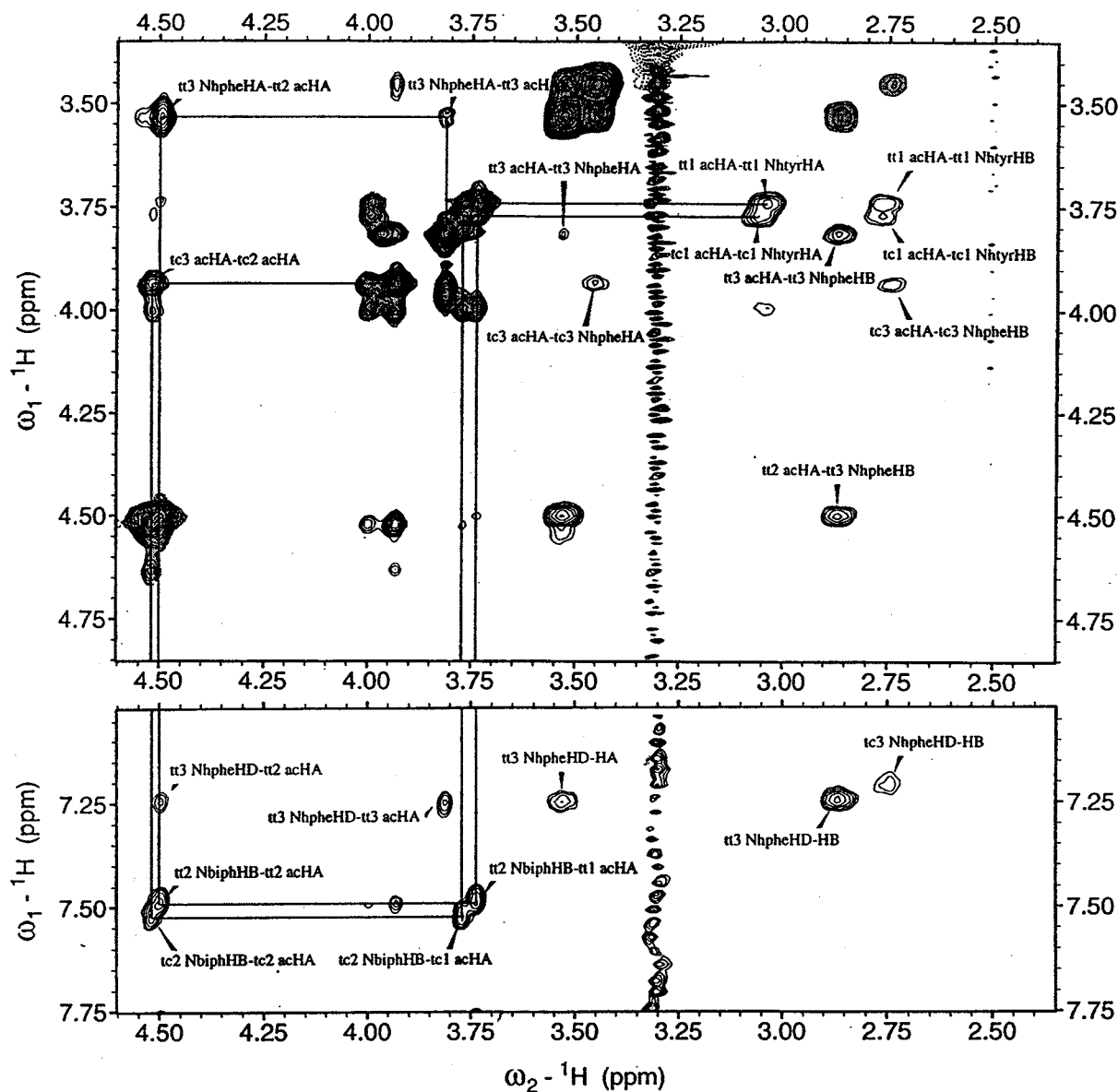
