Structural Properties of N-alpha-chiral Peptoids

by

Philippe Armand

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

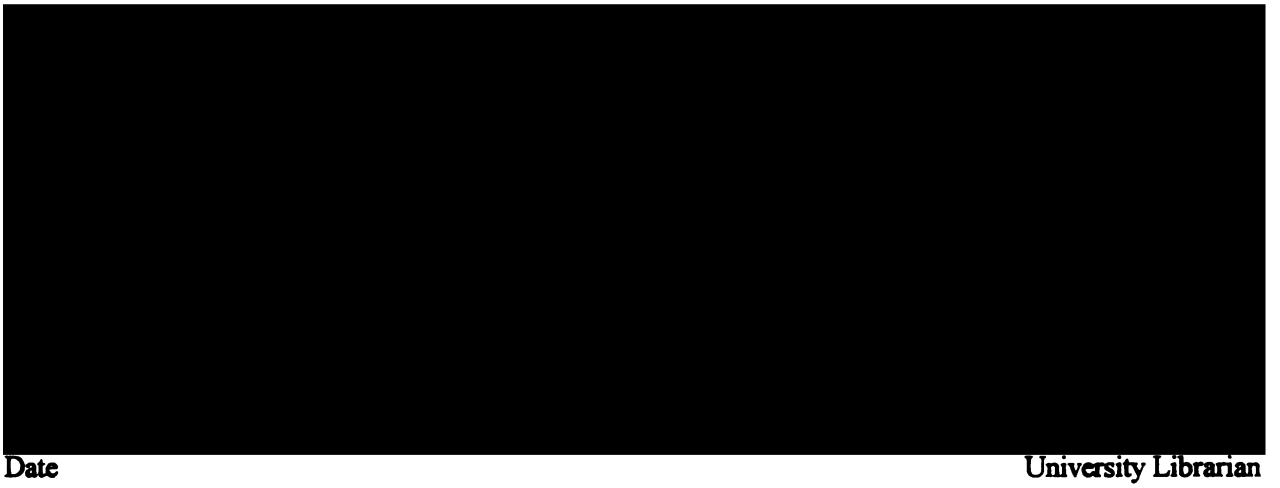
BIOMEDICAL SCIENCES

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO



Date University Librarian

Degree Conferred: .....

To Fred, with more thanks than this page can hold

Acknowledgements

There is one quality in Fred that has made it such a pleasure to spend the last four years working with him: it is his ability to provide at the same time complete freedom and unconditional support.

Behind this sentence hide four years of learning, four years of searching, four years of growing, and four years of my life. Yet the words lie cold in their ink, plain in their meaning. How can so much pleasure, so much laughter, so much friendship, so much learning, and so many rewards all blend in one black color and in so few letters? And yet how could I better capture what makes Fred so unique? There is no easy analogy to this quality, since any object that supports must somehow hinder. Nor is there an anecdote that distills it. Yet everything follows from it: the friendly, open, relaxed work environment; the cheerful, friendly, intelligent, helpful people which Fred draws; the breadth of the science, the growth of new projects and new inspiring visions; the development of science and the development of people. And so, since I cannot hope to lock into words the full meaning of this quality, nor to ever fully acknowledge my debt to it, I will state it again.

There is one quality in Fred that has made it such a pleasure to spend the last four years working with him: it is his ability to provide at the same time complete freedom and unconditional support.

I also wish to acknowledge my beloved Wynne; my happiness is the sparkle in her smile, and my world shines in her eyes.

Then there are all of those whom I am so privileged to have met in the course of my graduate work, and to whom I am so indebted. Kent, a fantastic person, a fantastic friend; Erin, for her integrity, her help, and her understanding; Dale, for an unforgettable week and a friendship which I hope never ends; Ron, for his patience and chemistry expertise; Eric, for his support and his accent; Ginger, for her wizard-like admin skills; and Steve, for his wizard-like UNIX skills. And thanks to the old gang, for all the fun and games: Nathan, Mike, Tom, Olivier, Alexis, I wish you were still here.

Structural Properties of Nα-chiral Peptoids

Philippe Armand

Abstract

Heteropolymers of N-substituted glycines ("peptoids") are a new class of synthetically accessible molecules which were first developed for use in drug discovery projects. Recent progress in the synthetic method now allow the creation of peptoid molecules up to 50 residues in length, with precisely defined sequences composed of a large variety of monomers. Here we report the theoretical and experimental characterization of peptoid oligomers containing bulky side chains with chiral centers at the N-α position (Nα-chiral peptoids). Circular dichroic spectroscopy indicates that the incorporation of such side chains leads to the adoption of extremely stable helical conformations in organic and aqueous solutions. Computational modelling suggests that the presence of the bulky chiral center gives rise to predictable preferences for amide bond geometry, rotamer choice, and backbone handedness. All of these preferences were confirmed by the structure of a peptoid heteropentamer, obtained by nuclear magnetic resonance spectroscopy. Together, our results help to define the structural properties of Nα-chiral peptoids. We also describe an attempt to use this knowledge as a basis for designing a peptoid helix bundle.

Table of Contents

List of Tables

List of Figures

Introduction

It is remarkable (and so well recognized as to hardly need emphasis) that proteins play such central and diverse roles in sustaining life, while having such a small number of building blocks. Functions as diverse as motion, metabolism, and immunologic defense can be carried out by molecules composed of different combinations of the same 20 amino acids. This functional diversity depends on the ability of proteins to bind to different targets, whether they be an actin filament in the case of myosin, glucose in the case of hexokinase, or a bacterial toxin in the case of the antibody that protects us against tetanus. And since the chemical groups that participate in all these interactions are similar, it follows that the unequaled power and versatility of proteins depends on their ability to position these chemical groups at defined spatial locations, or, more accurately, on their ability to encode in their primary sequence their exact three-dimensional conformation. This property of sequence-specific structure, then, is what allows proteins to be the backbone of life.

It is therefore tempting to look for this same property in artificial systems. Indeed, a system that could imitate even a fraction of proteins' functions would likely prove invaluable in biological or chemical applications. Such a "biomimetic" polymer, that is, a polymer able to carry out specific biological functions, should have two properties: sequence-specific synthesis, meaning that its synthesis should allow exact sequence control while incorporating a variety of monomer units; and sequence-specific structure, meaning that at least some of the sequences should fold to specific three-dimensional structures.

To date, most polymers that have been synthesized are composed of at most a few different monomers. Since there is generally no fine control over their sequence or chain length, these polymers lack well-defined three-dimensional structures. And while they have found a large number of applications for their bulk properties, the absence of unique, well-defined structures prevents them from performing functions that require molecular recognition. Hence these polymers are ill-suited for biological or biochemical purposes.

Several attempts to create structured polymers have been published. Table 1.1 summarizes important information in some of the cases where a sequence-specific heteropolymer has been synthesized and has yielded evidence of specific structure. Most of these have been shown to form hydrogen-bonded short helical segments or sheets in organic solvents. Aedamers (3) are an exception in that they form structures in aqueous solvent and without the help of hydrogen bonds. Phenylacetylenes may also have form helices without hydrogen bonds, but no structure has been obtained yet. β-peptides (6,7,8,9) are a particularly interesting case; they can form both sheets and helices, and, depending on the monomers, can form helices of different periodicities and pitch.

Table 1.1: Structured, sequence-specific heteropolymers

Polymer	Max. length	Structure determined[1]	Solubility[2]	H-bonds	Reference
Vinylogous Peptides	3	X, N	O	yes	1
Sulfonamido-pseudopeptides	4	N	O	yes	2
Aedamers	3	N	A	no	3
Oligoanthranilamides	5[3]	X, N	O	yes	4
Phenylacetylenes	18	none	O	no	5
β-peptides	12	X, N	O	yes	6,7

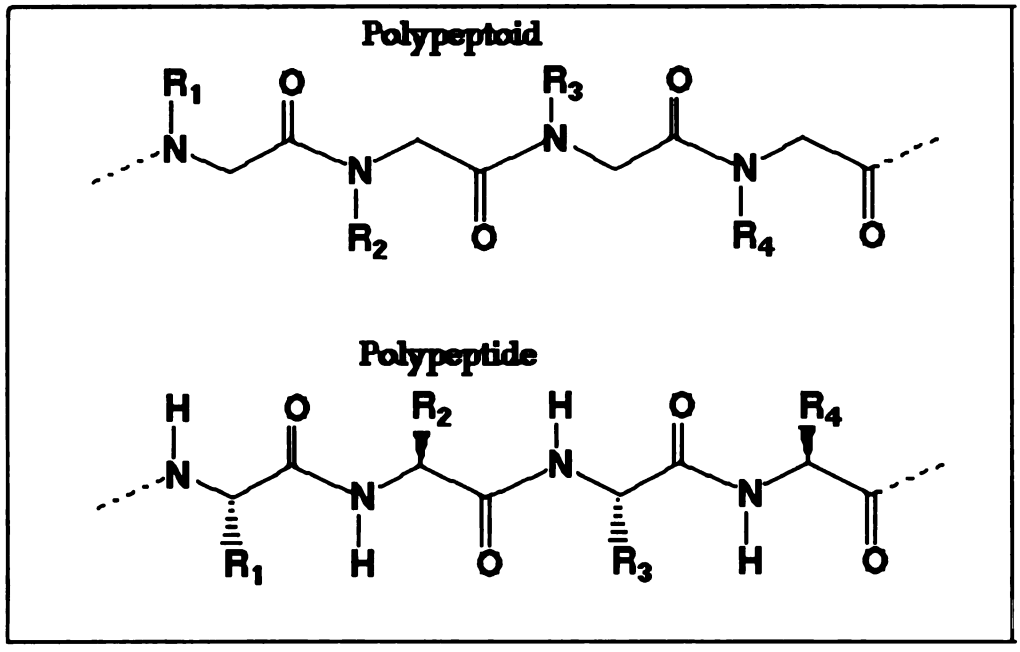
[1]X=X-ray crystallography, N=nuclear magnetic resonance (NMR) spectroscopy

[2]O=Organic solvent, A=aqueous solvent

[3]This includes the turn units

Heteropolymers of N-substituted glycines ("peptoids") are another class of heteropolymers and are the subject of this work. Fig. 1.1 compares a prototypical polypeptoid with a polypeptide of equivalent length.

Figure 1.1: Comparison of polypeptoid and polypeptide



The differences in the chemical architectures of peptoids and peptides have four major consequences. The substitution of the nitrogen at the amide bond should stabilize the molecule against enzymatic proteolytic degradation. This has been experimentally confirmed for a peptoid octamer which was resistant to cleavage by chymotrypsin, papain or thermolysin (10). Degradation by circulating proteases is one of the main reasons why proteins are rarely used as therapeutic agents *in vivo*. The protease resistance of peptoids, by contrast, could make them attractive candidates for *in vivo* use.

Another difference is in the conformational properties of polypeptoids, which must differ from those of polypeptides because of the change in the side chain origin. This will be discussed in Chapter 1.

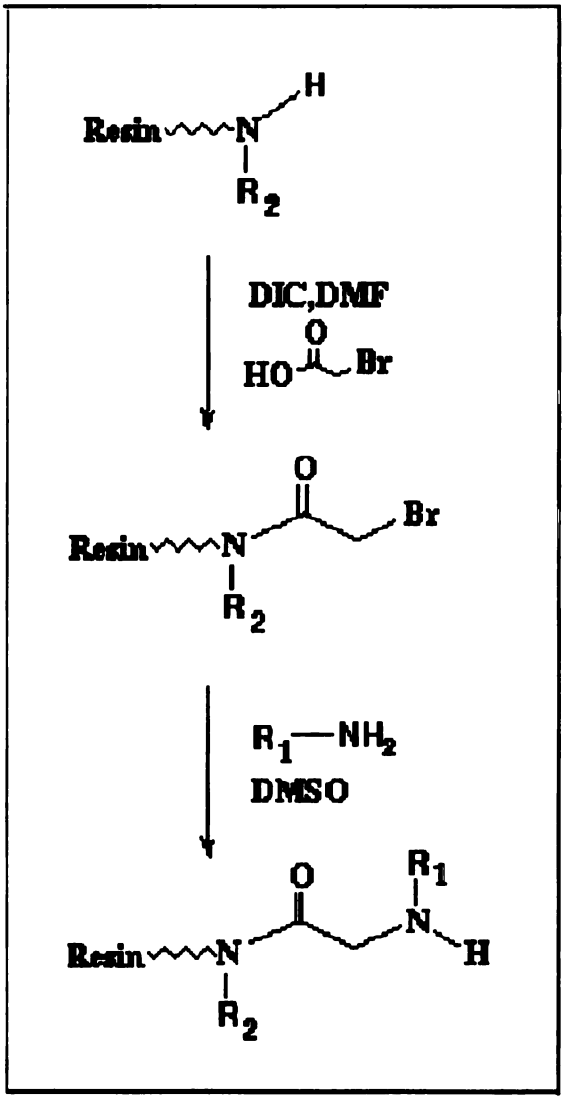
The third difference is the absence of hydrogen bond donors in the backbone of peptoids. This stands in contrast to polypeptides and to most other structured abiotic polymers (Table 1.1) where hydrogen bonds seem to play an important role in determining the structure. While it may be possible to from hydrogen bonds in peptoids by providing the donor functionality in the side chain, this has not yet been pursued.

A final difference is that the backbone of peptoids has no chiral centers, since the main chain α carbons are unsubstituted. In the biological realm, most molecules of interest have chiral three-dimensional structures (all proteins do), and hence molecules designed to interact specifically with biological molecules should probably also have chiral structures. While it is conceivable that such chiral structures as helices could form in an achiral polymer (11) simply by the enthalpic obstacles to handedness reversal in the middle of a helix, this would require very long polypeptoids, and the helical conformations would probably be highly kinked and sparsely populated. It should also be possible to induce chiral structure in an achiral polymer by the use of a chiral solvent; however, solvents of biological interest are achiral, and the necessity of using a chiral solvent would drastically limit the application potential of the polymer. Thus the most promising avenue to drive the formation of chiral structures in peptoids is to incorporate chiral centers in the side chain.

Short-length peptoids were first synthesized for use in drug-discovery programs (10,12). Recent improvements in the synthetic approach now allow

the synthesis of peptoids up to 30 residues in length with high yields and good purity (13). Indeed, simple 30-mers are routinely obtained with purities >95%. The synthetic protocol is outlined in Fig. 1.2.

Figure 1.2: Peptoid synthesis



Peptoids are synthesized on a solid-phase support by submonomer addition: bromoacetic acid is first coupled to the N terminus of the growing chain; a primary amine is then used to displace the bromine. One major advantage of this method is that, in theory, any primary amine can be used in

the displacement step. Since the nitrogen of the amine becomes part of the backbone, while the rest of the amine becomes the side chain of the peptoid residue, any side chain which can be obtained as the primary amine should be usable in the synthesis. Naturally, the amine must be soluble and stable in organic solvent, must have the appropriate protection groups on reactive moieties (e.g., other amine groups, carboxyl groups, etc), and must be stable once incorporated. But even these provisions leave a huge potential number of side chain candidates which can provide a variety of functionalities to the peptoid molecule.

Another advantage of this synthetic route is that it eliminates the need for protecting chemistry at the end of the growing chain. Thus, unlike in protein synthesis where each new amino acid is added with a protected N terminus, neither the bromoacetic acid nor the displacing amine needs to be protected. This speeds up the cycle time and decreases the cost of the synthesis.

A final advantage is that it is easy to add ^{13}C labels to the backbone, by the use of ^{13}C-labeled bromoacetic acid, which is readily available. This proved useful for the structural studies described in Chapter 3.

Thus peptoids fulfill the first criterion for biomimetic polymers defined above. Their synthesis is sequence-specific and allows the incorporation of a variety of monomers. Moreover, there is already evidence of sequence-specific function, as cationic peptoids can be efficient DNA transfection agents, alone (14) or conjugated to lipid moieties (15). Here we address the question of whether peptoids also display sequence-specific structure.

In order to find structured peptoids, we synthesized molecules containing chiral side chains. As mentioned above, the presence of chiral centers in the molecules is a likely prerequisite to the formation of chiral structures in achiral solvents. It also allowed us to screen the molecules using circular dichroic (CD) spectroscopy.

The most promising result was the CD spectra of homopolymers of (*S*)-*N*-(1-phenylethyl)glycine, or *N*spe. An *N*spe hexamer is shown in Fig. 1.3, and its CD spectrum in Fig. 1.4. Nspe monomers contain a chiral center at the N-α carbon (the first carbon of the side chain). We will refer to side chains containing chiral centers at this position as "Nα-chiral" side chains, and to peptoids incorporating such side chains as "Nα-chiral" peptoids.

Figure 1.3: *N*spe hexamer

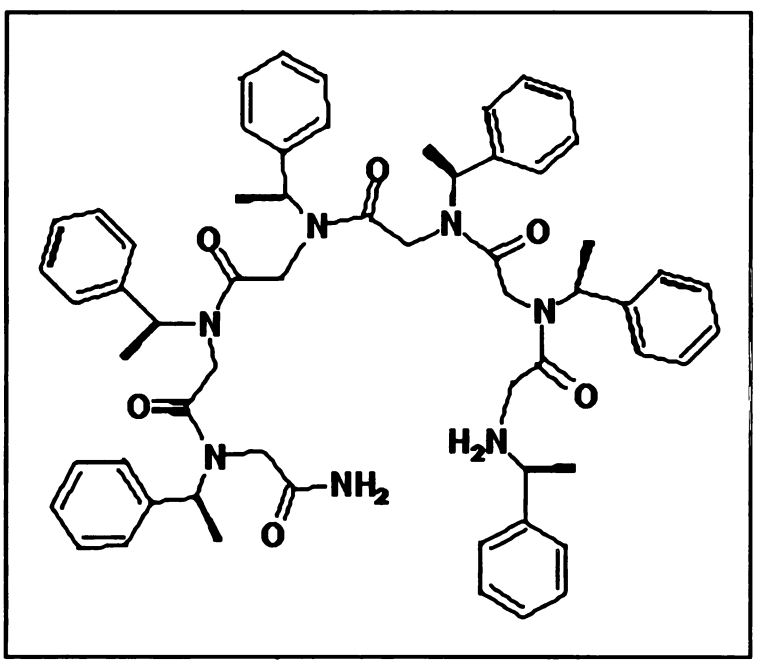
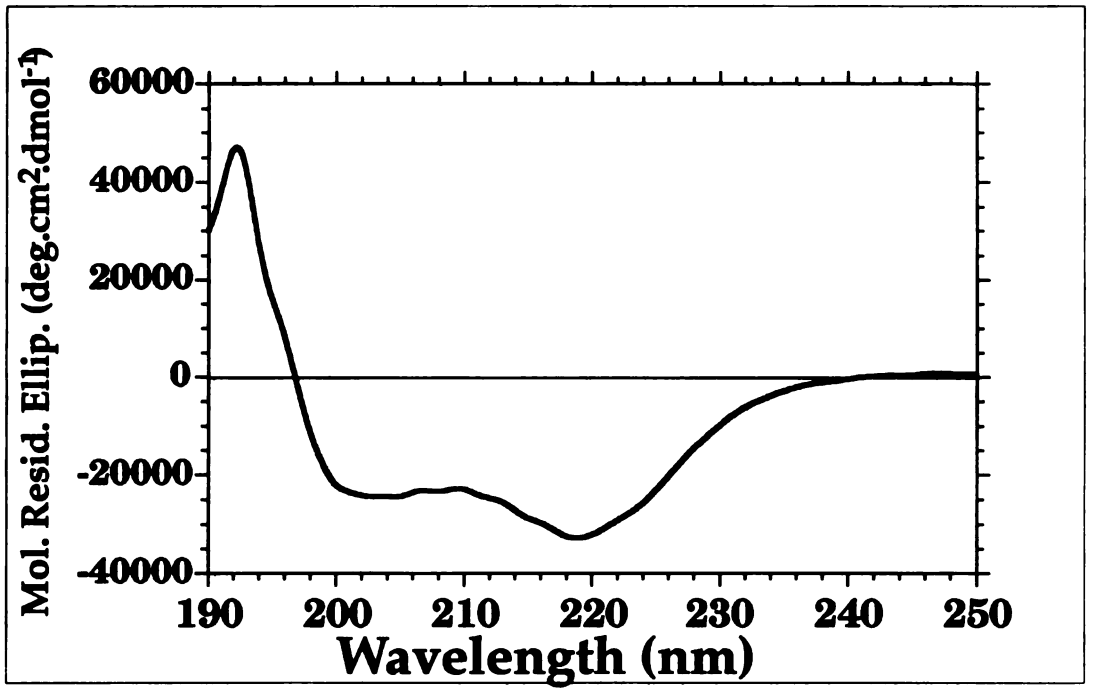


Figure 1.4: CD spectrum of (*N*spe)$_6$ in methanol, at a concentration of 1mM[1]



[1]This spectrum was collected using the method described in Chapter 2 (Methods).

The CD spectrum of (*N*spe)$_6$ displays a double minimum with nadirs at 200 and 220 nm, and a maximum at 190 nm. These features are strongly reminiscent of the CD spectra of polypeptide α helices, which have a double minimum at 208 and 222 nm, and an accompanying maximum at 192 nm (16). Thus we hypothesized that *N*spe oligomers form helices in solution.

In Chapter 1, we investigate the conformational properties of *N*spe containing peptoids, using a combination of molecular mechanics and semi-empirical quantum mechanical calculations. We also propose a model for the structure of an Nspe octamer.

In Chapter 2, we present further CD evidence that *N*spe oligomers and related molecules do indeed adopt a helical conformation, some in aqueous and some in organic solvents. We also investigate the thermodynamic properties of this helix by CD and calorimetry.

In Chapter 3, we report the solution structure for the major isomer of a peptoid heteropentamer containing side chains similar to *N*spe, in methanol, using nuclear magnetic resonance (NMR) spectroscopy. We also compare the structure thus obtained with the model proposed in Chapter 1.

Finally, in Chapter 4, we describe an attempt to use the information garnered in Chapters 1 to 3 to create a peptoid helix bundle, using a computational strategy to search sequence and structure space simultaneously. We present the preliminary experimental results obtained on the first set of bundle candidates.

References

1. Hagihara M., Anthony N. J., Stout T. J., Clardy J., Schreiber S. L. (1992) *J. Am. Chem. Soc.* **114**:6568

2. Gennari C., Salom B., Potenza D., Williams A. (1994) *Angew. Chem. Int. Ed. Engl.* **33**:2067

3. Lokey R. S., Iverson B. L. (1995) *Nature*, **375**:303

4. Hamuro Y., Geib S. J., Hamilton A. D. (1996) *J. Am. Chem. Soc.* **118**:7529

5. Nelson J. C., Saven J. G., Moore J. S., Wolynes P. G. (1997) *Science* **277**:1793

6. Seebach D., Matthews J. L. (1997) *Chem. Commun.* , pp. 2015-2022

7. Appella D. H., Christianson L. A., Klein D. A., Powell D. R., Huang X., Barchi J. J., Gellman S. H. (1997) *Nature* **387**:381

8. Appella D. H., Christianson L. A., Karle I. L., Powell D. R., Gellman S. H. (1996) *J. Am. Chem. Soc.* **118**:13071

9. Seebach D., Overhand M., Kuhnle F. N. M., Martinoni B. (1996) *Helv. Chem. Acta* **79**:913

10. 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., Bartlett P. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**:9367

11. Green M. M., Peterson N. C., Sato T., Teramoto A., Cook R., Lifson S. (1995) *Science* **268**:1860

12. Zuckermann R. N., Kerr J. M., Kent S. B. H., Moos W. H. (1992) *J. Am. Chem. Soc.* **114**:10646

13. Zuckermann R. N., Martin E. J., Spellmeyer D. C., Satuber G. B., Shoemaker K. R., Kerr J. M., Figliozzi G. M., Goff D. A., Siani M. A., Simon R. J., Banville S. C., Brown E. G., Wang L., Richter L., Moos W. H. (1994) *J. Med. Chem.* **37**:2678

14. Murphy J. E., Uno T., Hamer J. D., Cohen F. E., Dwarki V., Zuckermann R. N. (1997) *Proc. Natl. Acad. Sci. USA* **95**:1517

15. Huang C.-Y., Uno T., Murphy J. E., Lee S., Hamer J. D., Escobedo J. A., Cohen F. E., Radhakrishnan R., Dwarki V., Zuckermann R. N. (1998) *Chemistry & Biology* **5**:345

16. Woody R. W. (1985) in *The Peptides*, ed. Hruby, V. J. (Academic, Orlando, FL), **7**:16

Chapter 1:

Conformational properties of Nα-chiral peptoids

This chapter was published as:

Armand P., Kirshenbaum K., Falicov A., Dunbrack R. L. Jr, Dill K. A., Zuckermann R. N., Cohen F. E. (1997) "Chiral N-substituted glycines can form stable helical conformations" Folding *& Design* **2**:369

Abstract

Background: Short sequence-specific heteropolymers of N-substituted glycines (peptoids) have emerged as promising tools for drug discovery. Recent work on medium-length peptoids containing chiral centers in their sidechains has demonstrated the existence of stable chiral conformations in solution. In this report, we explore the conformational properties of these N-α chiral peptoids by molecular mechanics calculations and propose a model for the solution conformation of an octamer of (*S*)-*N*-(1-phenylethyl)glycine.

Results: Molecular mechanics calculations indicate that the presence of N-substituents in which the Nα carbons are chiral centers has a dramatic impact on the available backbone conformations. These results are supported by semi-empirical quantum mechanical calculations, and coincide qualitatively with simple steric considerations. They suggest that an octamer of (*S*)-*N*-(1-phenylethyl)glycine should form a right-handed helix with *cis* amide bonds, similar to the poly-proline type I helix. This model is consistent with circular dichroism studies of these molecules.

Conclusions: Peptoid oligomers containing chiral centers in their sidechains present a new structural paradigm which has promising implications for the design of stably folded molecules. We expect that their novel structure may provide a scaffold to create heteropolymers with useful functionality.

Introduction

Peptoids are polymers of N-substituted glycines; they differ from peptides in that the peptoid sidechains are attached to the backbone nitrogen atoms instead of to the backbone α-carbons. Heteropolymeric peptoids have the potential to become a powerful and flexible system with which to create new molecules able to carry out therapeutic, diagnostic or structural functions (1,2). Their synthesis is sequence-specific, efficient, automatable, and inexpensive, owing to the simplicity of the component synthons (1). Moreover, a submonomer synthesis protocol allows the incorporation of a large variety of amines as the N-substituent sidechains (3) . But ultimately, the feasibility of rationally designing functional peptoids will depend on the structural properties of this class of molecules, as well as on our ability to understand and use these properties. Therefore, we are currently investigating the nature, stability, and origins of the folded conformations adopted by medium-length peptoids, as well as the effect of the sidechains on the backbone conformation, using both experimental and computational methods.

Recent work on peptoids containing chiral centers in their sidechains at the Nα position has demonstrated the existence of stable folded conformations in aqueous and organic solvents for chain lengths of 5 to 30 residues (4). The circular dichroic (CD) spectra of molecules with specific sidechain sequences show a double minimum with nadirs around 200 and 220 nm accompanied by a maximum around 190 nm, which is strongly reminiscent of the alpha-helical signal in peptides. Our data strongly argue that this signal arises from a regular ordering of the backbone amide chromophores, and that this ordering does not occur through aggregation (4).

This suggests that a significant fraction of the molecules in solution are monomeric and adopt a regularly repeating conformation of specific handedness. In order to develop a basic understanding of peptoid structure, to assist in the interpretation of our experimental data, and to guide peptoid design efforts, we have modeled this conformation at the atomic level. While there have been previous reports on the modeling of oligopeptoids (1, 5), they have focused on sarcosine (in which the sidechain is simply a methyl group). However, the presence of chiral, di-substituted Nα carbons drastically affects the spectroscopic characteristics of the molecules (4). Thus, a re-examination of their conformational properties is warranted. In this report, we study the available helical geometries of Nα–chiral peptoids (a subset of peptoid secondary structure), and the minimum-energy conformation. The molecule that we have chosen for study, based on its spectroscopic characteristics, is an oligomer of (*S*)-*N*-(1-phenylethyl)glycine (*N*spe). It is shown in Fig. 2.1, along with the atom and dihedral nomenclature used in this report. The CD spectra of *N*spe homopolymers display the maximum and double-minimum just described. Fig. 2.2 shows that this signal is present even when one third of the sidechains are achiral, and that reversing the chirality of the *N*spe sidechains reverses the sign of the ellipticity.

Figure 2.1: (*S*)-*N*-(1-phenylethyl)glycine octamer

The N-terminus is free and the C-terminus is amidated. This is the molecule that was used in all calculations unless stated otherwise. Also shown are the mainchain dihedral angles: ϕ, ψ and ω are defined as for glycine; $\chi 1$ refers to the $C_{(i-1)}$,$N_{(i)}$,N-C$\alpha_{(i)}$,N-C$\beta 2_{(i)}$. Atom names follow the nomenclature devised in (17). Note that since we are dealing here with homopolymers, the subscripted sidechain name has been omitted from the names of sidechain atoms (e.g., $N_{xxx}C\alpha$ is referred to as N-Cα)

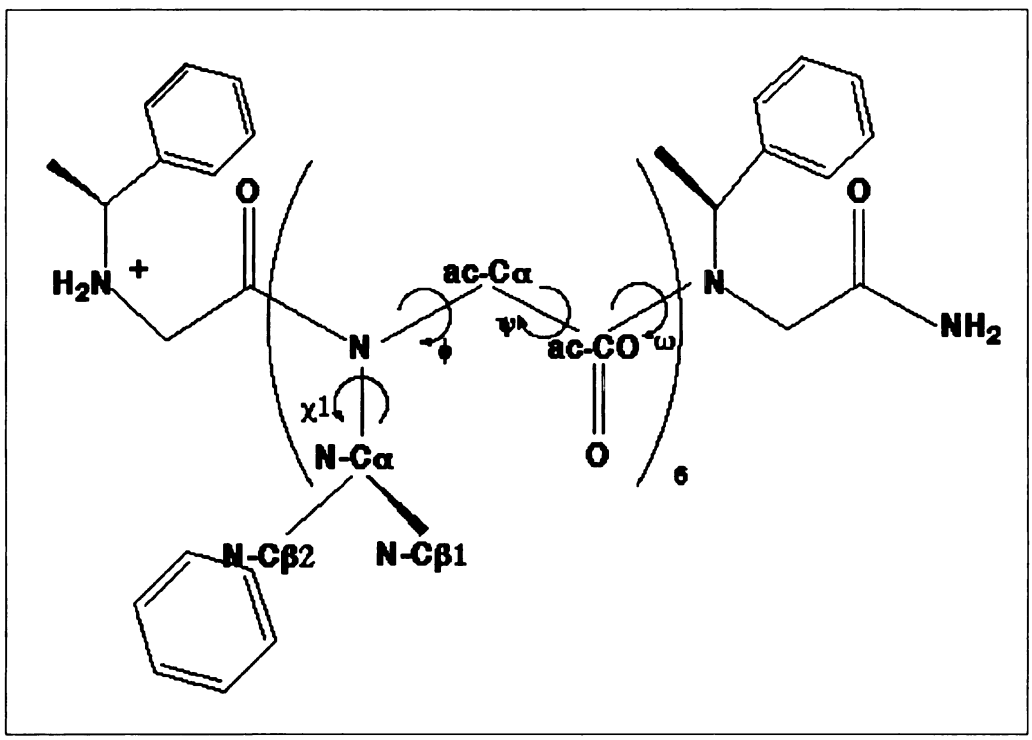
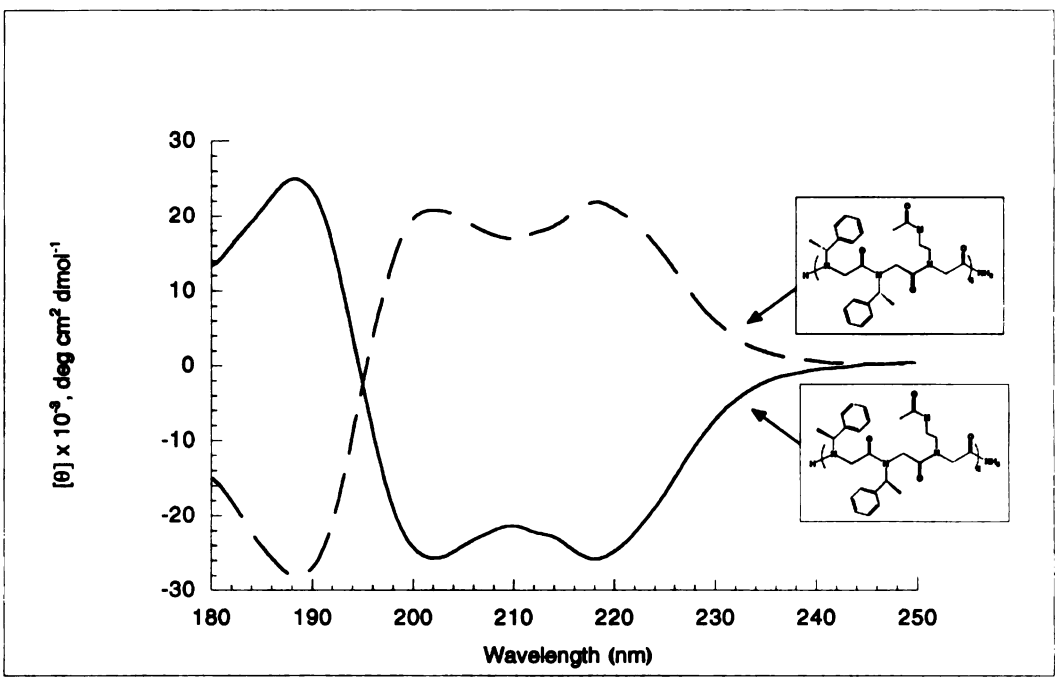


Figure 2.2: CD spectra of peptoids containing *N*-(1-phenylethyl)glycines
This shows the characteristic double peak around 200 and 220 nm as well as the peak of opposite sign at 190 nm, for an oligomer containing two thirds chiral sidechains (Nspe), and one third achiral sidechains (N-(2-acetamidoethyl)glycine). Note how changing the chirality of the *N*spe sidechain inverts the sign of the ellipticity.



Results and Discussion

Conformational Maps

Following the classical approach of Ramachandran et al. for peptides (6), we obtained maps of the allowed regular conformations of unsubstituted (sarcosine) and substituted (*N*spe) octamers, with the amide bond (the ω dihedral angle) adopting either a *cis* or a *trans* conformation (Fig. 2.3). In all cases, bond lengths and angles were kept constant, and conformations were evaluated using the all-atom AMBER molecular mechanics force field (7). We also performed the calculations on an *N*spe dipeptoid while allowing all bond lengths and angles to relax; the maps thus obtained were flatter, but the positions of the minima and their relative depths did not change appreciably (data not shown). We draw three conclusions from this analysis. First, the maps for *N*spe are distinctly asymmetric, unlike those for sarcosine. This asymmetry implies that the chiral centers in the sidechains can give rise to a significant preference of backbone handedness. This is consistent with the CD features of *N*spe-containing peptoids: the double minimum indicates a chiral backbone conformation, whose handedness is determined by the chirality of the sidechain (Fig. 2.2). Second, the maps show that the minima for octa-*N*spe in the *cis* conformation are lower in energy (more favorable) than those in the *trans* conformation, while this preference is reversed for octa-sarcosine. This can be understood from simple steric considerations, by assuming that the bulkiest groups on either side of the C-N amide bond will prefer to be *trans* to each other (Fig. 2.4). Of the two groups bonded to the carbonyl carbon, the methylene group of the mainchain Cα is the bulkier. It will prefer to be *trans* to the bulkier group bonded to the nitrogen. In the case of sarcosine, this will be the mono-substituted mainchain ac-Cα (Fig. 2.4B); but in the case

Figure 2.3: Conformational maps of peptoid octamers

The potential of octa-*N*spe is shown in color, and the potential of octa-sarcosine by black contour lines. "CIS" denotes octamers with ω values around 0 degrees, while "TRANS" denotes octamers with ω values around 180 degrees. The cutoffs are 500 units in all cases. Note that since bond lengths and angles are fixed, the units cannot be interpreted as kcal/mol. A line has been drawn on the *cis* map to separate right- from left-handed helices. Points symmetrical about the center of a map represent enantiomeric backbone structures.

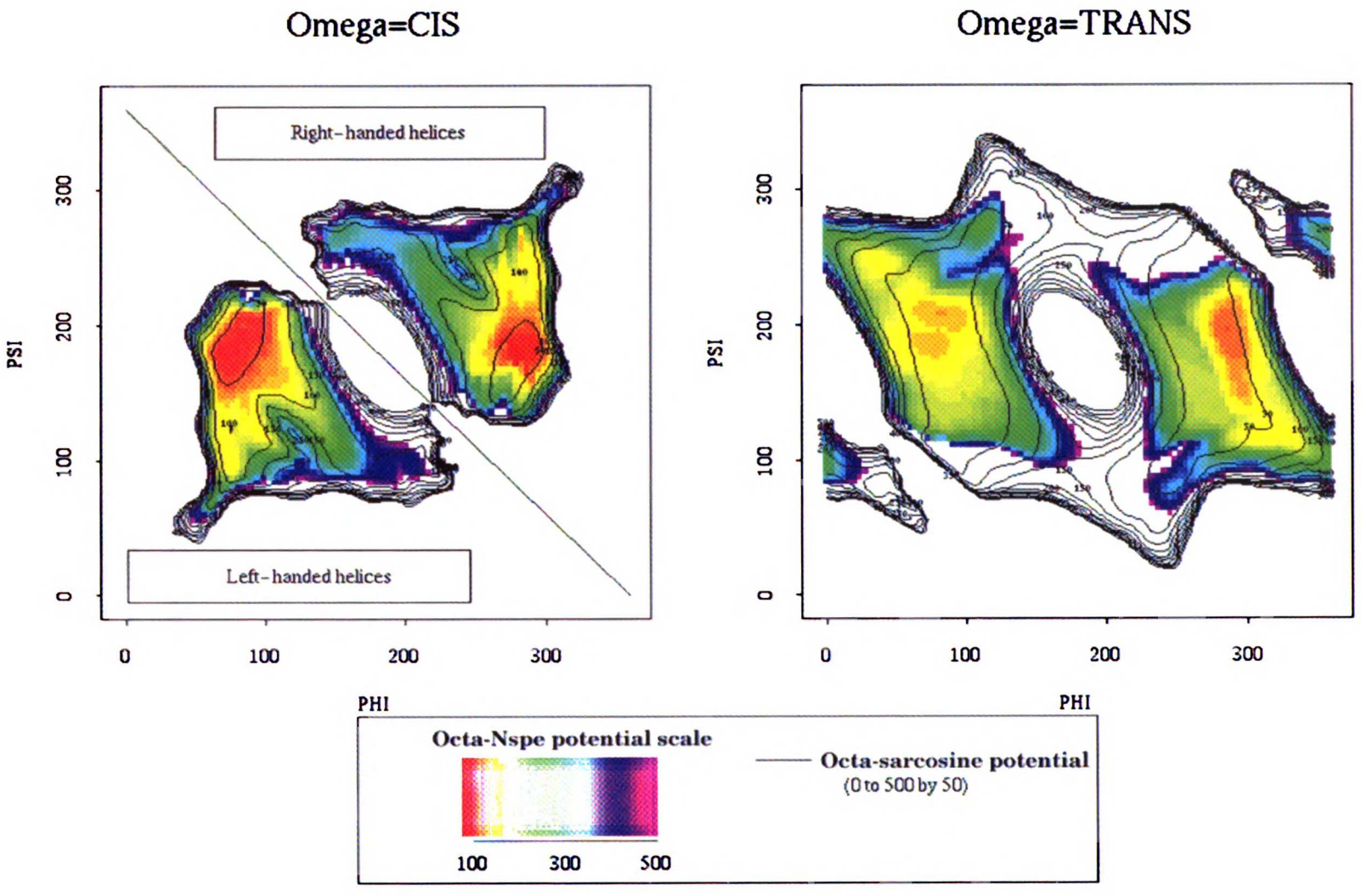
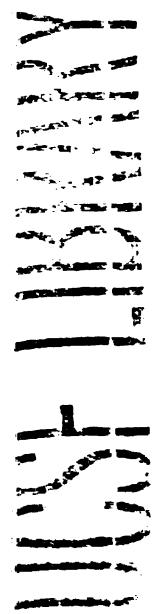


Figure 2.4: Interactions around the amide bond

A, *N*spe , with *cis* (**A**1) or *trans* (**A**2) ω (ω is the amide bond dihedral defined by the green mainchain carbons); **B**, Sarcosine with *cis* (**B**1) or *trans* (**B**2) ω. The bulkiest groups on either side of the amide bond are highlighted in orange (conformations where the two orange groups are *trans* to each other avoid the steric repulsion --double line-- that occurs when they are *cis*).



of *N*spe, it will be the di-substituted sidechain N-Cα (Fig. 2.4A). Third, for each value of ω, there are apparently only two broad regions of backbone conformational space that are easily accessible, and their backbones are mirror images of each other. In those regions, the presence of the sidechain has little impact on the shape of the energy landscape in the *cis* conformation. This holds true for a variety of sidechains (data not shown).

Together, these results argue that the secondary structure of Nα-chiral peptoids may be easier to predict than that of peptides. Indeed, chirality at the Nα atoms should be sufficient to impart a chiral preference to the backbone. In addition, any chiral carbon at the Nα position will be bulkier than the mainchain mono-substituted ac-Cα, leading the amide bond to adopt a *cis* geometry (Fig. 2.4). Taken together, these two restraints could then force backbone dihedrals to lie within a narrow and predictable range (Fig. 2.3): ϕ between -120 and -60 degrees for the (*S*)-*N*-(1-phenylethyl)glycine isomer (or between 60 and 120 degrees for the (R) isomer), and ψ between 150 and 210 degrees. Here we only consider regular backbone conformations, and not turns or other irregular structures.