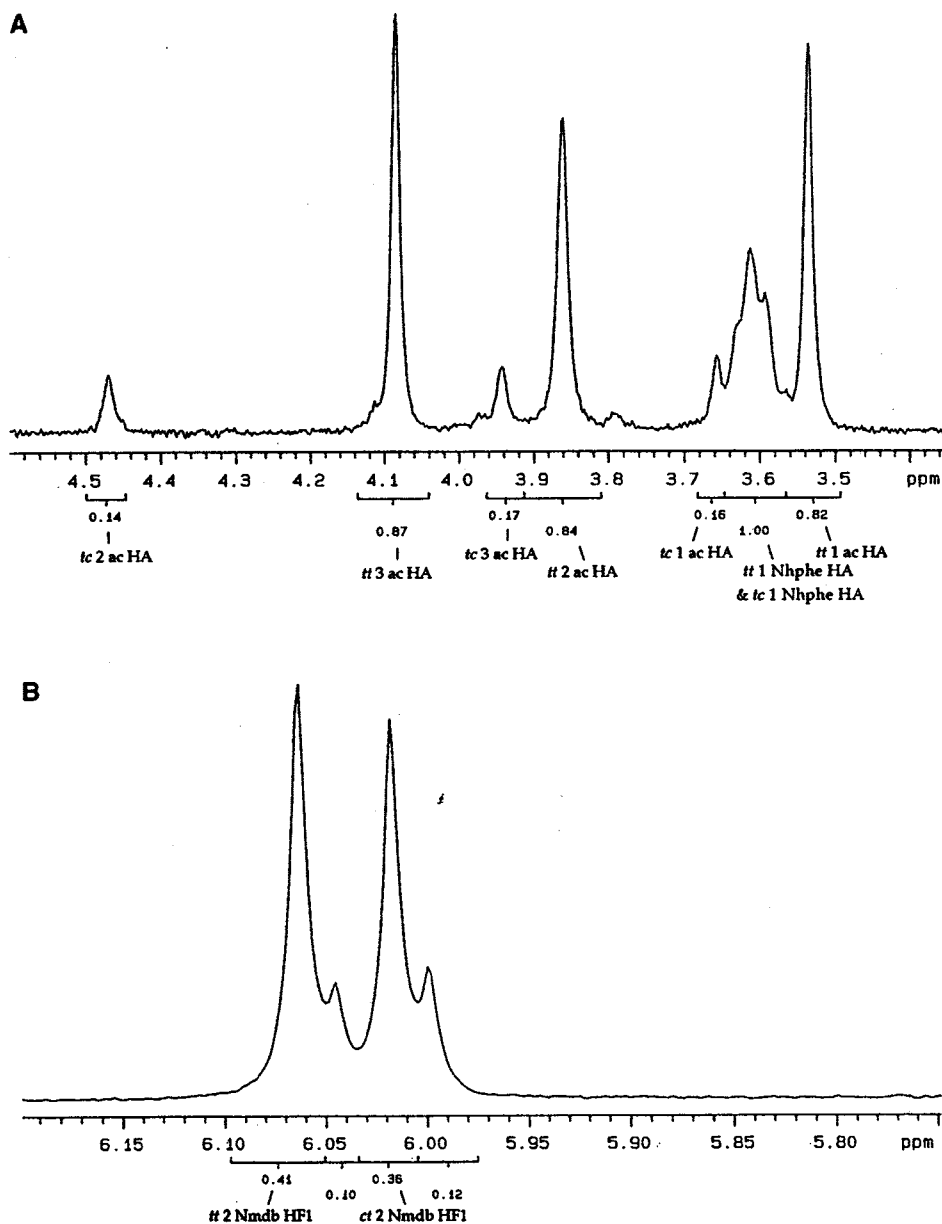