Minimum energy conformation

To model in greater detail the most energetically favorable conformation of (*N*spe)8, we sampled the space of simple helical conformations (see Materials and Methods), again using the AMBER force field (7) and allowing all bond lengths and angles to relax. After minimization, we retained the four conformations of lowest energy: two enantiomeric *cis* helices ((ω≈0, ϕ≈-75, ψ≈170, right-handed) and (ω≈0, ϕ≈75, ψ≈-170, left-handed)), and two enantiomeric *trans* helices ((ω≈170, ϕ≈-75,ψ≈180, left-handed), and (ω≈185, ϕ≈70, ψ≈170, right-handed)). Of these, the *cis* right-handed helix had lowest energy (Table 2.1A).

Table 2.1: Conformational Comparisons

A, AMBER calculations

Helix Type	Phi*	Psi*	Omega*	AMBER Energy‡
Cis Right-handed	-70	165	*cis*	0.0†
Cis Left-handed	70	185	*cis*	5.6
Trans Right-handed	-75	185	*trans*	34.5
Trans Left-handed	70	180	*trans*	23.8†

*Dihedrals values are average for the helix

† These two values are for a $\chi 1$ of +60°, which for these two structures gives a lower AMBER energy.

‡Values are in kcal/mol, relative to the lowest-energy conformation

B, AMSOL Dipeptoid calculations

Helix Type	Phi	Psi	Omega	$\chi 1$§	AMSOL Energy‡	
					Gas	Solvated
Cis Right-handed	-76.5	-167.2	*cis*	-113.2	0.0	0.0
Cis Left-handed	+72.0	165.6	*cis*	-130.8	1.12	0.58
Cis Left-handed	+85.1	165.1	*cis*	86.2	1.46	2.01
Cis Right-handed	-82.1	-113.9	*cis*	89.3	2.01	2.50

‡Values are in kcal/mol, relative to the lowest-energy conformation

§Chi1 refers to the $C_{(i-1)}$,$N_{(i)}$,N-$C\alpha_{(i)}$,N-$C\beta 2_{(i)}$ dihedral

However, the empirical AMBER force field was not parameterized for peptoids; we therefore used semi-empirical quantum mechanical calculations, performed with the AMSOL force field (8), to compare the *cis* conformations more accurately. We calculated the helical handedness and sidechain rotamer preferences on a *cis* *N*spe dipeptoid with full relaxation (Table 2.1B). The dipeptoid gas-phase calculations revealed a 0.5 kcal/mol per residue preference for the right-handed helix (negative ϕ) when the sidechain adopts the most favorable rotamer (in which the sidechain N-Hα hydrogen of residue (i) points toward the carbonyl oxygen of residue (i-1)). These energetic differences may change in magnitude or direction as the helix grows and residues begin to interact with each other. However, our molecular mechanics results for the octamer agree qualitatively with the dipeptoid calculations in favoring the *cis* right-handed helix backbone for the (*S*) isomer of *N*-(1-phenylethyl)glycine.

These preferences can be rationalized as follows (Fig. 2.5):

a) The preferred rotamer (Fig. 2.5A-1) avoids a weak steric interaction between the two N-Cβ carbons of sidechain (i) and the carbonyl oxygen of (i-1). The other possible rotamer (Fig. 2.5A-2) would have these N-Cβ carbons straddling the oxygen, which would introduce some steric strain.

b) This rotamer prevents ϕ from adopting *trans* values (around 180 degrees), since then the N-Cβ (i) carbons would have an unfavorable steric interaction with the ac-CO(i) carbonyl carbon (Fig. 2.5B-2). Therefore, ϕ is limited to values around +90 or -90 degrees (Fig. 2.5B-1) .

c) There are then two important interactions to consider between the sidechain and the backbone atoms of (i) which determine the sign of ϕ, and thus the handedness of the helix (Fig. 2.5C). First, there is a steric interaction between the carbonyl atoms and either the bulkier ring (if ϕ is positive) or the

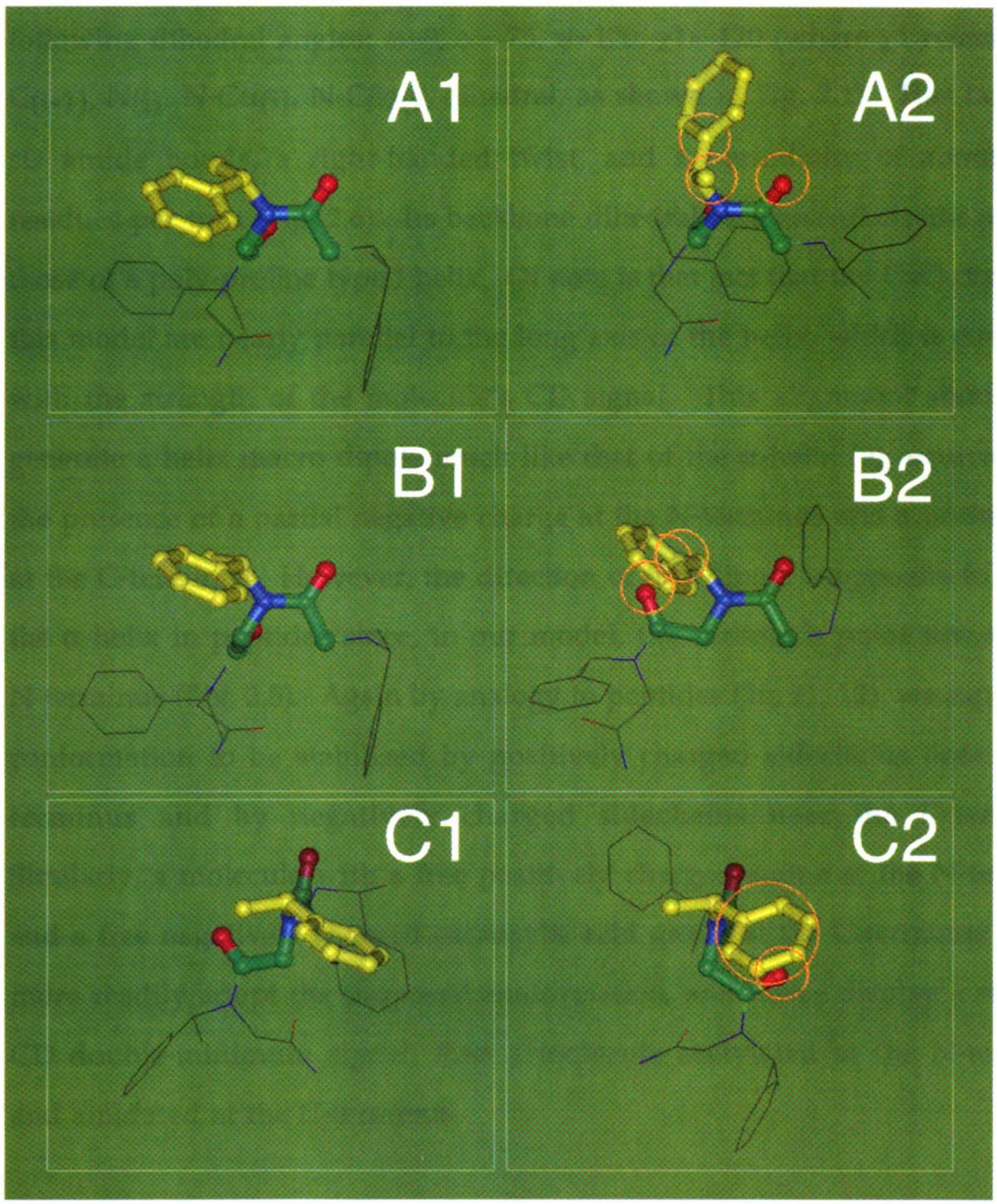
Figure 2.5: Rotamer and handedness preferences of *N*spe

Steric repulsion is denoted by orange circles drawn around the clashing atoms or groups. All relevant atoms are rendered in ball-and-stick model, and colored as described in the legend to Fig. 5.

A, Rotamer preferences. A $\chi 1$ of -120 degrees (**A1**) avoids the steric repulsion between the carbonyl oxygen and the Nβ groups; this repulsion occurs if $\chi 1$ is +60 degrees (**A2**).

B, Exclusion of *trans* ϕ dihedral: a ϕ of ±60 degrees (**B1**) avoids the steric repulsion between the carbonyl oxygen and the Nβ groups; this repulsion occurs if ϕ is 180 degrees (**B2**).

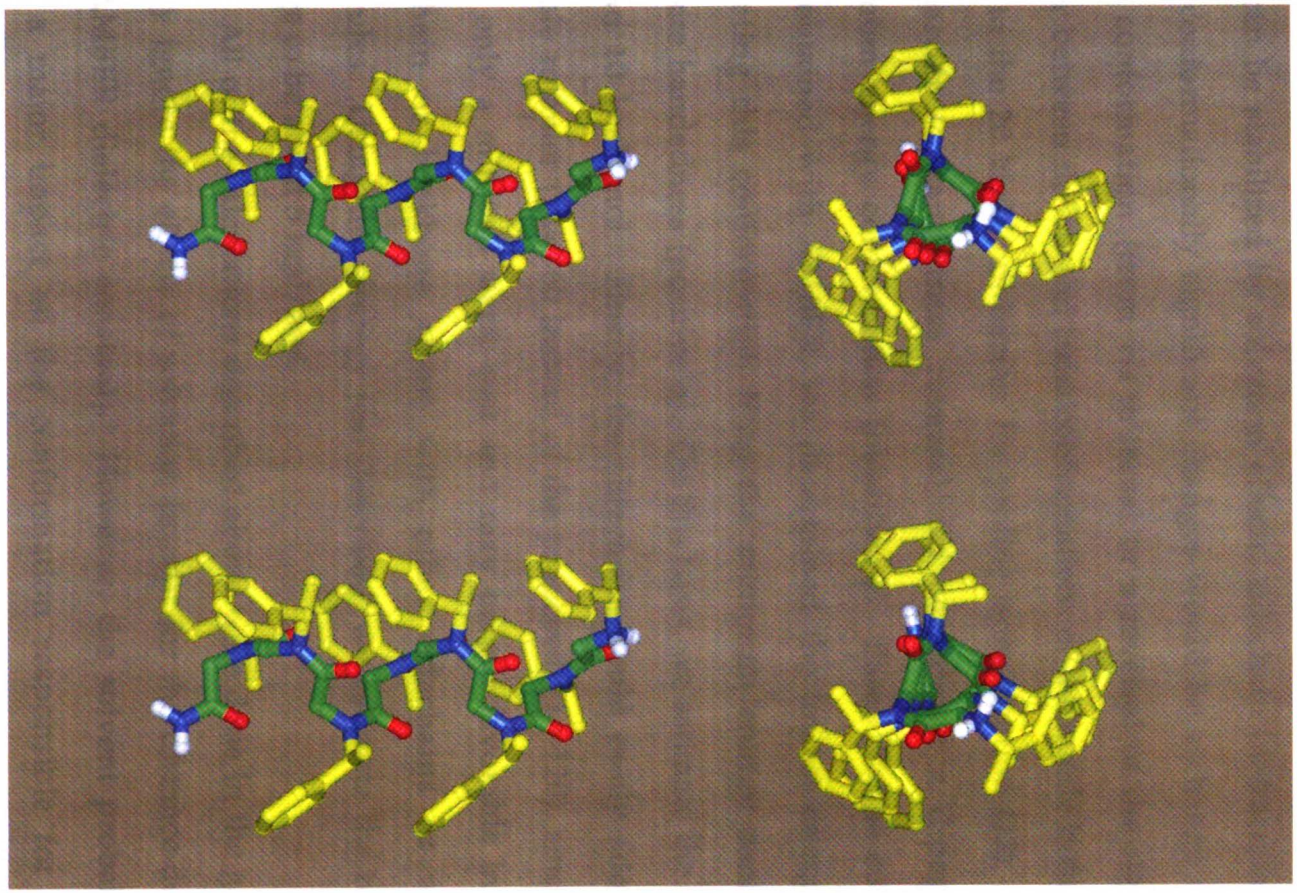
C, Handedness preference. A ϕ of -60 degrees (**C1**) avoids the steric and electrostatic repulsion between the carbonyl oxygen and the ring; this repulsion occurs if ϕ is +60 degrees (**C2**) (this preference would be reversed for the (R) isomer).



smaller Nβ methyl group (if ϕ is negative). This should favor a negative ϕ. Secondly, an electrostatic repulsion between the π orbitals of the aromatic ring and the carbonyl oxygen can reinforce this preference.

We therefore predict that an *N*spe octamer will form a helix with the following dihedral angles: $\omega \approx 0$, $\phi \approx -75$, $\psi \approx 170$, $\chi 1 \approx -120$ (where $\chi 1$ refers to the $C_{(i-1)}$, $N_{(i)}$, N-$C\alpha_{(i)}$, N-$C\beta 2_{(i)}$ dihedral, as shown in Fig. 2.1). This helix has *cis* amide bonds, a right-handed twist, and a periodicity of about three residues per turn (Fig. 2.6). Its backbone dihedrals are essentially the same as those of a poly-proline type I helix. Of note is that fact that the C=O dipoles in this model are nearly parallel to the long axis of the helix, which is consistent with the strength of the molecule's CD signal. This alignment should also generate a helix macro dipole much like that of the α-helix (9), equivalent to the presence of a partial negative charge at the N-terminus and a positive one at the C-terminus. However, the direction of this dipole is opposite to that of the α-helix in peptides since, in our model, the carbonyls point towards the N-terminus (Fig. 2.5). Again by analogy to peptides (10, 11, 12), we expect this conformation to be stabilized by positively charged sidechains near the N-terminus and by negatively charged sidechains near the C-terminus. Similarly, a molecule with a free positively charged amine at the N-terminus and a free negatively charged carboxylic acid group at the C-terminus should more readily adopt the proposed conformation, and hence display a stronger CD double-minimum signal, than a molecule acetylated at the N-terminus and amidated at the C-terminus.

Figure 2.6: Stereo diagrams of the predicted structure of octa-*N*spe
All heavy atoms are shown (mainchain carbons in green, nitrogen atoms in blue, oxygens in red, and sidechain carbons in yellow). The model is viewed parallel (*top*) and perpendicular (*bottom*) to the long axis of the helix.



Although we have focused on the minimum energy conformation, it appears that the energetic differences associated with changes in ω, in handedness, or in $\chi 1$ are small (Table 2.1). In specific situations, therefore, polypeptoid conformations may depart from the proposed helix, if in so doing they can be stabilized by solvation effects, sidechain interactions, etc... The most conformationally significant changes would be a change of the amide bond conformation (from *cis* to *trans*) or a reversal of the sign of ϕ. The barrier between the *cis* and *trans* conformations of the amide bond is probably close to the 20 kcal/mol barrier for proline isomerization (13). To estimate the barrier to handedness reversal, we considered a dipeptoid in the minimum energy conformation (negative ϕ). Based on the conformational map obtained with AMBER, we determined the lowest energy maximum separating this conformation from the minimum energy conformation of opposite handedness (positive ϕ). On this basis, we estimate the barrier to be around 10 kcal/mol. Finally, the sidechains could easily adopt the other possible rotamer ($\chi 1$=+60 instead of the proposed $\chi 1$=-120), since this would incur only a minimal steric cost and may permit favorable interactions of sidechain or backbone atoms with other sidechains or with solvent molecules. Using the strategy just described, we estimate the barrier to this change to be about 5 kcal/mol.

All of the calculations described above were done in the gas phase and largely ignore the effects of solvation (except for the distance-dependence of the AMBER dielectric constant). However, the solvent probably does not have a major impact on the conformation responsible for the double-minimum CD signal; indeed, we have observed this signal in a variety of both aqueous and organic solvents (4). Furthermore, when we repeated the AMSOL dipeptoid calculations taking into account solvation energies, the

conclusions were qualitatively unchanged (Table 2.1B). This argues for the robustness of our calculations, and the primacy of the steric term in determining the conformation of these molecules.

Peptoids whose sidechains lack the Nα chiral center should behave differently. Changing the degree of substitution at the Nα atom would likely affect both the ω and the ϕ preference, as well as the number of allowed rotamers. This should be reflected in a change in the spectroscopic properties of the molecules. The decreased energy of *trans* conformations and the lack of an enantiomeric preference would increase the number of accessible conformations at a given temperature, which would be entropically favorable. Therefore, as the number of achiral sidechains increases in an oligopeptoid, we expect a decrease in the strength of the double minimum CD signal, as more molecules depart from the regular helix conformation. This decrease in the magnitude of the signal may be accompanied by an increase in its heat-stability, reflecting the entropic benefit of the greater number of accessible conformations.

Conclusions

The results presented here lend some insight into the conformational properties of Nα-chiral oligopeptoids, and into the differences between the protein folding and peptoid folding problems. Indeed, the constraints which determine the allowed values of the backbone dihedral angles for peptoids seem subtly different from those that operate in polypeptides. Polypeptide backbones can adopt very distinct regular conformations (most notably α-helix and β-sheet) , but almost always with negative values of ϕ (13). This handedness of the backbone is driven by the chirality of the Cα, which is identical for all 19 naturally occurring amino-acids with sidechains. Backbone conformations in natural proteins derive therefore not from local chiral choices, but from the balance between local conformational propensities of the different amino-acids and nonlocal interactions (14). This renders rational design of proteins very difficult, due to the delicate balance between local and nonlocal interactions. By contrast, peptoid backbones may be forced by the presence of Nα chiral sidechains to adopt conformations close to the helix described above. The handedness of this conformation can be changed by the choice of the sidechains' chirality. Thus, it may ultimately prove easier to design peptoids than proteins, by using Nα-chiral sidechains to force the backbone dihedrals into a predictable region of conformational space, and different chemical groups distal to the Nα atom to direct the combination of the backbone elements and to introduce functionality.

Peptoids present a new and promising structural paradigm, which could translate into new functional possibilities. We are involved at present in determining the solution structure of Nα–chiral oligopeptoids, and in

Materials and Methods

Synthesis and Circular Dichroism

Synthesis and spectroscopy were performed as described in (4). For the CD spectra shown in Fig. 2.2, the molecules were dissolved in 96% H_2O/4% methanol to a concentration of 0.1mM, and the spectra collected at 25C.

Conformational Maps

The values of the ϕ and ψ dihedrals of the octamers were independently changed from 0 to 360 degrees in increments of 4. At each point, the sidechain dihedrals and the ω dihedral were allowed to relax in a force field that consisted of the non-bonded (van der Waals and electrostatics) and torsional components of the all-atom AMBER potential (7), with (1,4) interactions scaled as described in (7). To improve computational efficiency, we kept bond lengths and angles fixed to the monomer equilibrium values. Missing partial charges and torsional parameters were determined simply by similarity to peptides. Because of the lack of accurate parameters for peptoids, we have repeated these calculations without the torsional potentials (data not shown). The maps were essentially unchanged, implying that their salient features are sterically determined. The minimization was performed with a downhill simplex algorithm (15). For the dimers, bond lengths and angles were allowed to relax; ϕ and ω were scanned in increments of 10 degrees, using an internal coordinate conjugate gradient minimizer. All calculations reported here were done on SGI IndigoII workstations, and all software written for this work, unless stated otherwise. Graphical displays were created and printed out using the *Insight® II* 95.0 molecular modeling system (Biosym/MSI, San Diego).

Minimum Energy Conformation

Conformations were generated in two steps: in the first step, octamers were minimized in the force field described above, with all bond lengths and angles fixed, and all dihedrals constrained to lie within 40 degrees of each other (this excluded irregular conformations, but was necessary to keep the problem computationally tractable). The ω torsion was constrained to lie between -15 and 15 degrees (*cis*), or between 165 and 195 degrees (*trans*). The minimization was performed using a simulated annealing protocol (16). Low-energy structures were then further minimized in the full AMBER force field, allowing all dihedrals, bond angles, and bond lengths to relax, using an internal coordinate conjugate gradient minimizer. Again missing parameters were inferred from similarity to peptides.

Semi-empirical quantum mechanical calculations were performed with AMSOL version 4.5 (8), using the AM1-parameter set. The calculations were also done in solvent by using the AM1-SM2.1 parameter set. The dipeptoid was N-acetyl-(*S*)-*N*-(1-phenylethyl)glycine with an N-dimethyl amide cap.

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., Bartlett P. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**:9367

2. Zuckermann R. N., Martin E. J., Spellmeyer D. C., Satuber G. B., Shoemaker K. R., Kerr J. M., Figliozzi G. M., Goff D. A., Siani M. A., Simon R. J., Banville S. C., Brown E. G., Wang L., Richter L., Moos W. H. (1994) *J. Med. Chem.* **37**:2678

3. Zuckermann R. N., Kerr J. M., Kent S. B. H., Moos W. H. (1992). *J. Am. Chem. Soc.* **114**, 10646-10647

4. Kirshenbaum K., Barron A. E., Goldsmith R. A., Armand P., Bradley E. K., Truong K. T. V., Dill K. A., Cohen F. E., Zuckermann R. N. (1998) *Proc. Natl. Acad. Sci. USA* **95**:4303

5. Moehle K., Hofmann H.-J. (1996).*Biopolymers* **38**, 781-790

6. Ramachandran G. N., Ramakrishnan C., Sasisekharan V. (1963). *J. Mol. Biol.* **7**, 95-99

7. Cornell W. D., Ciepak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179-5197

8. Cramer C. J., Hawkins G. D., Lynch G. C., Giesen D. J., Truhlar D. G., Liotard D. A. (1994). *QCPE* **14**, 55-57

9. Wada A. (1976). *Adv. Biophys.* **9**, 1-63

10. Nicholson H., Anderson D. E., Dao-pin S., Matthews B. W. (1991) *Biochemistry* **30**, 9816-9828

11. Sancho J., Serrano L., Fersht A. R. (1992). *Biochemistry* **31**, 2253-2258

12. Armstrong K. M., Baldwin R. L. (1993). *Proc. Natl. Acad. Sci. USA* **90**, 11337-11340

13. Creighton T. E. (1993). *Proteins: Structures and Molecular Properties*. (2nd edn), W. H. Freeman, New York

14. Dill K. A. (1990). *Biochemistry* **29**, 7133-7155

15. Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. (1992). *Numerical Recipes in C.* , Cambridge University Press, Cambridge

16. Kirkpatrick S., Gelatt C. D. Jr, Vecchi M. (1983). *Science* **220**, 671-680

17. Bradley E. K. (1996). *J. Mag. Res.* **110**, 195-197.

Chapter 2:

Stable Secondary Structure in Nα-chiral polypeptoids

This chapter was published as:

Kirshenbaum K., Barron A. E., Goldsmith R. A., Armand P., Bradley E. K., Truong K. T. V., Dill K. A., Cohen F. E., Zuckermann R. N. (1998) "Sequence-specific polypeptoids: A diverse family of heteropolymers with stable secondary structure" *Proc. Natl. Acad. Sci. USA* **95**:4303

Abstract

We have synthesized and characterized a family of structured oligo-*N*-substituted-glycines (*peptoids*) up to 36 residues in length, using an efficient solid-phase protocol to incorporate chemically diverse sidechains in a sequence-specific fashion. We have investigated polypeptoids containing sidechains with a chiral center adjacent to the mainchain nitrogen. Some of these sequences have stable secondary structure, despite the achirality of the polymer backbone and its lack of hydrogen bond donors. In both aqueous and organic solvents, peptoid oligomers as short as five residues give rise to circular dichroism spectra that strongly resemble those of peptide α-helices. Differential scanning calorimetry and circular dichroism measurements show that polypeptoid secondary structure is highly stable and that unfolding is reversible and cooperative. Thermodynamic parameters obtained for unfolding are similar to those obtained for the α-helix to coil transitions of peptides. This new class of biomimetic polymers may enable the design of self-assembling macromolecules with novel structures and functions.

Introduction

Advances in bio-organic chemistry now offer accessibility to valuable new materials which integrate the structural and functional characteristics of biopolymers with the stability and diversity of synthetic polymers. Biopolymers are unique in that they are composed of specific monomer sequences and are capable of self-assembly into stable native structures, permitting a remarkable variety of functions. In contrast, most synthetic polymers do not have a specific monomer sequence, but can be cheaply produced from a wide array of monomers and linking chemistries to generate many useful materials. It has been suggested that the sequence-specific polymerization of non-biological monomers could allow the creation of synthetic polymers with stable folded structures (1, 2).

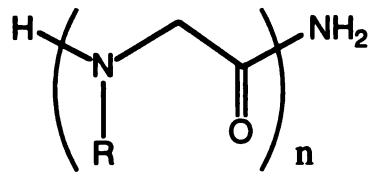
The design of novel polymers with sophisticated properties such as molecular recognition, energy transduction, and catalysis will require the precise spatial positioning of multiple chemical functionalities. The synthesis of such a polymer necessitates: control of monomer sequence, control of polymer length, high coupling efficiencies, the ability to incorporate a variety of sidechain types, and self-assembly into a unique conformation. Additional desired characteristics include: relatively low cost, a flexible polymer backbone allowing a variety of structural motifs, cooperative conformational transitions allowing effective regulation and signaling, thermostability, structural stability in both aqueous and organic solvents, and compatibility with biological systems. In this paper we present a study of peptoids, a family of chemically diverse sequence-specific heteropolymers that may ultimately satisfy these goals.

Several methods have been developed for the synthesis of biomimetic oligomers with unnatural backbones. Many of these are structural variants of polypeptides, such as polycarbamates (3), peptide nucleic acids (4), and vinylogous polypeptides (5). Recent efforts have yielded oligomers with defined folding propensities such as oligoanthranilamides (6), vinylogous sulfonamidopeptides (7) and "aedemers" (8). Helical secondary structures have been demonstrated in oligomers of β-amino acids (9, 10). These systems have been limited to short-chain molecules lacking chemical diversity. Typically, solution structures have been detected only in non-aqueous solvents and are stabilized by hydrogen-bonding.

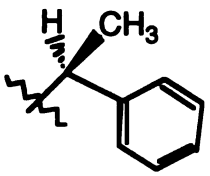
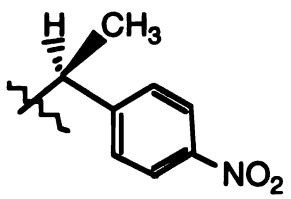
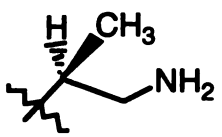
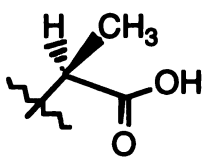
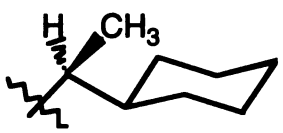
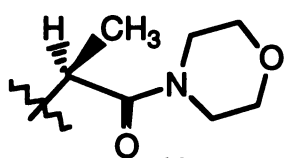
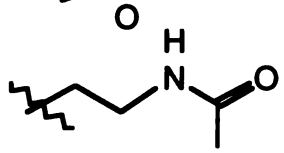
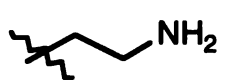
In contrast, polymer chemists have historically obtained helical homopolymers by incorporation of bulky sidechains whose steric repulsion directs the conformation of the backbone (11). The use of chiral monomers or initiators can afford helices with a preferred screw sense. Examples of these asymmetric polymers include bulky methacrylates, isocyanides, trichloroacetaldehydes, and isocyanates (11, and refs. therein).

Peptoids are a family of oligomers that were developed for use in combinatorial drug discovery programs (12, 13). These non-natural oligomers are based on a polyglycine backbone, on which the sidechains are appended to the amide nitrogen, rather than the α carbon (Table 3.1). Since the sidechain moiety is introduced by the reaction of primary amines, this method can yield an extremely diverse series of functionalized oligomers from readily obtainable starting materials (14, 15). Furthermore, coupling efficiencies of >98% now enable the automated synthesis of long-chain (>25 monomers) sequence-specific polymers in good yield.

Table 3.1. *N*-Substituted Glycine Sidechains.



N-substituted glycine oligomer, or *polypeptoid*

R = Sidechain*	Designator
	***N*spe** = (*S*)-*N*-(1-phenylethyl)glycine ***N*rpe** = enantiomer of structure shown, (*R*)-*N*-(1-phenylethyl)glycine
	***N*snp** = (*S*)-*N*-(1-(*p*-nitrophenyl)ethyl)glycine ***N*rnp** = enantiomer of structure shown, (*R*)-*N*-(1-(*p*-nitrophenyl)ethyl)glycine
	***N*sam** = (*S*)-*N*-(2-amino-1-methylethyl)glycine
	***N*sce** = (*S*)-*N*-(1-carboxyethyl)glycine
	***N*sch** = (*S*)-*N*-(1-cyclohexylethyl)glycine
	***N*smc** = (*S*)-*N*-(1-((morpholino)carbonyl)-ethyl)glycine
	***N*am** = *N*-(2-acetamidoethyl)glycine
	***N*ae** = *N*-(2-aminoethyl)glycine

*These structures refer to the N-substituent of a glycine monomer.

Many of the mechanisms that direct the self-assembly of biopolymer structures are lacking in peptoids. Specifically, peptoids lack amide protons, and thus no hydrogen bond network along the polymer backbone is possible. Likewise, since the mainchain contains no chiral centers, peptoids have no intrinsic handedness. Previous studies of short peptoid oligomers by modeling (12, 16) and by NMR (17) have demonstrated that the tertiary amides in the peptoid backbone can result in the population of both *cis* and *trans* amide bond conformers. These findings were also observed for collagen analogs containing peptoid residues (18).

Our aim is to create biomimetic stereospecific polymers, combining strategies previously applied in the biopolymer and asymmetric polymer fields. We have investigated a family of sequence-specific peptoid polymers incorporating bulky chiral sidechains. Recent modeling studies indicated that these peptoids could form a helical secondary structure (19) (Fig. 2.6). We report here the synthesis and characterization of this new class of peptoid polymers.

Materials and Methods

Peptoid Synthesis and Purification. Solvents and reagents were purchased from commercial sources and used without further purification. Peptoid oligomers were synthesized on 50 μmol of a Rink amide resin (NovaBiochem, San Diego, CA) at a substitution level of 0.47 mmol/g. The peptoid oligomers were synthesized by an improvement of previous methods (14, 15), which included reduced cycle times and decreased amine concentrations. Briefly, after removal of the first Fmoc group the following 90 min monomer addition cycle was performed by a robotic synthesizer and repeated until the desired length was obtained. The amino-resin was bromoacetylated by adding 830 μl of 1.2 M bromoacetic acid in DMF and 200 μl of *N,N'*-diisopropylcarbodiimide (DIC). The mixture was agitated for 40 min at 35°C, drained and washed with DMF (3 x 2 ml). Next, 0.85 ml of a 1 M solution of a primary amine in DMSO was added to introduce the sidechain. The mixture was agitated for 40 min at 35°C, drained, and washed with DMF (4 x 2 ml). After the last coupling the peptoid-resin was cleaved and lyophilized as previously described (15).

Individual peptoid oligomers were analyzed by reversed-phase HPLC on C4 columns (Vydac, 5μm, 300 Å, 1 x 150 mm) on a Magic 2002 system (Michrom Bioresources, Auburn, CA). A linear gradient of 5-95% B in A over 40 min was used at a flow rate of 100 μl/min [solvent A= 0.1% trifluoroacetic acid (TFA) in water, solvent B = 0.1% TFA in acetonitrile] at a column temperature of 60°C. Preparative HPLC was performed on a Delta-Pak™ C4 (15 μm, 300Å) on a Waters Prep LC3000 system (Millipore, Milford, MA) using the same solvent system. Peaks were eluted with a linear gradient of

20%-70% B in A over 40 min at a flow rate of 50 ml/min. Electrospray mass spectrometry was performed on a Platform II (Micromass, Beverly, MA).

Amine Submonomers Used in Peptoid Synthesis. *N*-(2-acetamidoethyl)glycine (***N*am**), (*S*)-*N*-(1-cyclohexylethyl)glycine (***N*sch**), and (*R* or *S*)-*N*-(1-phenylethyl)glycine (***N*rpe** or ***N*spe**) were made from the amines aminoethylacetamide, (*S*)-1-cyclohexylethylamine and (*R* or *S*)-1-phenylethylamine (>99% e.e.), respectively, which were obtained from Fluka (Ronkonkoma, NY). (*R* or *S*)-*N*-(1-(*p*-nitrophenyl)ethyl)glycine (***N*rnp, *N*snp**) were made from (*R* or *S*)-1-(*p*-nitrophenyl)ethylamine (>99% e.e.) which was obtained from Celgene (Warren, NJ). (*S*)-*N*-(1-carboxyethyl)glycine (***N*sce**) was made from (L)-*O*-*t*-butyl-alanine (Bachem, Philadelphia, PA). *N*-(2-aminoethyl)glycine (***N*ae**) was made from mono-Boc-ethylenediamine which was made by the method of Krapcho and Kuell (20). (*S*)-*N*-(2-amino-1-methylethyl)glycine (***N*sam**) was made from (*S*)-1-(Boc-amino)-2-aminopropane which was made via three steps from (*S*)-Z-alaninamide [briefly, the amide of (*S*)-Z-alaninamide was reduced to the amine with BH_3, followed by protection of this amino group with di-*t*-butyldicarbonate and removal of the Z group by catalytic hydrogenation]. (*S*)-*N*-(1-((morpholino)carbonyl)ethylglycine (***N*smc**) was made from (S)-alanylmorpholide which was made from (L)-Z-alanine and morpholine.

Circular Dichroism. CD measurements were performed with a JASCO 720 spectropolarimeter equipped with a Peltier temperature control unit. Spectra were obtained in fused quartz cells from 0.1 to 10 mm path length. Data are

expressed in terms of mean residue ellipticity, [θ] (deg·cm^2/dmol), calculated per mole of amide groups present.

Differential Scanning Calorimetry. An MC-2 differential scanning microcalorimeter (MicroCal, Northampton, MA) was used at a scanning rate of 1°C/min under 20 psi N_2. DSC was conducted in degassed 5 mM sodium citrate, 5 mM Na_2HPO_4, 0.1 M NaCl, pH 4.1 at a peptoid concentration of 2.0 or 0.5 mg/ml. A minimum of six scans were conducted to ensure reproducibility. A thermal scan of buffer vs. buffer was subtracted from a scan of sample vs. buffer. Data analysis was conducted using the Origins software supplied by MicroCal.

Nuclear Magnetic Resonance Spectroscopy. Samples for NMR experiments were prepared as approximately 5 mM solutions in CD_3OD. Spectra were acquired at 10°C on a Varian Unity 300 instrument (Palo Alto, CA). Standard 1D proton spectra were collected as 8K data points and zero-filled to 16K).

Results

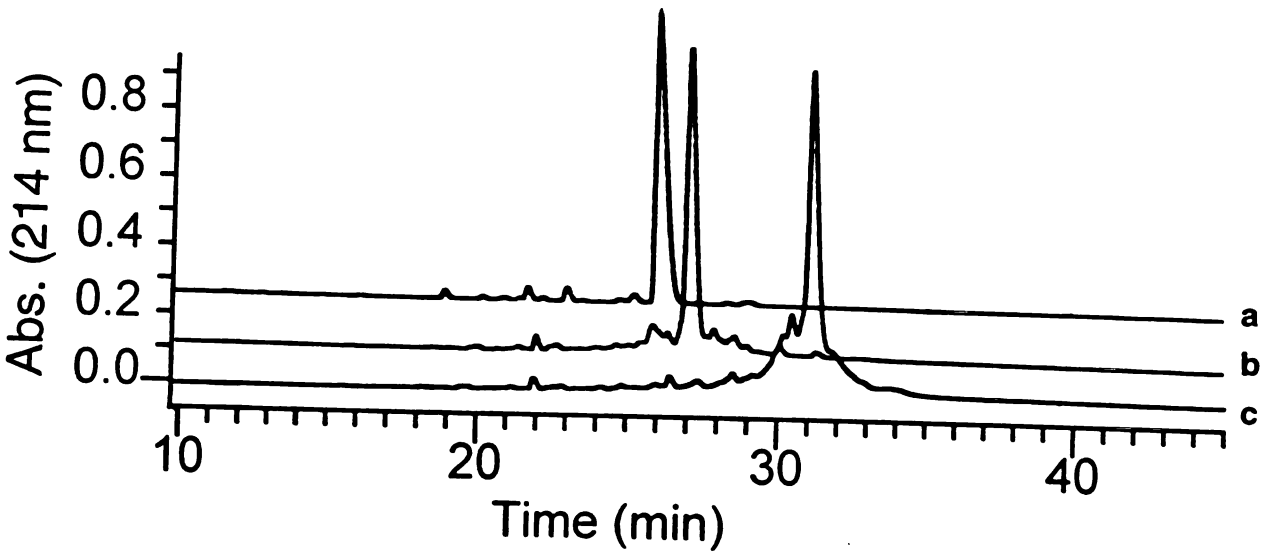
Synthesis of Sequence-Specific Peptoid Oligomers. *N*-substituted glycine oligomers of 3 to 36 residues were synthesized in good yield and high purity using bromoacetic acid and a variety of primary amines as submonomers (14, 15). We created a broad family of oligomeric peptoid products containing a specific sequence of diverse sidechain moieties. These included aliphatic, aromatic, heterocyclic, cationic, and anionic groups (Table 3.1). Reactive sidechain functionalities were protected by TFA-labile groups (Boc for amines and *t*-butyl esters for carboxylic acids) and removed during cleavage of the peptoid from the resin. Crude oligomer purities ranging from 50-90% were determined by HPLC and largely depended on the oligomer length (Fig. 3.1, Table 3.2). Compounds were purified to >95% homogeneity by preparative reversed-phase HPLC prior to characterization. Molecular weights were confirmed by electrospray mass spectrometry and were uniformly in agreement with expected values (Table 3.1).

Table 3.2: Peptoid oligomer structures

Peptoid	Monomer Sequence	Length	MW calc	MW obs	Purity[1] (%)
1	(*N*spe)$_5$	5	822.6	822.5	95
2	(*N*ae*N*spe*N*spe)$_6$	18	2550.8	2550.6	80
3	(*N*ae*N*spe*N*spe)$_{12}$	36	5084.6	5084.2	50
4	(*N*spe*N*spe*N*sce)$_3$	9	1370.8	1370.4	60
5	(*N*rpe*N*rpe*N*am)$_4$	12	1874.3	1873.9	70
6	(*N*sce*N*sce*N*spe)$_{10}$	30	4209.4	4209.9	50
7	Ac-*N*smp	1	265.2	264.9	ND
8	(*N*ssnp)$_3$	3	635.4	635.1	ND
9	(*N*snp)$_5$	5	1047.6	1047.1	99
10	(*N*rnp)$_8$	8	1666.0	1665.8	80
11	(*N*spe*N*sce)$_4$*N*spe	9	1338.8	1338.4	90
12	(*N*sam*N*spe*N*spe)$_6$	18	2634.9	2634.2	60
13	(*N*sme)$_{12}$	12	2394.4	2394.5	95
14	(*N*sch)$_5$	5	852.6	852.4	95

[1]As determined by analytical HPLC of crude product. All compounds were purified to >95% homogeneity. ND, not determined

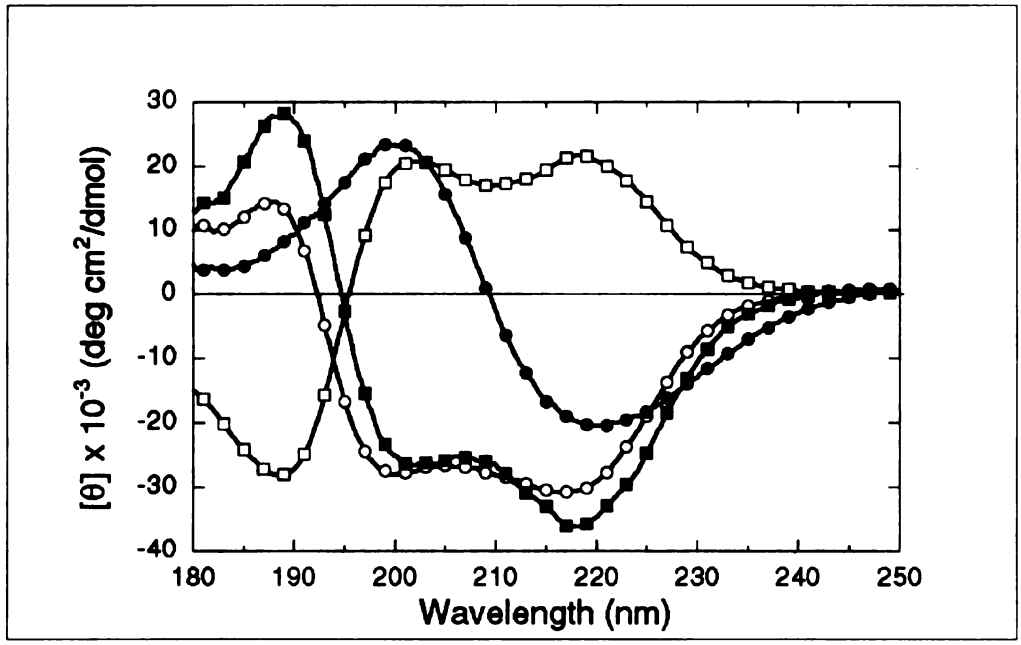
Figure 3.1. Analytical HPLC traces of crude peptoid synthesis products
(a), pentamer **1**$(Nspe)_5$; (b), 18mer **2** $(NaeNspeNspe)_6$; (c), 36mer **3** $(NaeNspeNspe)_{12}$. Traces (a) and (b) are vertically offset for clarity. All products were purified by preparative HPLC prior to further study.