Figure 6. Portions of the TROESY [9] spectra of CHIR-2279 in DMSO ( $d_6$ ) at 30 °C. These spectra contain the sequential connectivities of both the major and minor species, and were used in conjunction with main chain assignments to determine amide bond geometries. The assignments are shown, with tt denoting the trans-trans main chain geometry, tc the trans-cis, ct the cis-trans, and cc the cis-cis. Note that 2 Nbiph has no HA (Figure 1), so the diagnostic cross peak for the trans amide bond is 1 ac HA - 2 Nbiph HB (lower portion of the spectra). In DMSO, because of the chemical shift overlap, the critical peaks for determining main chain geometry at the second amide are in the upper left corner.

energy difference between the cis and trans isomers of amide bonds is low for N-substituted amide bonds, NSGs are expected to be more flexible than peptides. Both cis and trans amide bond geometries would be accessible, similar to proline (tertiary amide). Because the group directly connected to the nitrogen is the

same for the side chain and the main chain (methylene groups,  $\text{CH}_2$ ), it was expected that the NSGs would sample both geometries roughly equally. This would result in four discernible main chain geometries for a trimer (with two amide bonds): trans-trans, trans-cis, cis-cis, and cis-trans. These four geometries are



**Figure 7.** Portions of standard  $^1\text{H}$  spectra used to estimate the occupancy of each main chain geometry. The assignments are shown below the peaks along with the peak integration, with *tt* denoting the trans-trans main chain geometry, and *tc* the trans-cis, *ct* the cis-trans, and *cc* the cis-cis. (A) CHIR-2279 in 25%  $\text{CH}_3\text{CN}$  ( $d_3$ )/75%  $\text{D}_2\text{O}$ , 10  $^\circ\text{C}$ . Integrated values are normalized to 4 for the estimation of populations based on four resonances (1 ac HA, 2 ac HA, 3 ac HA, and 3 Nhphe HA). The peak integrations indicate that CHIR-2279 has a major conformer ( $\sim 85\%$ ) with trans-trans geometry and a minor conformer ( $\sim 15\%$ ) with trans-cis geometry. (B) CHIR-4531 in 25%  $\text{CH}_3\text{CN}$  ( $d_3$ )/75%  $\text{D}_2\text{O}$ , 10  $^\circ\text{C}$ . Integrated values are normalized to 1 for the estimation of populations based on one resonance (2 Nmdb HF1). The peak integrations indicate that CHIR-4531 has two major conformers ( $\sim 40\%$  each), trans-trans main chain geometry and cis-trans. There are also two minor conformers, trans-cis and cis-cis, where the assignments of the 2 Nmdb HF1 resonance could not be made unambiguously ( $\sim 10\%$  each). (C) CHIR-2279 in DMSO, 30  $^\circ\text{C}$ . Integrated values are normalized to 4 for the estimation of populations based on four resonances, as in (A). These peak integrations indicate that in DMSO, CHIR-2279 now occupies two major conformers equally, trans-trans and trans-cis ( $\sim 45\%$  each based on the 2 ac HA peaks at 4.5 ppm). Note that there are some broad unassigned peaks in this region that may be indicative of short-lived states where the cis geometry is occupied at the first amide bond.

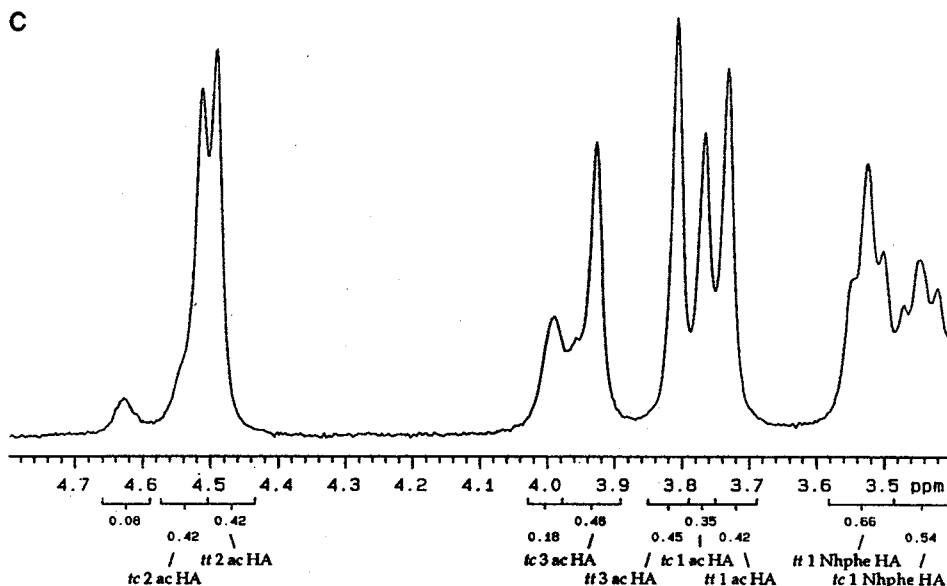
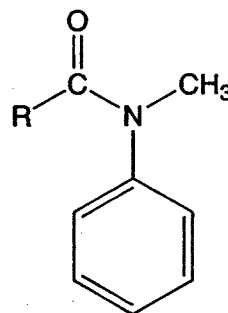