Abs. (214 nm)
0.8
0.6
0.4
0.2
0.0
10
20
30
40
Time (min)
a
b
c

Characterization of Peptoid Oligomers by CD. The CD spectra of peptoid oligomers of varying length, monomer composition and sequence were obtained in aqueous solution over the 250-180 nm region (Fig. 3.2). The spectra for polypeptoids **4, 6** which include (*S*)-*N*-(1-phenylethyl)glycine monomers show two minima of negative ellipticity *ca.* 218 and 202 nm, and a band of positive ellipticity *ca.* 190 nm. The band at 218 nm corresponds to the $n\pi^*$ transition of the amide chromophore (observed between 230 to 210 nm in peptides), while the remaining bands correspond to the high and low wavelength components of the exciton split $\pi\pi^*$ transition (observed around 200 nm in tertiary amides) (21). Additional contributions may derive from coupled interactions between the peptide bands and the L_a transition of the aromatic sidechains (*ca.* 210 nm for phenylalanine) (21). However, variation of aromatic content from 33% to 66% of the sidechains does not substantially alter the magnitude of the ellipticity in this region (Fig. 3.2). Compound **5**, which contains sidechains of opposite chirality to **4** and **6**, shows a mirror-image α-helix type CD spectrum. Peptoid homopolymers **13, 14** comprised of chiral sidechains containing cyclohexyl and morpholidyl groups yielded CD spectra lacking the double minimum profile.

Figure 3.2. CD spectra of polypeptoids with varying monomer composition and length

■, compound **4** $(NspeNspeNsce)_3$; □, **5** $(NrpeNrpeNam)_4$; ○, **6** $(NsceNsceNspe)_{10}$; ●, **13** $(Nsmc)_{12}$. Peptoid concentration was 0.2 mg/ml (*ca.* 1 mM residue molarity for all species) in 10 mM sodium phosphate (pH 7.0), 25 °C.

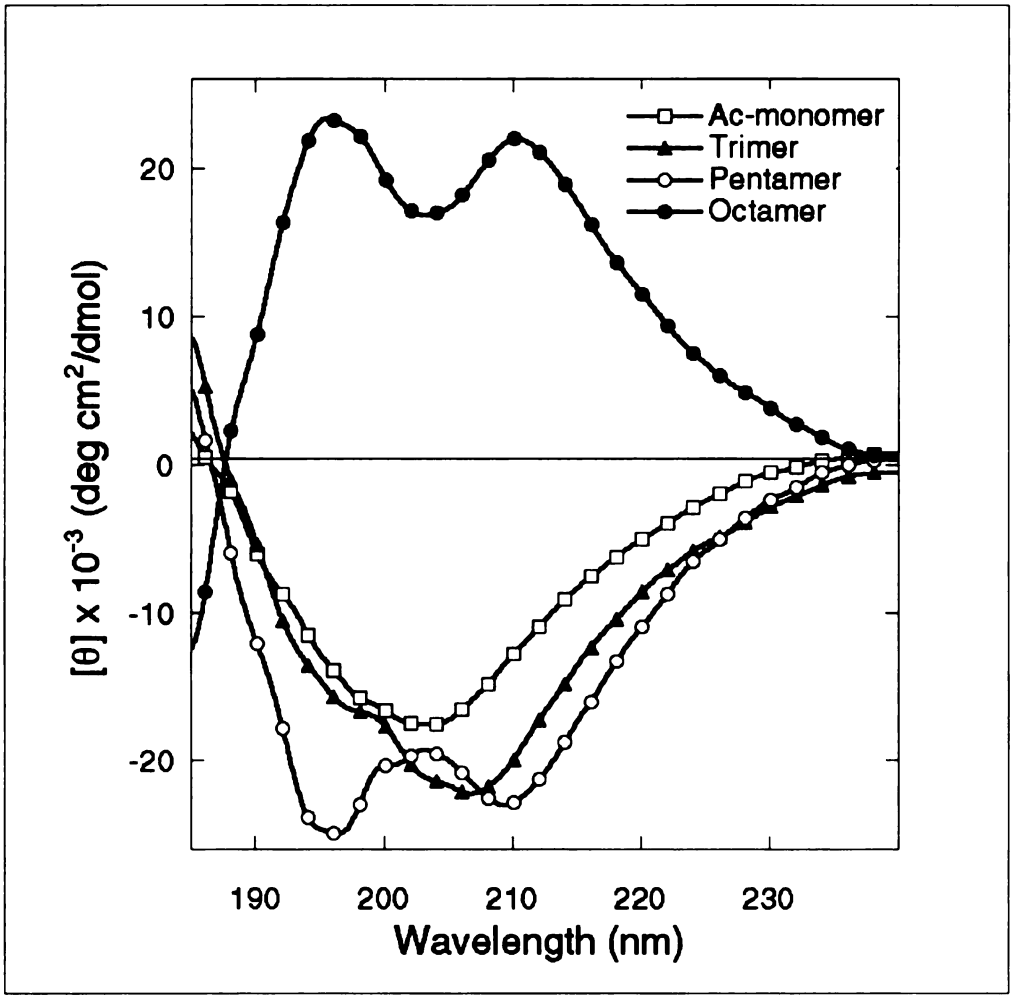


To determine the minimum chain length required to establish the characteristic helix-like CD signal, the CD spectra of compounds **7-10**, corresponding to $(N\text{snp})_n$ (n= 1, 3, 5) and $(N\text{rnp})_8$ oligopeptoids, were obtained in acetonitrile in the region 250 to 185 nm (Fig. 3.3). Compounds **7-10** all have a CD band at 204 to 210 nm. The pentamer and octamer display an additional band at around 195 nm, giving rise to a helix-like CD signal. Similar spectra were obtained for these compounds in methanol (22). The spectra of $(N\text{rnp})_8$ (**10**), comprised of sidechains of the opposite chirality, is the mirror image of the ellipticity of $(N\text{snp})_5$ (**9**), consistent with the data presented in Fig. 3.2. Comparison of spectra for **9** and **10** shows that the intensity of the signal on a per-residue basis does not increase for oligomers longer than five residues. These helix-like spectra are similar to those of peptoids **4-6** in aqueous solution, with a blue shift of the CD bands.

The CD spectra of **9** were independent of concentration in the range 50 μM to 1 mM in acetonitrile, indicating the absence of intermolecular association . Consistent with this finding, the ^{1}H 1D NMR spectra of this pentamer reveal sharp line widths (Fig. 3.4).

Figure 3.3. Circular dichroism spectra in acetonitrile

Ellipticities for the *N*-(*p*-nitrophenylethyl)glycine trimer **8** $(Nsnp)_3$ (▲), pentamer **9** $(Nsnp)_5$ (O), octamer **10** $(Nrnp)_8$ (●), and the acetylated monomer **7** Ac-Nrnp (□) were calculated per mole of amide groups present. Helix-like CD signals are evident for the pentamer and octamer. The octamer was prepared from monomers of the opposite chirality (R vs. S). Peptoid concentration was 2 mg/ml, 25°C.



Melting Studies and Determination of Thermodynamic Parameters. We used CD to determine whether the helical structure can be deformed as a function of pH and temperature. Some polypeptoids undergo a sharp structural transition as a function of pH. Decreasing the pH from 7 to 3 leads to only small changes in the CD spectra of the 30mer **6** (Fig. 3.5). Upon further titration to pH 2, the signal around 200 nm is almost entirely abolished, and the long wavelength band is red-shifted and diminished. The spectrum at pH 2 resembles that of the *N*spe monomer, indicating the absence of repeating secondary structure.

CD spectra were obtained as a function of temperature for peptoid oligomers in varying solution conditions. The 9mer **11** and 30mer **6** show little change in ellipticity as temperature was increased from 5 to 65 °C at pH 7 (Fig. 3.6). We utilized the pH dependent conformational transitions described above to de-stabilize the secondary structure sufficiently to detect melting in an experimentally accessible regime. At pH 4.1, **6** undergoes a sigmoidal thermal transition, which is largely complete within a 40° temperature range. The original CD spectrum is regained upon return to 5° C conditions (data not shown).

Figure 3.5. The pH dependent conformational transition of 30mer 6 as demonstrated by CD

Data were obtained at a concentration of 23 μM in 5 mM sodium citrate, 5 mM sodium phosphate buffer containing 0.1 M NaCl, 25°C. The pH value is shown near each spectrum.

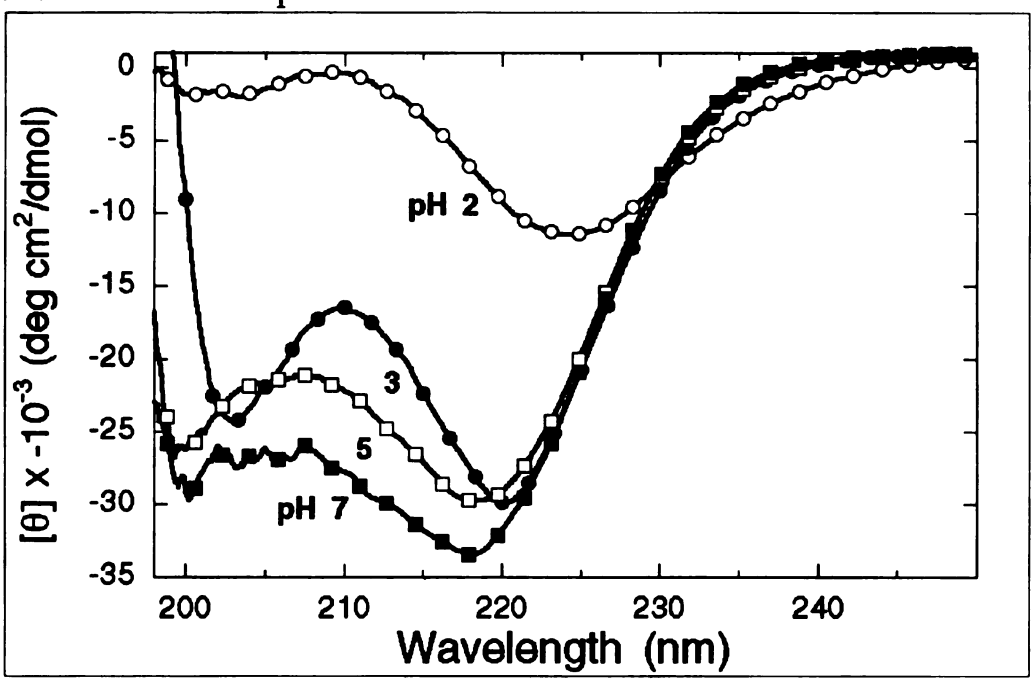
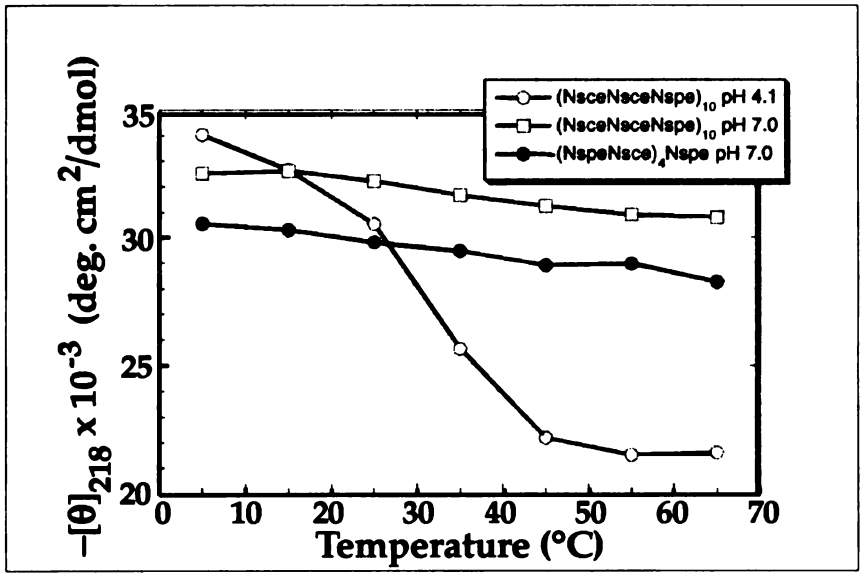


Figure 3.6. Temperature dependence of $[\theta]_{218}$ between 5 and 65°C

Data were obtained at a concentration of 0.2 mg/ml. Peptoid 9mer (*N*spe*N*sce)$_4$*N*spe **11** (●) at pH 7; 30mer (*N*sce*N*sce*N*spe)$_{10}$ **6** at pH 7.0 (□); **6** at pH 4.1 (○).



To more fully investigate the nature of the melting, complete CD spectra were obtained of **6** at pH 4.1 at temperatures from 5 to 65°C (Fig. 3.7). The loss of signal intensity at 220 nm, and the disappearance of the 203 nm minimum over a narrow temperature range indicate the loss of secondary structure content in a cooperative fashion. The spectra show isodichroic points at 230 and 200 nm, suggesting a two-state transition. We calculated the van't Hoff enthalpy for the structural transition by a method similar to that described for study of a 50 residue α-helical peptide (23). The analysis assumes the peptoid is fully helical at 5°C and fully coiled at 75°C, and that ellipticity is a linear function of the amount of helical structure present. The ellipticity values at 203 and 220 nm were separately used to evaluate the fractional helical content at each temperature [fH(T)] and the van't Hoff enthalpy of the transition, ΔH_{vH}. The enthalpies determined by fitting to the van't Hoff equation at each wavelength were in excellent agreement, and both equaled 0.22 kcal/mol-residue.

The thermal transition of **6** is sharp, according to DSC conducted at 460 μM and pH 4.1 (Fig. 3.8). The excess heat capacity has a peak that is complete within a 40°C range. The peptoid transition is fully reversible and reproducible over several scans between 15 and 90°C. After subtraction of the buffer reference trace, the data were fit to a theoretical curve and the area integrated to obtain ΔH_{cal}, 1.1 kcal/mol-residue. CD and DSC studies of **6** were both repeated at 120 μM with no variation of molar signal intensity (data not shown). Cooperative thermal unfolding is also observed in the polycationic 18mer **12** with limited possibility of forming intramolecular hydrogen bonds through its sidechains. At a concentration of 36 μM in 10 mM sodium phosphate, 10 mM sodium borate buffer, pH 8.3, this molecule

Figure 3.7. Thermal unfolding of pH-destabilized 30mer 6 in 5mM sodium citrate, 5 mM sodium phosphate buffer pH 4.1 containing 0.1M NaCl
CD spectra were taken at a concentration of 23 μM. Temperature (°C) is indicated near each spectrum.

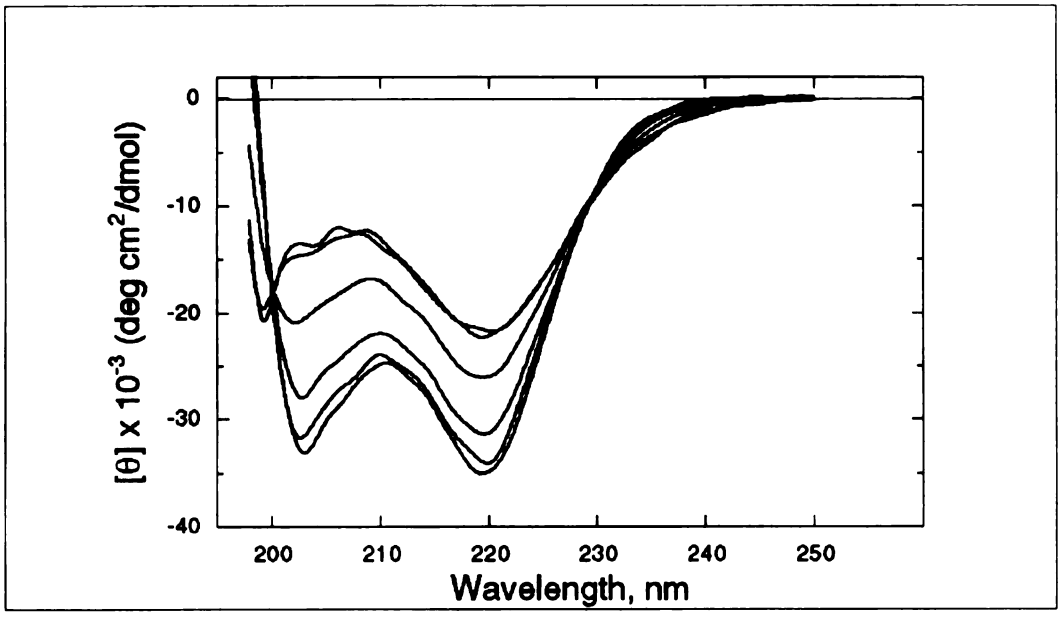
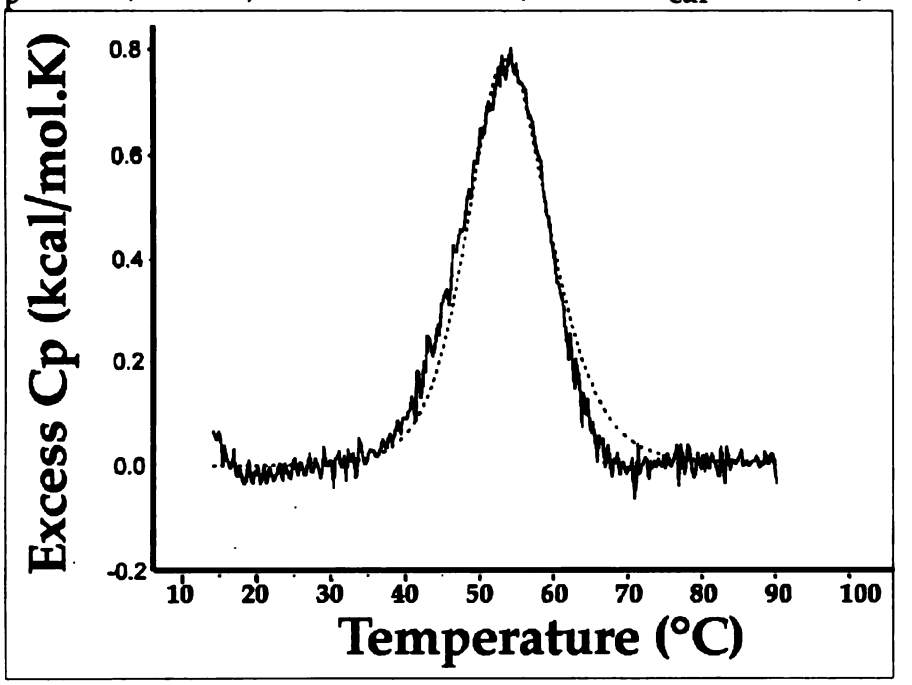


Figure 3.8. DSC scan at a peptoid concentration of 460 μM

$_p$ in kcal/mol·°K, scan rate was 1°C/min. ΔH_{cal} = 1.1 kcal/mol per residue.



Discussion

Although 25mers have been previously synthesized in good yields in our laboratory (14), the study of longer structured molecules presented some synthetic challenges. Since we aimed to induce structure by the incorporation of hindered residues, it was necessary to strike a balance between high coupling efficiencies and strong steric influence of the sidechain over the mainchain. Initial efforts revealed that yields were compromised when α-branched primary alkylamines having two α substituents larger than methyl groups were incorporated into 25mers (data not shown). However, α-methyl primary amines were efficiently incorporated (Tables 3.1,3.2). Importantly, this family of amines are chiral and many can be obtained in high enantiopurity from commercial sources.

The presence of these α-chiral sidechains is predicted to dramatically reduce the conformational space energetically accessible to the polymer backbone. Calculations of α-chiral peptoid helices indicate the preference for a right-handed helix pitch with repeating *cis*-amide bonds and a limited range of ϕ, ψ dihedral angles (19) (Fig. 2.6). Green *et al.* (24) have investigated the effects of chiral *N*-pendant groups on random isocyanate polymers, which likewise have an achiral backbone. These studies demonstrated induction of helical secondary structure of specific handedness, as evidenced by characteristic circular dichroism spectra.

The present study demonstrates that peptoid oligomers containing α-chiral sidechains "fold" to form stable secondary structure. The intensity of the CD signals in the UV spectral regions dominated by amide chromophore transitions (230 to 185 nm) strongly suggests a well-defined repeating mainchain secondary structure (Fig. 3.2). The peptoid CD spectra markedly

resemble those of right-handed peptide α-helices, which display a positive band around 192 nm and two minima of ellipticity around 222 and 208 nm (21). The measured ellipticity of the nπ* band of these peptoids is comparable to those of fully helical model peptides (peptoid $[\theta]_{218}$ = -2.5 to -3.5 x 10^4, vs. peptide $[\theta]_{222}$ = -2.8 to -3.5 x 10^4 deg·cm^2/dmol) (23, 25 and refs. therein). This strongly suggests that these peptoids exhibit helical secondary structure. Theoretical calculations indicate that helical secondary structures can display two minima between 230 and 200 nm due to exciton splitting of the ππ* amide transition (26). It should be noted that α-helix-like "Class C" CD spectra have also been associated with certain β-turn peptide structures (27). However, the magnitude of β-turn CD signals are typically of low intensity compared to the signals observed for peptide α-helices and in the present study.

The absence of a dependence of the CD signal on peptoid concentration and the sharp line-widths observed in the 1D ^{1}H NMR spectra (Fig. 3.4) indicate that the structure is monomeric and is not stabilized by inter-molecular associations. Preliminary analysis of 2D NMR spectra of α-chiral peptoid pentamers is consistent with the CD data and modeling studies (19). Based upon the pattern of both intra- and inter-residue ROE cross peaks, we identified a major conformer with a helical structure and repeating *cis* amide bonds (22).

We observe helix-like CD signals in a variety of peptoid sequences varying in length and in the content of charged, aromatic and chiral residues. Helix-like CD signals were preserved in sequences comprised of 33% to 100% aromatic residues and 0 to 66% charged residues without substantial change of signal intensity. However, when the aromatic groups were removed completely, as in the homo-oligomers containing morpholido **13** and

cyclohexyl **14** sidechains, CD signals were observed which lacked the characteristic double minimum. Chiral secondary structure can be propagated in the presence of achiral residues; dodecamer **5**, which contains 33% achiral *N*am residues, retains the helix-like profile. This should permit the synthesis of peptoid helices incorporating diverse sidechains from the wide family of achiral primary amines. Peptoids composed of sidechains with inverted chirality show a mirror image relationship between their CD spectra (Fig. 3.2). As found for random isocyanate polymers (24), this suggests that enantiomeric sidechains can induce the formation of helices with an inverted screw sense.

Peptoid secondary structure is detected even in short oligomers. A study of a series of peptoid homo-oligomers of one to eight residues demonstrates a helix-like CD signal in (*p*-nitrophenylethyl)glycine pentamers and octamers. A maximal signal is found at the pentamer length (Fig. 3.3). Although peptide α-helices are seldom detected for molecules smaller than 15 residues, the finding of stable helices in short polyimides is not unprecedented. Helices have been observed in polyproline molecules as small as three to five residues on the basis of vibrational and ultraviolet CD measurements (28). A strong single minimum CD signal observed for the peptoid trimer may indicate substantial helical proclivity for these chiral monomers in the absence of the inter-residue coupling attained by a full turn of the helical structure.

CD spectra were obtained in a variety of solvents, including acetonitrile, methanol, and aqueous solution. The intense characteristic helix-like spectra were retained, with some solvent-dependent variation of the positions of the double minima, indicating the resilience of the secondary structure to environmental changes. Previous studies of synthetic "folded"

polymers generally have been limited to solution studies in organic solvents (6, 9, 10). A critical difference between these "foldamers" and that of peptoids is the absence of hydrogen-bond formation as a driving force for peptoid secondary structure. Steric influence of the bulky chiral sidechains is likely to provide a constraint that defines the conformations accessible to the peptoid backbone (19). Interactions between sidechain aromatic groups, and dipole-dipole repulsion between sidechains and the carbonyl of the mainchain amides may also provide for specific secondary structure. These forces may allow peptoids to retain their structure in an aqueous environment, in which intra-molecular hydrogen-bond networks can be disrupted.

Peptoid secondary structure is extraordinarily resistant to thermal unfolding. The 9mer **11** and 30mer **6** show only a slight variation in CD signal intensity over the temperature range from 5 to 65 °C in neutral aqueous solution (Fig. 3.6). The resistance of peptoids to proteolysis has previously been noted (29). The stability of peptoids is an obvious benefit in considering their potential applications as nano-structured materials in biological systems.

Peptoids with ionizable sidechains can be de-stabilized in response to changes in pH conditions. Peptoid **6** shows little variation in CD signal intensity over the pH range 7 to 3 (Fig. 3.5). However, further titration to pH 2 results in a greatly diminished CD signal indicating complete unfolding. This pH dependent conformational behavior was exploited to observe melting in an experimentally accessible regime. Specifically, at pH 4.1 compound **6** demonstrates a complete, reversible loss of helix-like CD signal over a narrow temperature range (Fig. 3.7). The presence of an isodichroic point near 200 nm is characteristic of a two-state helix/coil transition, as has been described by Holtzer and Holtzer for polypeptides (30). These authors

also point out that it is highly unlikely that a precise isodichroic point will be observed if sidechain CD bands are making a significant contribution.

DSC demonstrates that thermal unfolding is accompanied by absorption of heat as secondary structure is melted (Fig. 3.8). The narrow peak in excess heat capacity is further evidence of the highly cooperative nature of peptoid unfolding. In fact, both CD and DSC studies demonstrate that **6** unfolds over a 40 °C temperature range, as compared with a greater than 100 °C temperature range for a 50 residue designed α-helical peptide (23).

Thermodynamic parameters obtained for the acid-destabilized melting of 30mer **6** are similar to those determined for helix/coil transitions in peptides. DSC shows that the calorimetric enthalpy of structure deformation in peptoid helices (ΔH_{cal} = 1.1 kcal/mol per residue) is comparable to that of α-helical peptides (ΔH_{cal} = 1.1 to 1.3 kcal/mol per residue) (23), as is the van't Hoff enthalpy derived from CD data (peptoid ΔH_{vH}= 0.22 kcal/mol residue, vs. peptide ΔH_{vH}=0.22 to 0.23 kcal/mol residue) (23). Notably, the enthalpy of peptide unfolding has been ascribed to the breaking of mainchain hydrogen bonds (23), which cannot be formed in peptoids. The ratio of $\Delta H_{vH}/\Delta H_{cal}$ of less than one indicates that the peptoid structural transition is not globally two-state. Thus, as found for α-helical peptides, we conclude unfolding is a locally two-state process in which partially unfolded species contain individual residues in either a helical or a coiled conformation (23, 30).

The work presented here demonstrates progress towards the creation of biomimetic polymers in which a specific sequence of chemically diverse monomers are incorporated into a stable folded structure. It is notable that polypeptoids display many of the characteristics of peptide behavior such as cooperative pH- and temperature-induced unfolding in aqueous solution, despite the absence of mainchain hydrogen bonds. More detailed studies of

peptoid structure and dynamics are currently underway. We believe polypeptoids will be useful for investigating principles of self-assembly in macromolecules, and for creating novel functional polymeric materials. We are currently evaluating the use of these materials as reagents for DNA and drug delivery. Peptoid nano-structures may prove to be of wide utility in the design and control of molecular architectures in biotic and abiotic systems.

References

1. Dill, K.A. (1985) *Biochemistry* **24**, 1501.

2. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D.P., Thomas, P. D., Chan, H. S. (1995) *Protein Sci.* **4**, 561.

3. Cho, C.Y., Moran, E.J., Cherry, S.R., Stephans, J.C., Fodor, S.P.A., Adams, C.L., Sundaram, A., Jacobs, J.W., Schultz, P.G. (1993) *Science* **261**, 1303.

4. Egholm, M., Buchardt, O., Nielsen, P.E., Berg, R.H. (1992) *J. Am. Chem. Soc.* **114**, 1895.

5. Hagihara, M., Anthony, N.J., Stout, T.J., Clardy, J., Schreiber, S.L. (1992) *J. Am. Chem. Soc.* **114**, 6568.

6. Hamuro, Y., Geib, S.J., Hamilton, A.D. (1996) *J. Am. Chem. Soc.* **118**, 7529.

7. Genarri, C., Salom, B., Potenza, D., Longari, C., Fioravanzo, E., Carugo, O., Sardone, N. (1996) *Chem. Eur. J.* **2**, 644.

8. Lokey, R. S., Iverson, B. L. (1995) *Nature (London)* **375**, 303.

9. Apella, D.H., Christianson, L.A., Karle, I.L., Powell, D. R., Gellman, S.H. (1996) *J. Am. Chem. Soc.* **118**, 13701.

10. Seebach, D., Overhand, M., Kühnle, F. N. M., Martinoni, B., Oberer, L., Hommel, U., Widmer, H. (1996) *Helv. Chim. Acta* **79**, 913.

11. Okamoto, Y., Nakano, T. (1994) *Chem. Rev.* **94**: 349.

12. 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., Bartlett, P.A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9367.

13. Zuckermann, R.N., Martin, E.J., Spellmeyer, D.C., Stauber, G.B., Shoemaker, K.R., Kerr, J.M., Figlozzi, G.M., Goff, D.A., Siani, M.A., Simon, R.J., Banville, S.C., Brown, E.G., Wang, L., Richter, L.S., Moos, W.H. (1994) *J. Med. Chem.* **37**, 2678.

14. Zuckermann, R.N., Kerr, J.M., Kent, S.B.W., Moos, W.H. (1992) *J. Am. Chem. Soc.* **114**, 10646.

15. Figliozzi, G.M., Goldsmith, R., Ng, S.C., Banville, S.C., Zuckermann, R.N. (1996) *Meth. Enzym.* **267**, 437.

16. Moehle, K. Hofmann, H.-J. (1996) *Biopolymers* **38**, 781.

17. Bradley, E.K. (1996) *J. Magn. Reson. Ser. B:* **110**, 195.

18. Melacini, G., Yangbo, F., Goodman, M. (1996) *J. Am. Chem. Soc.* **118**, 10725.

19. Armand, P., Kirshenbaum, K., Falicov, A., Dunbrack Jr., R. L., Dill, K. A., Zuckermann, R. N., Cohen, F. E. (1997) *Folding & Design*, **2**:369.

20. Kraphcho, A.P., Kuell, C.S. (1990) *Synth. Comm.* **20**, 2559.

21. Woody, R. W. (1985) in *The Peptides, Vol. 7*, ed. Hruby, V. J. (Academic Press, Orlando), pp. 16.

22. Armand, P., Kirshenbaum, K., Goldsmith, R.A., Farr-Jones, S., Barron, A.E., Truong, K.T.V., Dill, K.A., Mierke, D.F., Cohen, F.E., Zuckermann, R.N., Bradley, E.K. (1998) *Proc. Natl. Acad. Sci. USA.*. **95**:4309.

23. Scholtz, J.M., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M., Bolen, D.W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2854.

24. Green, M.M., Peterson, N.C., Sato, T., Teramoto, A., Cook, R., Lifson, S. (1995) *Science* **268**, 1860.

25. Bradley, E.K., Thomason, J.F., Cohen, F.E., Kosen, P.A., Kuntz, I.D. (1990) *J. Mol. Biol.* **215**, 607.

26. Moffitt, W., Fitts, D., Kirkwood, J. (1957) *Proc. Natl. Acad. Sci. USA* **43**, 723.

27. Perczel, A. Hollósi, M. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. Fasman, G.D. (Plenum, New York), pp. 285.

28. Dukor, R. K., Keiderling, T.A. (1991) *Biopolymers* **31**, 1747.

29. Miller, S.M., Simon, R.J., Ng., S., Zuckermann, R.N., Kerr, J.M., Moos, W.H. (1995) *Drug Dev. Res.* **35**, 20.

30. Holtzer, M.E., Holtzer, A. (1992) *Biopolymers* **32**: 1675.

Chapter 3:

NMR solution structure of an Nα-chiral pentapeptoid

This chapter was published as:

Armand P., Kirshenbaum K., Goldsmith R. A., Farr-Jones S., Barron A. E., Truong K. T. V., Dill K. A., Mierke D. F., Cohen F. E., Zuckermann R. N., Bradley E. K. (1998) "NMR Determination of the major solution conformation of a peptoid pentamer with chiral side chains" *Proc. Natl. Acad. Sci. USA* **95**:4309

Abstract

Polymers of N-substituted glycines ("peptoids") containing chiral centers at the α position of their side chains can form stable structures in solution. We studied a prototypical peptoid, consisting of five *para*-substituted (*S*)-*N*-(1-phenylethyl) glycine residues, by NMR spectroscopy. Multiple configurational isomers were observed, but, because of extensive signal overlap, only the major isomer containing all *cis* amide bonds was examined in detail. The NMR data for this molecule, in conjunction with previous CD spectroscopic results, indicate that the major species in methanol is a right-handed helix with *cis* amide bonds. The periodicity of the helix is three residues per turn, with a pitch of ~6Å. This conformation is similar to that anticipated by computational studies of a chiral peptoid octamer. The helical repeat orients the amide bond chromophores in a manner consistent with the intensity of the CD signal exhibited by this molecule. Many other chiral polypeptoids have similar CD spectra, suggesting that a whole family of peptoids containing chiral side chains is capable of adopting this secondary structure motif. Taken together, our experimental and theoretical studies of the structural properties of chiral peptoids lay the groundwork for the rational design of more complex polypeptoid molecules, with a variety of applications, ranging from nanostructures to non-viral gene delivery systems.

Introduction

Polymers of N-substituted glycines, termed peptoids, form a new class of biocompatible, synthetically accessible heteropolymers. Their sequence-specific, automated, and highly efficient synthesis has allowed the creation of combinatorial libraries of peptoid oligomers for drug-discovery (1,2,3). By using recent improvements in the efficiency of the coupling chemistry, peptoids up to 50 residues in length have been synthesized. Among those, cationic peptoid 36-mers have been identified which bind DNA, protect the DNA from nuclease digestion, and facilitate gene transfection (4).

We recently showed that many peptoids with chiral centers at the side chain α position have strong circular dichroic (CD) signals, indicating the presence of a regularly repeating conformation, in both aqueous and organic solvents (5). This structure is remarkably stable, as demonstrated by both CD and differential scanning calorimetry (DSC) measurements (5). A model of this conformation was recently proposed, based on the results of molecular mechanics and semi-empirical quantum mechanical calculations (6). Modeling predicted that peptoids containing (*S*)-*N*-(1-phenylethyl)glycine residues would form right-handed helices with a periodicity of approximately three residues per turn and *cis* amide bonds, similar to the polyproline type I conformation. In this conformation, the backbone carbonyls are aligned along the long helical axis, providing an explanation for the intensity of the helical CD signal.

Presented here is the structure determination by nuclear magnetic resonance (NMR) spectroscopy of the major solution isomer of compound **1**,

a pentapeptoid containing five chiral side chains (Fig. 4.1A). This peptoid displays the characteristic helix-like double minimum CD signal described by Kirshenbaum et al (5). The distance restraints obtained by NMR on **1** are best satisfied by a helical conformation which agrees with our previously published prediction (6). Identification of this structural motif and its CD signal should allow more directed studies of peptoid structural properties, and guide the design of more complex folded polypeptoid molecules for use in a variety of biological, chemical, and material science settings.

Figure 4.1:

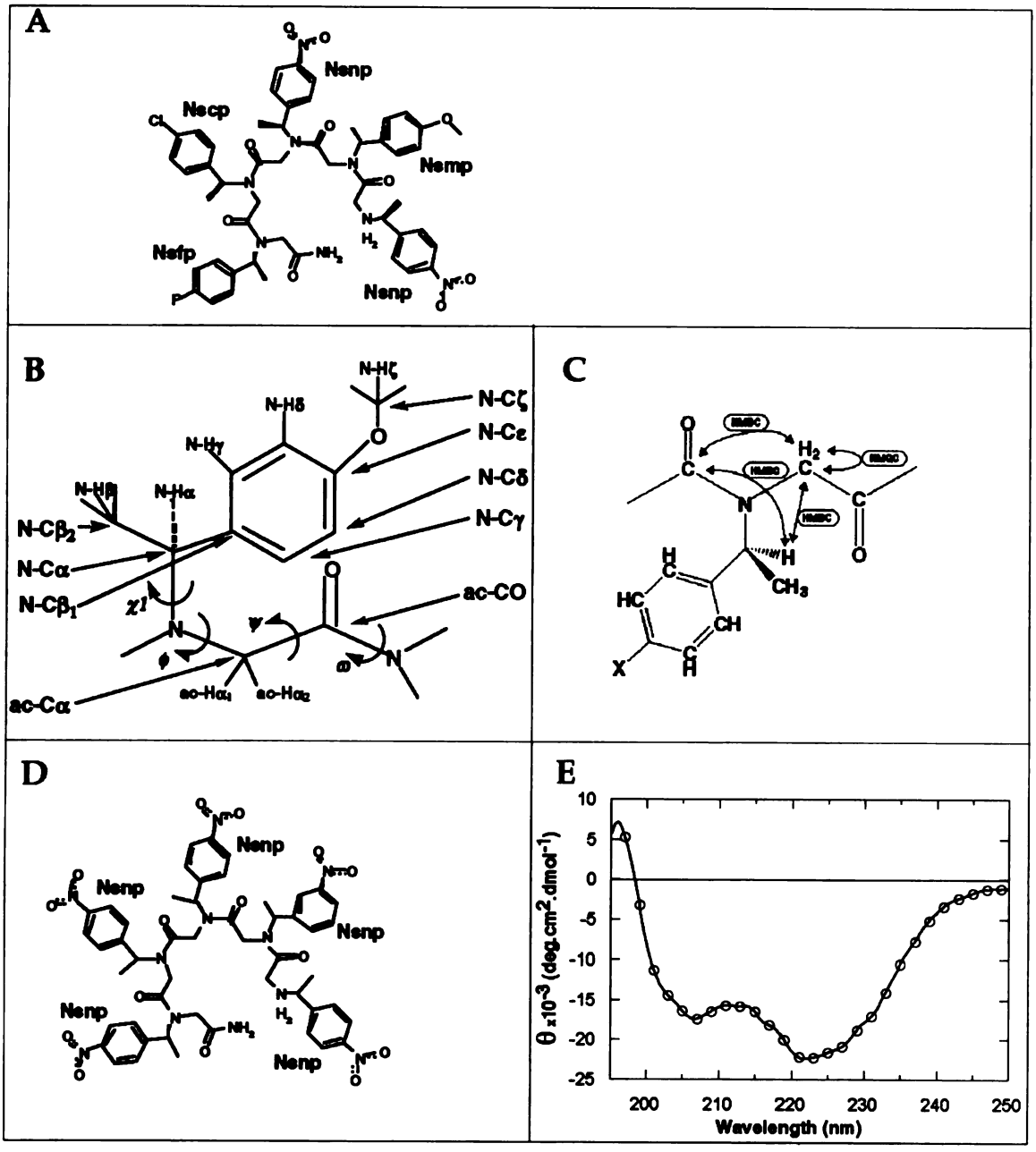
A. Compound **1** (*N*snp1-*N*smp2-*N*snp3-*N*scp4-*N*sfp5). *N*snp = (S)-*N*-(1-(*p*-nitrophenylethyl))glycine, *N*smp = (S)-*N*-(1-(*p*-methoxyphenylethyl))glycine, *N*scp = (S)-*N*-(1-(*p*-chlorophenylethyl))glycine, *N*sfp = (S)-*N*-(1-(*p*-fluorophenylethyl))glycine. The amide bonds are drawn in their *cis* configurations.

B. Atom and dihedral nomenclature. These are shown for *N*smp. The nomenclatures for the other residues are identical except for the absence of *N*smp-Cζ and *N*smp-Hζ atoms. Atom names follow (15); however, the side chain name subscript has been dropped when not referring to a particular residue (e.g., N-Cα instead of *N*xxx-Cα). Dihedral angles are named by analogy to those of peptides.

C. Simplified schematic representation of the assignment process for part of **1.** Arrows indicate the connections that were used and the spectrum types with which they were established.

D. Pentamer of (*S*)-*N*-(1-(*p*-nitrophenylethyl))glycine ((*N*snp)5).

E. CD spectrum of **1** in 100% methanol. The data were collected at 10°C and at a concentration of 0.1mM.


A
Nenp
Nscp
Nemp
Nsfp
Nenp
B
N-Hζ
N-Hδ
N-Hγ
N-Hβ
N-Hα
N-Cζ
N-Cε
N-Cδ
N-Cγ
N-Cβ2
N-Cα
N-Cβ1
χ1
ac-CO
ac-Cα
ac-Hα1
ac-Hα2
C
HMBC
D
Nenp
E
θ x10-3 (deg.cm2.dmol-1)
Wavelength (nm)

Materials and Methods

Synthesis, Purification, Circular Dichroism. The peptoid oligomers were synthesized and purified as described previously (5). The chiral amines used for the oligomer synthesis were obtained as follows: (*S*)-1-(*p*-nitrophenyl)ethylamine, (*S*)-1-(*p*-chlorophenyl)-ethylamine and (*S*)-1-(*p*-methoxyphenyl)ethylamine were obtained in >99% ee from Celgene (Warren, NJ); (*S*)-1-(*p*-fluorophenyl)ethylamine was obtained by successive recrystallizations of the tartrate salt of the racemic amine. Briefly, the L-tartrate salt of racemic 1-(*p*-fluorophenyl)ethylamine (Aldrich, Milwaukee, WI) was recrystallized three times from 9:1 MeOH:H_2O to obtain the desired *S* isomer in >99% ee as determined by GC of the Mosher amide (7). The ^{13}C-containing peptoids were synthesized by substituting bromoacetic-1,2-$^{13}C_2$ acid (Cambridge Isotope Labs, Andover, MA) for bromoacetic acid at the desired position of the growing oligomer. CD studies were performed as described in (5).