Figure 7. (Continued).

seen for CHIR-4531. But of more interest, the data presented here demonstrate that there is a continuous range of possible states, from one major main chain geometry (CHIR-2279 in aqueous co-solvent), through two major geometries (CHIR-2279 in organic solvent), two major and two minor geometries (CHIR-4531 in aqueous co-solvent), and finally four equally populated geometries (CHIR-4531 in organic solvent). The challenge then becomes determining what physical forces determine the population distribution. The ultimate goal is to gain some insight into bioactive conformations by understanding these contributions under the conditions where the compounds interact with their receptor/target.

The most promising experimental evidence that different NSGs would have different conformational preferences was the predominance of one main chain geometry for CHIR-2279 in aqueous co-solvent. In fact, the ROESY data for CHIR-2279 in DMSO have only two observable species, which are both trans at the first (aromatic) amide bond, unlike CHIR-4531, which occupies both cis and trans in populations at the first amide under both solvent conditions. This difference in amide bond geometries can be explained by differences in the amides themselves. The first amide of CHIR-2279 is aromatically substituted at the nitrogen. The presence of the aromatic ring of the Nbphe side chain attached to the amide nitrogen (Figure 1) alters the energy difference between cis and trans amide. Pre-

Figure 8. *N*-Methyl-*N*-acyl anilide model compounds used for X-ray crystal structure determination by Itai et al. [14,15].

viously, studies [14,15] with model compounds suggested that this type of amide bond prefers a geometry equivalent to the trans main chain of an NSG.

Itai et al. [15] studied a series of model compounds of the form *N*-acyl-*N*-methyl anilides (Figure 8). Determining the crystal structures, they found amide N-C bond lengths of  $\sim 1.35$  Å and a nitrogen substituent geometry consistent with  $sp^2$  hybridization. These two observations are consistent with the partial double bond nature of the N-C bond in the amide. They also reported that interplanar angles between the phenyl ring and the amide group ranged from  $60^\circ$  when R was a benzyl ring, to  $90^\circ$  for alkyl substitutions like isopropyl and methyl. These values led Itai et al. [15] to conclude that there is no longer conjugation between the imi-

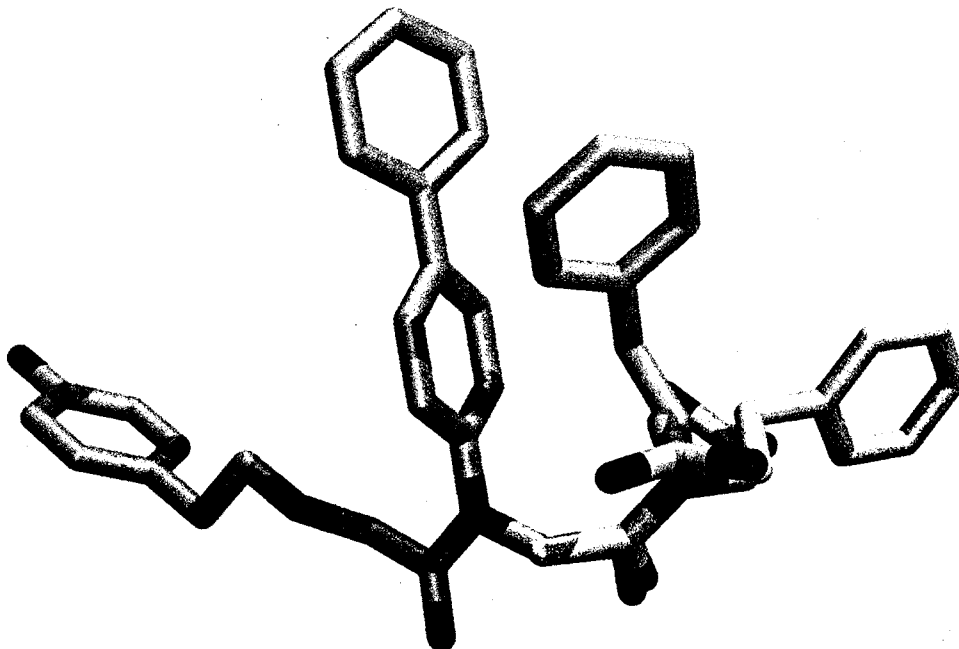


Figure 9. Two superimposed models of CHIR-2279. Carbon atoms are shown in gray (and light blue), nitrogen in blue, and oxygen in red. As stated in the text, there is not a single conformation in solution, but an ensemble. This was also reflected in molecular modeling (results not published) where a large number of low-energy conformations were seen. The two models shown here were chosen to demonstrate how differences in the amide bond geometry between positions 2 and 3 can affect hydrophobic packing. The major species (85%) in aqueous co-solvent has a trans amide, shown in the completely gray structure. The minor species (15%) has a cis amide bond in this position, shown by the light blue structure starting at the amide between 2 and 3.

no nitrogen and the aromatic ring. This conjugation is probably sacrificed to avoid steric hindrance between the phenyl group and the aliphatic group adjacent to the carbonyl. All of the compounds adopt this conformation with the R group trans to the methyl and cis to the aromatic ring (Figure 8). This is exactly what we see for CHIR-2279, where 1 ac occupies the R position and 2 ac takes the place of the methyl.

Aromatic substitution explains the trans geometry preference at the first amide of CHIR-2279. But what is causing the population differences in aqueous co-solvent for the other amide bonds? Under partially aqueous conditions, CHIR-2279 prefers trans amide geometry at the second amide bond (85:15), and CHIR-4531 prefers trans amide geometry at the second amide bond (40:10). As noted above, steric interactions are the main determining factor in peptides preferring trans main chain geometries. Since the bulk of a proton is much less than a CH, the main chain  $\alpha$ -carbons end up being trans to each other. This cannot be the explanation for NSGs, since the main chain and side chains are roughly the same bulk at their attachment to the nitrogen, both being methylenes ( $\text{CH}_2$ ).

Another factor that has been proposed as being important to unique folded conformations in peptides is 'hydrophobic collapse'. Wiley and Rich [7] have used this term to describe the result of molecules minimizing their exposed hydrophobic surfaces in water by packing their hydrophobic groups together. They have proposed that small flexible molecules will undergo hydrophobic collapse to form intramolecular hydrophobic interactions in aqueous solutions that are not present in organic solvents. They have also suggested that neither organic solution structures nor in vacuo simulations are accurate or relevant models of the ligand conformation in aqueous solution or bound to the receptor.

One example of this is the cyclosporin/cyclophilin story. The solution structures of cyclosporin determined in organic solvents [16–18] have been shown to be very different from that of cyclosporin complexed with cyclophilin [19–21], in aqueous solutions. In fact, the cyclophilin bound structure is very similar to an antibody bound form, which presumably reflects the aqueous solution structure [22]. Wiley and Rich propose that flexible ligands like cyclosporin undergo

Table 2.  $^{13}\text{C}$  and  $^1\text{H}$  chemical shift<sup>a</sup> assignments of CHIR-4531 in 25%  $\text{CH}_3\text{CN}/75\%$   $\text{D}_2\text{O}$  at 10 °C