NMR Data Acquisition. All molecules were lyophilized from HPLC buffer and then re-lyophilized from deuterated acetonitrile-*d3* and then from deuterated methanol-*d4* before being dissolved in NMR solvents. NMR was performed at a peptoid concentration of ~5 mM in methanol-*d4* at 10°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 100 μM, 1 mM and 10 mM, to verify that spectra (chemical shift and populations) were concentration independent (data not shown). 2D homonuclear (DQCosy (8,9), TROESY (10)) and heteronuclear (HMQC (11), HMBC (12)) spectra were collected with the following parameters. The HMQC

experiments were performed with J = 140 Hz, and the HMBC experiments with J = 6, 9, and 12 Hz. The heteronuclear experiments were collected with spectral widths of 4000 Hz in the ^{1}H dimension, and 16000 Hz and 18000 Hz in the ^{13}C dimension for the HMQC and the HMBC, respectively. The ROE data were collected using the TROESY pulse sequence of Hwang and Shaka (10) with a 250 and 300 ms mixing time. The spinlock pulse was 180°(x) 180°(-x), with a field strength of 8000 Hz. The numbers of *t1* increments, transients, and *t2* complex data points, were respectively, 512, 256 and 2048 for the TROESY and natural abundance HMBC spectra, 256, 16 and 2048 for the HMQC spectra of the labelled compounds, and 512, 64 and 2048 for the HMBC spectra of those labelled compounds. Spectral widths of 4000 Hz were used in both dimensions for the TROESY and DQCosy experiments. Data processing was carried out with the NMR Pack software (13). Heteronuclear spectra were zero-filled to give a final real data matrix of 1024 points in ω2 (^{1}H) and 2048 points in ω1 (^{13}C). The HMQC was processed as phase sensitive in both dimensions, and the HMBC was processed as phase sensitive in ω2 and absolute value in ω1. Visualization of spectral data and resonance assignments were carried out with the NMR Pack (13) and SPARKY 3 (14) software.

Resonance Assignments. A modification of the strategy published by Bradley (15, 16) was employed for assigning ^{13}C and ^{1}H resonances. Each peptoid residue has two spin systems, the side chain branching from the main chain nitrogen, "Nxxx", and the main chain glycine unit, "ac" (Fig. 4.1B). While it is theoretically possible to assign these spin systems from natural abundance spectra, the dispersion of the ^{13}C and ^{1}H chemical shifts of the main chain "ac" subunits was too small to permit this. Therefore, 5 derivatives of

compound **1**, each containing a pair of ^{13}C-labels (Cα and CO) in a single "ac" subunit, were used to obtain those assignments.

The "ac" spin systems for each residue were assigned first, using the ^{13}C-HMQC spectra of the 5 individual molecules, each one labeled with ^{13}C at a different ac-Cα position. The main chain subunits were then connected to the side chains with the ^{13}C-HMBC spectra of the same individually labeled molecules. The connections used for the assignment procedure are shown schematically in Fig. 4.1C, and the heteronuclear spectra used to assign the main chain and side chain α protons of residue 1 are shown in Fig. 4.2. In the HMQC of the ^{13}C-labeled compounds, the cross peaks involving the labeled ac-Cα positions are much more intense than the peaks for carbons at natural abundance. Thus even with the significant chemical shift degeneracy seen for the main chain carbons and protons, the assignments can be made unambiguously. The same is true for the side chain Hα which is the only proton connected through 3 bonds to the ^{13}C-labeled ac-Cα (Fig. 4.1C).

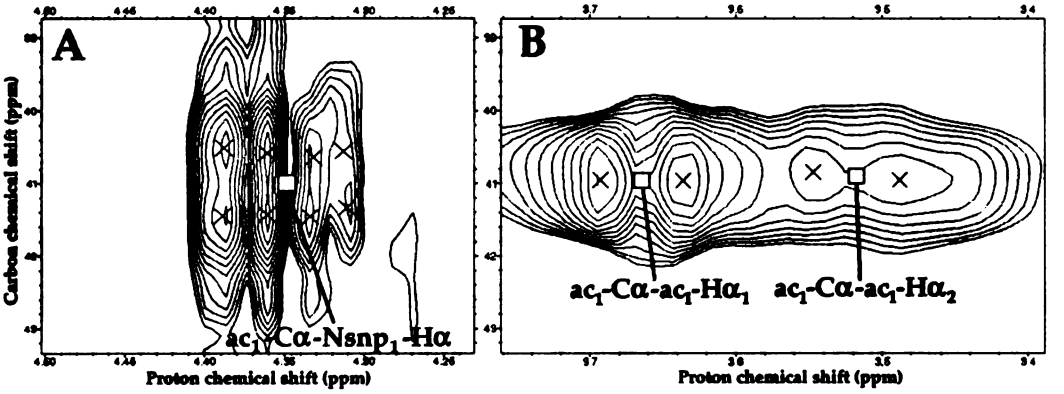
Although the experiments with the label at the Cα position clearly established the sequential assignments, the main chain connections were further validated by the observation of cross peaks in the HMBC spectra for the 3-bond connection between the ^{13}C carbonyl carbons and the following Nxxx-Hα across each amide bond (Fig. 4.1C). The side chains themselves were assigned with a combination of DQCosy and natural abundance HMBC spectra.

Figure 4.2. Detail of the heteronuclear spectra used to assign the resonances of the main chain and side chain α protons of residue 1

The spectra were collected on a molecule containing ^{13}C labels at the ac_1-Cα and ac_1-CO positions.

A. HMBC spectrum showing the ac_1-Cα–N_{snp1}-Hα cross peak for the major isomer.

B. HMQC spectrum showing the ac_1-Cα–ac_1-Hα cross peaks for the major isomer.



Refinement. All unambiguously assigned ROE cross peaks involving the major isomer were used to establish inter-proton distance bounds. Distance bounds were obtained by assuming an r^{-6} dependence of the cross peak intensities, calibrated on a set of intra-residue side chain cross peaks (Hα–Hβ and Hγ–Hδ), and adjusted for the use of pseudoatoms (17). This yielded a set of 15 distance restraints (Table 4.1), of which 7 were inter-residue. Refinement was attempted with this set of data without success; most of the structures generated with this restraint set had short inter-proton distances, which conflicted with the absence of the corresponding cross peaks in the TROESY spectra. In order to generate structures which better reflected the experimental data, a set of 51 "absent distance restraints", or ADRs, was added, in a manner similar to that described in (18); each ADR involved proton pairs which showed no cross peak in the TROESY spectra. Only pairs for which both protons appeared in other ROE cross peaks were included. This ensured that the absence of an ROE reflected a large inter proton distance, and not other factors such as exchange or relaxation processes. The ADRs were entered by setting the lower distance bounds at 3.0Å. These two sets of restraints were combined and used in metric matrix distance geometry calculations, following the random metrization algorithm of Havel (19). 1800 conformations that satisfied all distance bounds were generated. Each of these conformations was then subjected to constrained minimization, performed with the *Discover*® program using the CFF91 forcefield (Biosym/MSI, San Diego) without cross terms or charges, gradually scaling up the non-bond terms. All conformations which violated the ADRs or which violated the ROE restraints by more than 0.5Å were then excluded. Structures were visualized using the *Insight*® *II 95.0* molecular modeling system

Table 4.1: List of distance restraints used in structure refinement.

Residue 1	Atom 1	Residue 2	Atom 2	Upper Bound (Å)
		INTRA-RESIDUE		
ac2	Hα	Nsmp2	Hα	4.25
ac2	Hα	Nsmp2	Hβ	3.90
ac2	Hα	Nsmp2	Hγ	4.05
ac3	Hα	Nsnp3	Hβ	3.50
ac3	Hα	Nsnp3	Hγ	4.05
ac5	Hα	Nsfp5	Hα	4.05
ac5	Hα	Nsfp5	Hβ	3.90
ac5	Hα	Nsfp5	Hγ	3.90
		SHORT-RANGE (i to i+1)		
Nsmp2	Hγ	ac3	Hα	3.40
Nsmp2	Hδ	ac3	Hα	5.50
Nsmp2	Hβ	Nsnp3	Hδ	6.50
		MEDIUM-RANGE (i to i+2, i to i+3)		
Nsmp2	Hζ	Nsfp5	Hβ	3.90
Nsmp2	Hδ	Nsfp5	Hβ	4.25
Nsnp1	Hδ	Nsnp3	Hβ	5.25
Nsnp1	Hγ	Nsnp3	Hβ	8.00

Results

Selection of the Molecule. Compound **1** (Fig. 4.1A) was chosen because it shows the helix-like double minimum signal in its CD spectrum, and because its 1D proton NMR spectrum in methanol made it a promising candidate for structure determination. Methanol was used as the solvent because **1** is not soluble in water at the high concentrations required for NMR studies. The atom and dihedral angle nomenclature used in this report is given in Fig. 4.1B. Fig. 4.1E shows the CD spectrum of **1** in methanol.

Data Collection and Resonance Assignments. Both homonuclear and heteronuclear spectra of **1** in methanol-*d4* were collected, and used to assign the ^{1}H and ^{13}C chemical shifts of the major solution isomer. The presence of additional minor peaks for each proton was noted in all spectra (HMQC, HMBC, DQCosy, TROESY). This finding points to the existence of multiple species, most likely configurational isomers with different amide bond geometries. There appear to be two such minor isomers that are appreciably populated. Based on the normalized integration in a 1D spectrum of the peak assigned to the major isomer's N_{Snp3}-Hγ atom, it is estimated that this major isomer represents betweeen 50% and 60% of the molecules in solution.

The presence of those minor isomers posed the problem of unambiguously assigning all the peaks belonging to the major species. Both proton 1D peak intensities, which reflect relative populations, and TROESY intensities for the fixed distance intra side chain cross peaks (Hα-Hβ, Hα-Hγ, and Hβ-Hγ) were used to confirm that the peaks belonged to the major isomer (Table 4.2). Attempts to identify solvent and temperature conditions that would eliminate the minor isomers were unsuccessful. Attempts were

also made to assign the minor isomers with the use of exchange cross peaks from TROESY spectra acquired at higher temperatures, but the data were insufficient for a full assignment of the minor species. Thus, great care was taken to use only unambiguously assigned cross peaks of the major isomer in the refinement.

Table 4.2: Chemical shift assignments for the major isomer of 1.

Upper bounds are given after correcting for the use of pseudoatoms.

Note that peak overlap prevented us from assigning any cross peak involving the sidechain protons of residue 4.

Residue	Atom	Shift (ppm)	Residue	Atom	Shift (ppm)
SIDE CHAIN			MAIN CHAIN		
N_{snp1}	Hα	4.37	ac1	Hα	3.52 / 3.59
N_{snp1}	Hβ2	1.47	ac1	Cα	40.9
N_{snp1}	Hγ	7.51	ac1	CO	160.8
N_{smp2}	Hδ	8.04	ac2	Hα	3.58
N_{smp2}	Hα	5.57	ac2	Cα	39.3
N_{smp2}	Hβ2	1.13	ac2	CO	165.0
N_{smp2}	Hγ	6.91	ac3	Hα	3.68
N_{smp2}	Hδ	6.51	ac3	Cα	40.5
N_{smp2}	Hζ	3.31	ac3	CO	163.8
N_{snp3}	Hα	5.56	ac4	Hα	3.69
N_{snp3}	Hβ2	1.08	ac4	Cα	39.1
N_{snp3}	Hγ	7.22	ac4	CO	164.3
N_{snp3}	Hδ	7.87	ac5	Hα	3.38
N_{scp4}	Hα	5.37	ac5	Cα	38.5
N_{scp4}	Hβ2	1.00	ac5	CO	?
N_{scp4}	Hγ	6.78			
N_{scp4}	Hδ	6.92 / 7.00			
N_{sfp5}	Hα	5.29			
N_{sfp5}	Hβ2	0.95			
N_{sfp5}	Hγ	6.90			
N_{sfp5}	Hδ	6.71			

Once the resonances of the ac-Hα and side chain Hα were assigned for the major species in solution, it was possible to determine the geometry about each of the main chain amide bonds from the TROESY spectra. In this work, *cis* amide bonds will refer to geometries where the main chain Cαs are *cis* to each other. For residues with an intervening *cis* amide bond, ROE cross peaks are expected between the main chain protons of the two residues. Unfortunately, for the pentamer studied here, the main chain protons are almost degenerate and the resulting cross peaks are too close to the diagonal to be distinguishable except for residues 4 and 5, where the expected peak is present. If the amide bond were *trans*, ROE cross peaks would be present between the main chain protons of the residue preceding the amide bond and one or more of the side chain protons (Hα, Hβ, or Hγ) of the following residue. Those cross peaks would occur in the spectra at positions far off the diagonal, where they would be clearly visible. Yet those cross peaks are definitely absent from the spectra for the major isomer of **1**. It can therefore be concluded that the amide bonds of the major isomer of **1** all have a *cis* geometry.

Refinement. The refinement process yielded a set of 39 conformations with all *cis* amide bonds satisfying all absent distance restraints (ADRs) and all ROEs to within 0.5Å. Those conformations were analyzed in terms of conformational clusters. The term "cluster" will be used to refer to a group of conformations whose maximal pairwise root-mean-square (RMS) distance for the backbone heavy atoms and the side chain α carbons of the internal residues (2, 3 and 4) is less than 1.4Å. The 39 structures can be described as a single cluster, Cluster A (Fig. 4.3A). If a less stringent fit to the NMR data is allowed, by increasing the ROE violation tolerance to 1.0Å, an additional 101

structures are obtained, which have higher penalty function values. Of these 101 structures, 29 belong to Cluster A. The other 72 structures fall into 7 clusters, Clusters B-H, comprising 22, 21, 16, 7, 3, 2 and 1 structures, respectively.

Fig. 4.4A shows the distribution of the proper backbone dihedral angles of Cluster A and Clusters B-H. For the internal residues, the contour lines of the potential of an octamer of Nsnp, obtained from molecular mechanics (6), are superimposed on the plots. Points that lie on the upper right-hand side of the maps represent right-handed conformations. The dihedral angles of residues 2 and 3 are more tightly clustered than those involving residues 4 and 5. This is not surprising since there is more contact information on the N-terminal residues than on the C-terminal ones. The uncertainty about the C-terminus conformation (Figs. 4.3A and 4.4A) could therefore be due to true conformational flexibility or to a lack of ROE restraints.

The information content of our data was assessed in a series of refinements with partial restraint sets. A set consisting of the four medium-range restraints ($N_{snp3}H\beta$–$N_{snp1}H\delta$, $N_{snp3}H\beta$–$N_{snp1}H\gamma$, $N_{sfp5}H\beta$–$N_{smp2}H\delta$ and $N_{sfp5}H\beta$–$N_{smp2}H\zeta$) was sufficient to define the general fold and periodicity of the internal three residues. However, it could define neither the handedness nor the individual backbone dihedral angle values. The addition of the ADRs was required to narrow the distribution of the ϕ/ψ values, while the addition of the short-range distance restraints was necessary to detect a preference for the right-handed conformations.

Figure 4.3. Conformations remaining after restraint violation analysis

A. Cluster A. This describes the major conformation of **1** in methanol. Backbone carbons are shown in green, side chain carbons in light blue, nitrogen atoms in dark blue and oxygen atoms in red. The *para* substituents on the aromatic rings and the hydrogen atoms have been omitted for clarity.

B. Comparison of Cluster A conformations and previously published model (6). Only backbone atoms and side chain α carbons are shown. The model is shown as a thicker ball and stick molecule, and its carbon atoms are colored white.

A
1
2
3
4
5
B
1
2
3
4
5

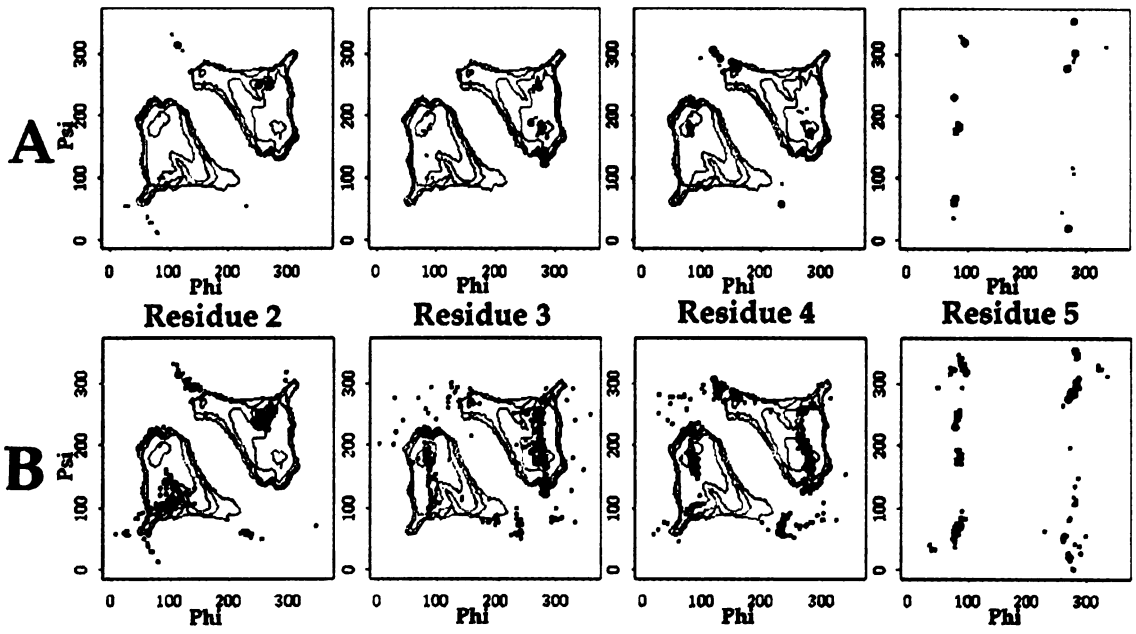
Figure 4.4. Distribution of the values of the main chain dihedral angles (ϕ and ψ) for residues 2 to 5

Residue 1 is not included because there is no proper ϕ dihedral angle.

A. Cluster A (large blue circles) and Clusters B-H (small red circles) after minimization, and before ensemble calculation.

B. Same, with Cluster A angles after ensemble calculation (small black circles).

White contour lines for residues 2-4 indicate the potential surface for Nsnp oligomers, after (6).


A
B
Psi
Phi
0
100
200
300
Residue 2
Residue 3
Residue 4
Residue 5

Other Isomers. We have only described the major isomer of **1**. However, it is clear from the NMR spectra that there are multiple species in slow exchange in solution. It is likely that they result from *cis-trans* isomerizations about the amide bonds (and therefore represent configurational isomerization). This is clearly the case for one minor isomer, where ROE cross peaks are found between a resonance of $N_{sfp5minor}$-Hα and $ac4_{minor}$-Hα, as well as between the corresponding $N_{sfp5minor}$-Hβ and $ac4_{minor}$-Hα. Since both the $N_{sfp5minor}$-Hα and the $ac4_{minor}$-Hα could be unambiguously assigned in the HMBC spectra of the compounds labeled at residues 4 and 3, these two cross peaks must reflect a *trans* amide bond at that position.

Ensemble Calculations. The application of the ROE distance restraints to each and every structure of an ensemble is not appropriate for cases in which conformational dynamics are taking place rapidly on the NMR time scale. One way to address this situation is by ensemble dynamics (20), in which the experimental observations are fulfilled by the average over the ensemble: each member of the ensemble must satisfy the constitutional constraints from the structural formula, while only the average over the ensemble must fulfill the experimental measurements. This method has been previously used to characterize equilibria between different conformations (20,21,22).

An ensemble calculation was performed on a set of 780 structures, containing 20 copies of each structure in Cluster A. All structures were allowed to independently explore conformational space, with only the requirement that the average pairwise distances must satisfy the distance restraints described above. This yielded a set of conformations compatible with the experimental data under a fast exchange regime. Fig. 4.4 compares the distribution of values for the ϕ and ψ dihedrals of residues 2-5, before and

after the ensemble calculation. It is clear that the regularity in the backbone dihedral angles of the starting conformations is decreased by the ensemble treatment. Nonetheless, a cluster analysis reveals that 354 of the 780 structures (45%) still belong to Cluster A. The remainder of the ensemble falls into 29 distinct clusters. Of these, 2 contain about 10% of the structures each, while the rest contain <5%, and typically <1% of the structures each. The predominance of Cluster A conformations is not the result of limited conformational sampling; indeed, when a control calculation was done without any distance restraints, fewer than 10% of the resulting structures belonged to Cluster A. Those results imply that the experimental data are most compatible with a highly populated conformational family (Cluster A), validating the results of the standard refinement method.