Submonomer	trans-trans		cis-trans		
	$^{13}\text{C}$	$^1\text{H}$	$^{13}\text{C}$	$^1\text{H}$	
1 Ndpe	A	50.1	3.77	49.83.74	
	B	46.7	4.44	46.6	4.35
	C	138.6		138.6	
	D	126.5	7.39	126.4	7.35
	E	128.4 <sup>b</sup>	7.39 <sup>b</sup>	128.4 <sup>b</sup>	7.35 <sup>b</sup>
	F	126.9 <sup>b</sup>	7.31 <sup>b</sup>	126.9 <sup>b</sup>	7.31 <sup>b</sup>
1 ac	A	46.3	3.65	46.5	3.96
	CO	165.3		165.1	
2 Nmdb	A	49.3	4.20	50.0	4.11
	B	128.1		127.2	
	C1	107.2	6.70	106.1	6.64
	C2	120.6	6.64	119.3	6.59
	D1	146.5		146.8	
	D2	107.3	6.86	107.7	6.90
	E2	145.6		145.9	
	F1	100.3	6.02	100.5	5.96
2 ac	A	46.7	3.69	46.7	3.81
	CO	167.6		168.5	
3 Nhtyr	A	49.0	3.46	49.0	3.52
	B	31.4	2.70	31.4	2.73
	C	128.3		128.2	
	D	129.3	7.00	129.3	6.97
	E	114.6	6.76	114.4	6.60
	F	153.7		153.7	
3 ac	A	47.5	4.09	47.6	4.05
	CO	171.4		171.7	

<sup>a</sup> Units of ppm  $^{13}\text{C}$  referenced to 117.7 ppm  $\text{CH}_3\text{CN}$ ;  $^1\text{H}$  referenced to 0.0 ppm TSP.

<sup>b</sup> Estimated because of severe peak overlap.

Table 3.  $^{13}\text{C}$  and  $^1\text{H}$  chemical shift<sup>a</sup> assignments of CHIR-2279 in DMSO at 30 °C

Submonomer	trans-trans		trans-cis		
	$^{13}\text{C}$	$^1\text{H}$	$^{13}\text{C}$	$^1\text{H}$	
1 Nhtyr	A	47.0	3.05	47.0	3.05
	B	29.8	2.78	29.8	2.78
	C	125.9		125.9	
	D	128.5	6.97	128.5	6.97
	E	114.4	6.71	114.4	6.71
	F	155.0		155.0	
1 ac	A	46.6	3.74	46.6	3.77
	CO	164.2		164.2	
2 Nbiph	A	139.0		138.6	
	B	127.9	7.47	127.9	7.52
	C	126.8	7.77	126.8	7.79
	D	139.0		139.0	
	E	139.0		139.0	
	F	125.6	7.71	125.6	7.74
	G	127.9	7.50	127.9	7.50
	H	126.7	7.44	126.7	7.44
2 ac	A	50.1	4.50	50.1	4.52
	CO	166.1		165.9	
3 Nhpe	A	48.5	3.53	48.5	3.45
	B	32.9	2.87	32.0	2.75
	C	137.4		138.0	
	D	127.6	7.23	127.6	7.23
	E	127.7	7.25	127.7	7.25
	F	125.3	7.17	125.3	7.17
3 ac	A	48.8	3.81	47.5	3.94
	CO	168.8		169.0	

<sup>a</sup> Units of ppm  $^{13}\text{C}$  referenced to 117.7 ppm  $\text{CH}_3\text{CN}$ ;  $^1\text{H}$  referenced to 0.0 ppm TSP.

'hydrophobic collapse' when dissolved in water, and, like cyclophilin, the bioactive conformation would be similar to the aqueous conformation. So, the organic solution structures would not be a good model for drug design.

The NSGs described here are very likely to undergo hydrophobic collapse. They are very flexible molecules with very hydrophobic side chains. However, there is no direct evidence from ROESY data to support this. Neither CHIR-2279 nor CHIR-4531 shows cross peaks between their hydrophobic side chains. The ROESY spectra contain only short-range cross peaks. This is consistent with the results found for small peptides. Peptides are usually present in solution as an ensemble

of conformations, with relatively free rotation about the phi and psi angles [23].

Initially, it might not appear that the main chain geometry preferences result from hydrophobic collapse. However, the shift in conformer populations when the solvent conditions are altered is indirect evidence of hydrophobic collapse. In the aqueous co-solvent system of 25%  $\text{CH}_3\text{CN}$ , there is a preference for certain main chain geometries, which can be seen as population differences in the 1D spectra. In an aqueous co-solvent system, both CHIR-2279 and CHIR-4531 have preferred amide bond geometries. CHIR-2279 has one preferred main chain geometry (~85%), and CHIR-4531 has two (~40% each).

In DMSO, the hydrophobic effect is not a driving force, and both the polar main chain and the hydrophobic side chains are fully solvated. When the same NMR analysis is performed in this organic solvent, there is equal occupancy of amide bond geometries. CHIR-2279 shifts from an 85:15 preference to 50:50 for the amide bond that is not aromatically substituted at the nitrogen. In DMSO, CHIR-4531 has no main chain conformational preference at either amide bond. Each combination of amide geometries is occupied equally, trans-trans (~25%), trans-cis (~25%), cis-cis (25%), and cis-trans (25%). This shift in conformer populations suggests that although there is not a unique conformation, 'hydrophobic collapse' [7] does occur for these small molecules. This hydrophobic driving force limits the sampling of conformational space by these molecules, which in turn favors certain combinations of amide bond geometries over others. This is suggested by the model of CHIR-2279 in Figure 9. The trans amide bond geometry results in conformations where the side chains of residues 2 and 3 can approach each other closely, facilitating hydrophobic collapse, whereas the cis amide bond has a more extended main chain, possibly preventing or obstructing association of the hydrophobic side chains. The effect of hydrophobic collapse on the distribution of amide bond geometries, and limiting of side chain conformational space, is one compelling reason to pursue the structure of NSGs, or any other small molecule of pharmaceutical interest, in at least partially aqueous solutions.

We have shown that it is possible to probe for hydrophobic collapse of small flexible molecules, like NSGs, by comparing NMR data collected under different solvent conditions. This method provided us with intriguing experimental evidence that these NSGs undergo hydrophobic collapse. The changes in the occupancy of the different amide bond geometries, under different solvent conditions, confirm that aqueous solvents may favor certain geometries, and possibly conformations, over others. Thus, hydrophobic collapse results in a restriction of conformational space, but the lack of ROESY cross peaks indicates there is still averaging in that space. The hydrophobic collapse theory of Wiley and Rich [7] would further suggest that the bioactive conformations are contained in this restricted space.

## Acknowledgements

The authors would like to thank Dr. I.D. Kuntz and Dr. G.A. Montelione for insightful discussions.

## References

1. Simon, R.J., Kania, R.S., Zuckermann, R.N., Huebner, V.D., Jewell, D.A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C.K., Spellmeyer, D.C., Tan, R., Frankel, A.D., Santi, D.V., Cohen, F.E. and Bartlett, P.A., *Peptoids - A modular approach to drug discovery*, Proc. Natl. Acad. Sci. USA, 89 (1992) 9367-9371.
2. Zuckermann, R.N., Kerr, J.M., Kent, S.B.H. and Moos, W.H., *Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis*, J. Am. Chem. Soc., 114 (1992) 10646-10647.
3. Zuckermann, R.N., Martin, E.J., Spellmeyer, D.C., Stauber, G.B., Shoemaker, K.R., Kerr, J.M., Figliozzi, G.M., Goff, D.A., Siani, M.A., Simon, R.J., Banville, S., Brown, E.G., Wang, L., Richter, L.S. and Moos, W.H., *Discovery of nanomolar ligands for 7-transmembrane G-protein coupled receptors from a diverse (N-substituted) glycine peptoid library*, J. Med. Chem., 37 (1994) 2678-2685.
4. Bradley, E.K., *A method for sequential NMR assignments of <sup>1</sup>H and <sup>13</sup>C resonances of N-substituted glycine peptoids*, J. Magn. Reson., B110 (1996) 195-197.
5. Bax, A. and Summers, M.F., *<sup>1</sup>H and <sup>13</sup>C assignments for sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR*, J. Am. Chem. Soc., 107 (1986) 2093-2094.
6. Mueller, L., *Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence*, J. Am. Chem. Soc., 101 (1979) 4481-4484.
7. Wiley, R.A. and Rich, D.H., *Peptidomimetics derived from natural products*, Med. Res. Rev., 13 (1993) 327-384.
8. Atkinson, R.A. and Pelton, J.T., *Conformational study of cyclo[D-Trp-D-Asp-Pro-D-Val-Leu], and endothelin-A receptor-selective antagonist*, FEBS Lett., 296 (1992) 1-6.
9. Hwang, T.-L. and Shaka, A.J., *Cross relaxation without TOCSY: Transverse rotating-frame Overhauser effect spectroscopy*, J. Am. Chem. Soc., 114 (1992) 3157-3159.
10. Day, M., Kneller, D. and Kuntz, I.D., *NMR Pack: The programs are available from the Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, U.S.A., 1993.*
11. Marion, D. and Wüthrich, K., *Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurement of <sup>1</sup>H-<sup>1</sup>H spin coupling constants in proteins*, Biochem. Biophys. Res. Commun., 113 (1983) 967-974.
12. Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K., *Improved spectral resolution in COSY <sup>1</sup>H NMR spectra of proteins via double quantum filtering*, Biochem. Biophys. Res. Commun., 117 (1983) 479-485.
13. Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.L., *Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame*, J. Am. Chem. Soc., 106 (1984) 811-813.

14. Itai, A., Toriumi, Y., Tomioka, N., Kagechika, H., Azmaya, I. and Shudo, K., *Stereochemistry of N-methylbenzanilide and benzanilide*, *Tetrahedron Lett.*, 30 (1989) 6177-6180.
15. Itai, A., Toriumi, Y., Saito, S., Kagechika, H. and Shudo, K., *Preference of cis-amide structure in N-acyl-N-methylanilines*, *J. Am. Chem. Soc.*, 114 (1992) 10649-10650.
16. Kessler, H., Loosli, H.-R. and Oschkinat, H., *Assignment of the  $^1\text{H}$ -,  $^{13}\text{C}$ -, and  $^{15}\text{N}$ -NMR spectra of cyclosporin A in  $\text{CDCl}_3$  and  $\text{C}_6\text{D}_6$  by a combination of homo- and heteronuclear two-dimensional techniques*, *Helv. Chim. Acta*, 68 (1985) 661-681.
17. Loosli, H.-R., Kessler, H., Oschkinat, H., Weber, H.-P., Petcher, T.J. and Widmer, A., *The conformation of cyclosporin A in the crystal and in solution*, *Helv. Chim. Acta*, 68 (1985) 682-701.
18. Kessler, H., Gehrke, M., Lantz, J., Kock, M., Seebach, D. and Thaler, A., *Complexation and medium effects of the conformation of cyclosporin A studied by NMR spectroscopy and molecular dynamics calculations*, *Biochem. Pharmacol.*, 40 (1990) 169-173; erratum, 2185-2186.
19. Fesick, S.W., Gampe, R.T., Holzman, T.F., Egan, D.A., Edalji, R., Luly, J.R., Simmer, R., Helfrich, R., Kishore, V. and Rich, D.H., *Isotope-edited NMR of cyclosporin A bound to cyclophilin: Evidence for a trans 9,10 amide bond*, *Science*, 250 (1990) 1406-1409.
20. Fesick, S.W., Gampe, R.T., Eaton, H.L., Gemmecker, G., Olejniczak, E.T., Neri, P., Holzman, T.F., Egan, D.A., Edalji, R., Simmer, R., Helfrich, R., Hochlowski, J. and Jackson, M., *NMR studies of [ $U$ - $^{13}\text{C}$ ]cyclosporin A bound to cyclophilin: Bound conformation and portions of cyclosporin involved in binding*, *Biochemistry*, 30 (1991) 6574-6583.
21. Weber, C., Wider, G., von Freyberg, B., Traber, R., Braun, W., Widmer, H. and Wüthrich, K., *The NMR structure of cyclosporin A bound to cyclophilin in aqueous solution*, *Biochemistry*, 30 (1991) 6563-6574.
22. Altschuh, D., Vix, O., Rees, R. and Thierry, J.-C., *A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex*, *Science*, 256 (1992) 92-94.
23. Dyson, J.H. and Wright, P.E., *Defining solution conformations of small linear peptides*, *Annu. Rev. Biophys. Biophys. Chem.*, 20 (1991) 519-538.