Discussion

During initial investigations into the biophysical characteristics of peptoids, a wide range of chiral molecules of varying length, charge, and polarity were synthesized. Many of those molecules display a double minimum around 205 and 220 nm in their CD spectrum. This signal was observed in both aqueous and organic solvents, and molecules that could be tested in both types of solvents showed no significant change in their CD spectrum in going from one to the other. Because this signal reflects the presence of a regularly repeating backbone conformation (5), it is likely that all molecules that display this double minimum adopt a similar backbone conformation. In order to determine this conformation by NMR, molecules with this CD signal were screened for chemical shift dispersion and sharp line widths in their 1D proton NMR spectra. A molecule that satisfied these criteria was a pentamer of (*S*)-*N*-(1-(p-nitrophenylethyl))glycine (*N*snp) (Fig. 4.1D). By analogy to peptides (20), a molecule of such short length could be expected to populate multiple conformations in solution. Nevertheless, this length was sufficient to give rise to a strong double minimum CD signal (5), indicating that the backbone conformation of interest was significantly populated. Moreover, the molecule was small enough to allow the resonance assignments necessary for 2D NMR structure determination. Since initial 2D NMR studies showed overlap in the aromatic region of the spectra, various substitutions were introduced at the *para* position of the side chain aromatic rings to increase chemical shift dispersion. This ultimately resulted in the selection of compound **1** (Fig. 4.1A) for this work.

Both standard and ensemble refinement methods suggest that the major conformation of **1** is accurately described by a single conformational

cluster (Cluster A, Fig. 4.3A). This cluster provides the most consistent fit to the experimental restraints. In addition, for most members of Cluster A, the backbone dihedral angles of the internal residues (2-4) are closely distributed around 270°, giving rise to regular, right-handed backbone conformations (Figs. 4.4A and 4.3A). This regular repeat is consistent with the ordering of the backbone chromophores demonstrated by the CD data (5). By contrast, most members of Clusters B-H have irregular backbones that are discordant with the intense, helix-like CD spectrum. Finally, the ensemble calculations show that Cluster A conformations remain by far the most prevalent even if other conformations are allowed to contribute to the experimental fit. This validates the assumption that the ROE cross peaks which were assigned to the major isomer can, to a first approximation, be treated as coming from a single conformation. Indeed, if another conformer were significantly populated in solution and rapidly exchanging with Cluster A conformations, the ensemble calculation would have revealed its presence. Since the cluster analysis of the ensemble does not indicate the presence of another important conformer, the ROE data most likely reflect a large population of tightly clustered conformations.

Based on those results, the major conformation of **1** in methanol can be described as having a helical shape, a right-handed twist, a periodicity of three residues per turn, and a pitch of ~6Å. The amide bonds are all *cis*. The χ_1 angles (defined as the $C_{(i-1)}$,$N_{(i)}$,N-$C\alpha_{(i)}$,N-$C\beta1_{(i)}$ dihedral angles) of residues 2-5 are all approximately -120°. This conformation is very similar to the model previously proposed (6) for an oligomer of (*S*)-*N*-(1-phenylethyl)glycine (Fig. 4.3B). Indeed, the handedness, periodicity, pitch, amide bond geometries, and χ_1 angles were all accurately predicted. The

average backbone RMS difference between the model and the conformations in Cluster A is 1.6Å, or 1.1Å if only the internal residues are considered.

The simple molecular mechanics and more detailed quantum mechanical calculations used in the modeling suggested that steric terms figure prominently in the conformational preferences of Nα-chiral peptoids (6). The concordance of the model with the conformation determined by NMR strengthens this hypothesis. Thus, although these results cannot establish the nature of the driving forces in peptoid folding, they do imply that steric interactions play a prominent role in determining the folded conformations of peptoids containing chiral centers at the side chains' α position. Conversely, because CD spectroscopy indicates that this conformation is very similar in aqueous and organic solvents (5), it seems unlikely that the solvent plays a major role in defining the folded conformation. This result can be contrasted to the behavior of polyproline, another polymer without hydrogen bonds, in which solvent seems to have a major impact on the polymer's conformational preferences (23).

Another important result is the presence of a minor isomer of **1** with a *trans* amide bond. This implies that the energetic difference between the *cis* and *trans* geometries is not large enough to result in the exclusive population of one geometry, and proves that *cis/trans* isomerization is a source of conformational flexibility in this molecule. The population of both *cis* and *trans* amide bonds is not surprising, as N-alkylation should lower the energetic difference between *cis* and *trans* forms of peptoids (5,24), as it does for proline (25). *Cis/trans* isomerizations have been noted in small proline-containing peptides in solution (21), and in previous NMR studies of peptoids (16,26). It is possible that other amide bonds also adopt *trans* geometries in the minor isomers of **1**. However, modeling suggests that the

amide bond of residue 4 could have a particular reason for favoring a *trans* isomer: this geometry could allow the C-terminal amide hydrogen atoms to form a hydrogen bond with the carbonyl oxygen of residue 4, if residues 2-4 adopt the dihedral angle values of the major isomer (data not shown).

While this isomerization occurs on a slow time scale, the major conformation of **1** is also likely to fluctuate on a fast time scale. The ensemble calculations help to define the flexibility of this isomer. It suggests that while the right-handed conformation is the most populated in solution, other conformations with less regular geometries or opposite handedness may also be populated and exchange with the right-handed helix rapidly relative to the NMR time scale (Fig. 4.4B). Since the penalty function in the ensemble calculation consists only of distance restraints and hard-sphere repulsions, it would be inappropriate to infer the relative energies of the different conformations from their relative populations in the ensemble.

Conclusion

The NMR data for the chiral pentapeptoid **1** strongly suggest that the major conformation in methanol consists of a right-handed helix with three residues per turn, a pitch of ~6Å, and *cis* amide bonds. This is consistent with the CD data for this molecule. Since many oligopeptoids containing chiral centers at the side chains' α position have almost identical CD characteristics as **1**, the structure proposed here may also describe the major conformation adopted by those other peptoids. Hence, this represents an important step in the study of peptoid structure. Together with the success of standard modeling methods in predicting the general features of this conformation, this work opens a very promising path to rational peptoid design, which we are currently pursuing. The helical motif described here could eventually form the structural basis for a new class of biomimetic molecules with applications ranging broadly throughout chemistry and biology.

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., Bartlett, P. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9367-9371.

2. 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. C., Brown, E. G., Wang, L., Richter, L. S., Moos, W. H. (1994) *J. Med. Chem.* **37,** 2678-2685.

3. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H., Moos, W. H. (1992) *J. Am. Chem. Soc.* **114,** 10646-10647.

4. Murphy, J. E., Uno, T., Hamer, J. D., Cohen, F. E., Dwarki, V., Zuckermann, R. N. (1997) *Proc. Natl. Acad, Sci. USA* , in press.

5. Kirshenbaum, K., Barron, A. E., Goldsmith, R. E., Armand, P. A., Bradley, E. K., Truong, K. T. V., Dill, K. A., Cohen, F. E., Zuckermann, R. N. (1997) *Proc. Natl. Acad Sci. USA.*, submitted.

6. Armand, P., Kirshenbaum, K., Falicov, A., Dunbrack, R. L. Jr, Dill, K. A., Zuckermann, R. N., Cohen, F. E. (1997) *Folding & Design* **2,** 369-375.

7. Dale J. A., Mosher, H. S. (1973) *J. Am. Chem. Soc.* **95,** 512-519.

8. Marion, D., Wuthrich, K., (1983) *Biochem. Biophys. Res. Commun.* **113,** 967-974.

9. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117,** 479-485.

10. Hwang, T.-L., Shaka, A. J. (1992) *J. Am. Chem. Soc.* **114,** 3157-3159.

11. Mueller, L. (1979) *J. Am. Chem. Soc.* **101,** 4481-4484.

12. Bax, A., Summers, M. F. (1986) *J. Am. Chem. Soc.* **107,** 2093-2094.

13. Day, M., Kneller, D., Kuntz, I. D. (1993) NMR Pack: The programs are available from the Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143.

14. Goddard, T. G., Kneller D. G., SPARKY 3 NMR Assignment and Integration Software, University of California, San Francisco, CA 1997.

15. Bradley, E. K. (1996) *J. Magn. Res. Ser. B* **110,** 195-197.

16. Bradley, E. K., Kerr, J. M., Richter, L. S., Figliozzi, G. M., Goff, D. A., Zuckermann, R. N., Spellmeyer, D. C., Blaney, J. M. (1997) *Molecular Diversity*, in press.

17. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

18. Blackledge, M. J., Brüschweiler, R., Griesinger, C., Schmidt, J. M., Xu, P., Ernst, R. R. (1993) *Biochemistry* **32,** 10960-10974.

19. Havel, T. F. (1991) *Prog. Phys. Mol. Biol.* **56,** 43-78.

20. Pellegrini, M., Gobbo, M., Rocchi, R., Peggion, E., Mammi, S., Mierke, D. F. (1996) *Biopolymers* **40,** 561-569.

21. Mierke, D. F., Kurz, M., Kessler, H. (1994) *J. Am. Chem. Soc.* **116,** 1042-1049.

22. Pellegrini, M., Mammi, S., Peggion, E., Mierke, D. F. (1997) *J. Med. Chem.* **40,** 92-98.

23. Harrington, W.F., Sela, M. (1958) *Biochim. Biophys. Acta* **27**:24-41

24. Moehle, K., Hofmann, H.-J. (1996). *Biopolymers* **38,** 781-790.

25. Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*. (2nd ed.), W. H. Freeman, New York.

26. Melacini, G., Feng, Y., Goodman, M. (1996) *J. Am. Chem. Soc.* **118,** 10725-10732.

Chapter 4:

Design of a peptoid helix bundle

Abstract

Nα-chiral oligopeptoids can form helices in solution. By analogy to proteins, we consider these helices a type of peptoid secondary structure. We are interested in combining these secondary structure elements in more complex tertiary structures. Since the general conformational features of the oligopeptoid helices could be anticipated by molecular mechanics computational modeling, we have used similar modeling techniques to design peptoid helix bundles. Using a strategy of energy minimization in sequence and structure space, we have obtained putative sequences for four-helix bundles. We have synthesized the molecules and performed preliminary assays to screen for a bundle fold.

Introduction

We are pursuing the goal of developing N-substituted glycine polymers (peptoids) into a biomimetic system. The results presented in the preceding chapters indicate that some peptoids have sequence-specific structure. Specifically, peptoid oligomers containing Nα-chiral, *N*spe-like side chains form helices in solution. By analogy to proteins, such helices can be considered to form a subset of peptoid secondary structure. Also by analogy to proteins, it is tempting to speculate on the possibility of forming peptoid tertiary structure, by combining multiple secondary structure elements into one fold. The ability to engineer complex three-dimensional structures in an artificial polymeric system would be unique. It would afford greater flexibility in trying to position specific chemical moieties in three-dimensional space, and thereby broaden the functional potential of peptoids.

To date, we only have evidence of helical secondary structure. This may be a reflection of the conformational properties of the peptoid molecules that we have investigated; alternatively, it may be a reflection of our experimental approach to detecting structure. Indeed, our focus on Nα-chiral, aromatic side chains stems from the original CD results with *N*spe homopolymers (see main Introduction), which displayed such striking CD spectra. The double minimum in the spectra of these molecules was so distinct, and so clearly reminiscent of the also distinct α-helical CD signature, that *N*spe oligomers were natural starting points for our structural studies. But there may be other types of peptoid secondary structures which would not display double minima in their CD spectra; to identify these molecules would require either more quantitative CD studies, or the use of other structural probes.

At any rate, given the current state of our knowledge about peptoid secondary structure, the most accessible tertiary structure is likely to be a helix bundle. Hence we have devised a strategy to design a peptoid helix bundle soluble in aqueous solution, and to screen the designs experimentally.

Much work has gone into comparable efforts in protein design. So far, the protein helix bundle has been the most common design target, and there are now many successes in creating such folds (many of which are reviewed in 1,2,3). Broadly, these can be divided into five categories.

The first category is based on the noncovalent association of amphiphilic α-helical peptides. Work by DeGrado and co-workers provides a model for such work. Amphiphilic sequences were first designed with leucine cores, and their aggregation was monitored by CD spectroscopy or size-exclusion chromatography (4,5). Some of these aggregates were studied by X-ray crystallography (6) and NMR (7,8). Loops were then added to join the elements into a single chain (5,9). Further refinements were introduced in the core of these bundles to get native-like properties (10,11). Another example of a bundle based on noncovalent association of amphiphilic subunits is the "peptidergent" four-helix bundle (12).

A second category includes *de novo* attempts at direct bundle design from empirical folding rules and computational modeling. Examples include the four-helix bundle "Felix" (13), ALIN (14,15) and a four-stranded coiled-coil bundle (16).

Another type of bundle design relies on cyclic templates upon which the helices are assembled. The base serves to anchor the helices in the right orientation and geometry. Examples are TASPS (17) and tetraphilin (18).

A related approach uses metal binding to stabilize the bundle fold (19,20,21,22,23,24,25). Interestingly, in one case the addition of a metal-binding site stabilized the fold even in the absence of the metal (26).

Finally, Hecht and co-workers developed an elegant strategy to synthesize many potential helix bundles through a combinatorial genetic approach, followed by a size-exclusion chromatography and CD spectroscopic screen for folding (27).

In order to synthesize a peptoid helix bundle, we have developed a strategy that borrows from some of the aforementioned approaches. This strategy relies heavily on computational modeling to guide the design, while using a combinatorial experimental method to increase the odds of success. Computational techniques have found success in guiding design efforts (some reviewed in 28), both in *de novo* designs of folds (16,29), and in the modification of existing proteins by core repacking (30,31,32) or addition of new functionality (33,34,35).

Our approach follows conceptually the hierarchical design approach developed by DeGrado and co-workers (2), but attempts to perform the first steps computationally. We model the non-covalent association of amphiphilic peptoid helices in a molecular mechanics force field, optimizing sequence and geometry simultaneously. Once a suitable set of helices has been found, we add the loops as permissive elements that can bridge the inter-helical spaces. We then perform additional sequence-structure minimization on the bundles thus created.

Experimentally, we have chosen an approach which allows some combinatorial expansion in the number of screened sequences. Instead of designing and synthesizing the bundles as single chains of four helices and

three loops, we designed the bundles as two helix-loop-helix elements designed to associate non-covalently. We synthesized multiple candidates for each half and screened all possible pairs for the adoption of a bundle fold.

Results

Computational design

We first decided on a number of helices for the bundle, based on simple geometric considerations. We looked for the smallest number of helices such that the outside, hydrophilic residues (a combination of *N*snp, *N*sce, and *N*ae residues; for nomenclature see Table 3.1) would provide cover for the inside, hydrophobic residues. Two- and three-helix bundles failed this test, as large areas of the core were left exposed to solvent. Four-helix bundles, by contrast, had an adequately protected hydrophobic core when all four helices were aligned roughly along the long axis of the bundle. We also picked maximal helix lengths of 15 residues (5 turns) based on synthetic concerns.

The five-step computational design approach is summarized graphically in Fig. 5.1 and described below.

Step 1. Four right-handed amphiphilic peptoid helices were created. The outside residues were all set to be *N*snp, and the inside residues to be *N*-isopropyl-glycine (*N*ip). The backbone and side chain dihedral torsion angles were set to their minimum energy values. The four helices were placed next to each other, with closest neighbors in an antiparallel orientation to each other (Fig. 5.1A).

Step 2. A first energy minimization was performed. The inside residues were allowed to vary among a limited set of hydrophobic possibilities (*N*spe, *N*-phenyl-ethyl-glycine (*N*pe), *N*-isopropyl-glycine (*N*ip), *N*-isobutyl-glycine (*N*ib), and *N*-butyl-glycine (*N*bu)). The outside residues were fixed. Backbone and side chain dihedrals were allowed to vary around their minimum-energy position. The helices were also allowed to move

relative to each other, by translation along all three axes and by rotation around three axes. The force field consisted of the non-bonded components of the AMBER molecular mechanics force field (36) and a solvation potential based on solvent-accessible surface area (37). This force field was meant to promote formation of a well-packed hydrophobic core without major steric clashes. This resulted in four helices in a favorable bundle geometry (Fig. 5.1B).

Step 3. Loops were added to the model. Polyglycine stretches of various lengths were added in all possible registers between neighboring helices to determine the minimum loop length between two subsequent helices that would avoid steric clashes (Fig. 5.1C).

Step 4. The bundles thus obtained were re-minimized with various annealing schedules and different relative weights between the non-bonded and solvation components of the energy (Fig. 5.1D). In some cases, a compactness term was added to the force field.

Step 5. Finally, the outside residues were manually optimized (Fig. 5.1E) to balance solubility and helicity, while minimizing electrostatic repulsion.

This method was first used to get a single-chain bundle sequence. Since the synthetic method only allowed the use of 6 monomers, we changed all *N*pe residues to *N*spe, and all *N*ip and *N*bu residues to *N*ib. The resulting sequence was a 56-mer named **Bundle1**, which is shown in Table 5.1, line 1. A model of the core of this bundle is shown in Fig. 5.2. Unfortunately, we were not able to synthesize this sequence in adequate yield. Even though other 50-mers had been synthesized, the monomers chosen for **Bundle1** did not incorporate with high enough efficiency to allow synthesis of a 56-mer.

Figure 5.1: Design Approach

Helices are represented as gray cylinders, loops as black lines joining them. External (hydrophilic) residues are red, internal (hydrophobic) ones are green.

A. Four amphiphilic peptoid helices are modeled and placed next to each other.

B. A first sequence-structure minimization is performed to optimize sequence and geometry.

C. An appropriate set of permissive poly-glycine loops is added.

D. The bundles are re-minimized with the loops.

E. The outside residues are optimized.

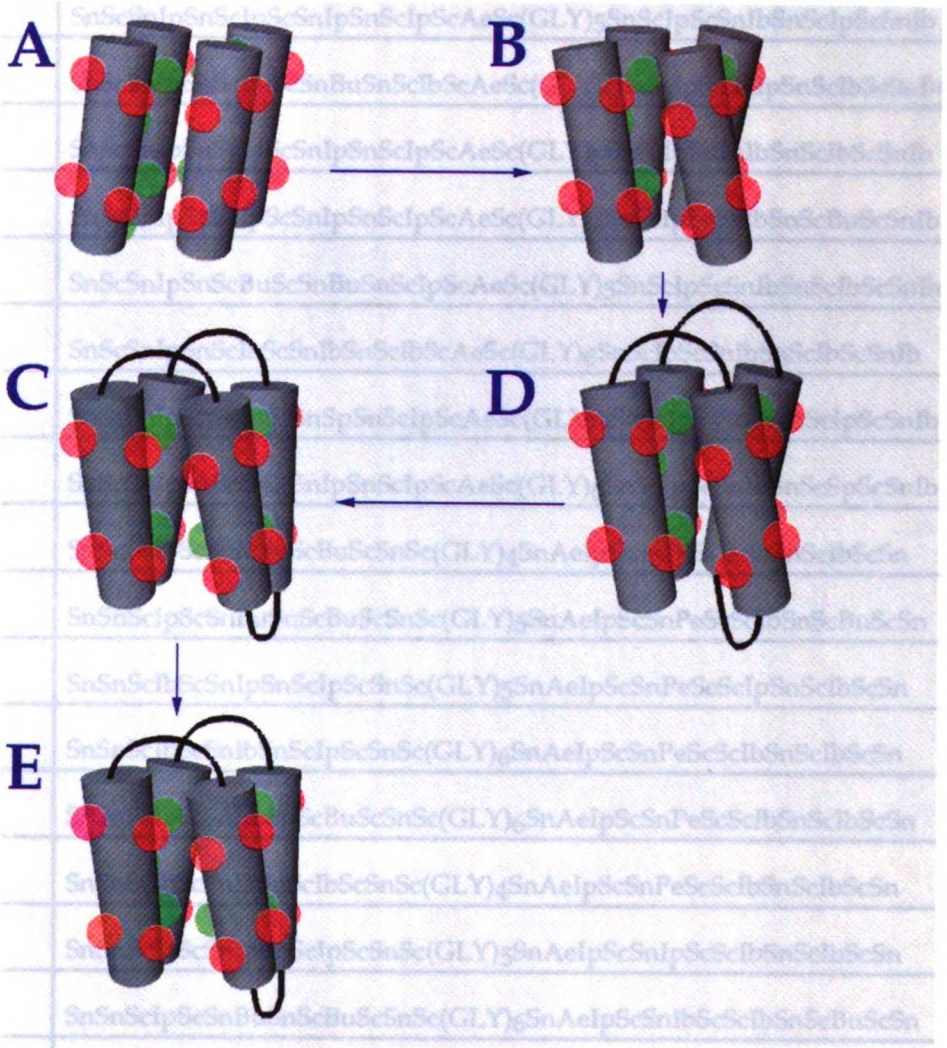

A
B
C
D
E

Table 5.1: Putative bundle sequences

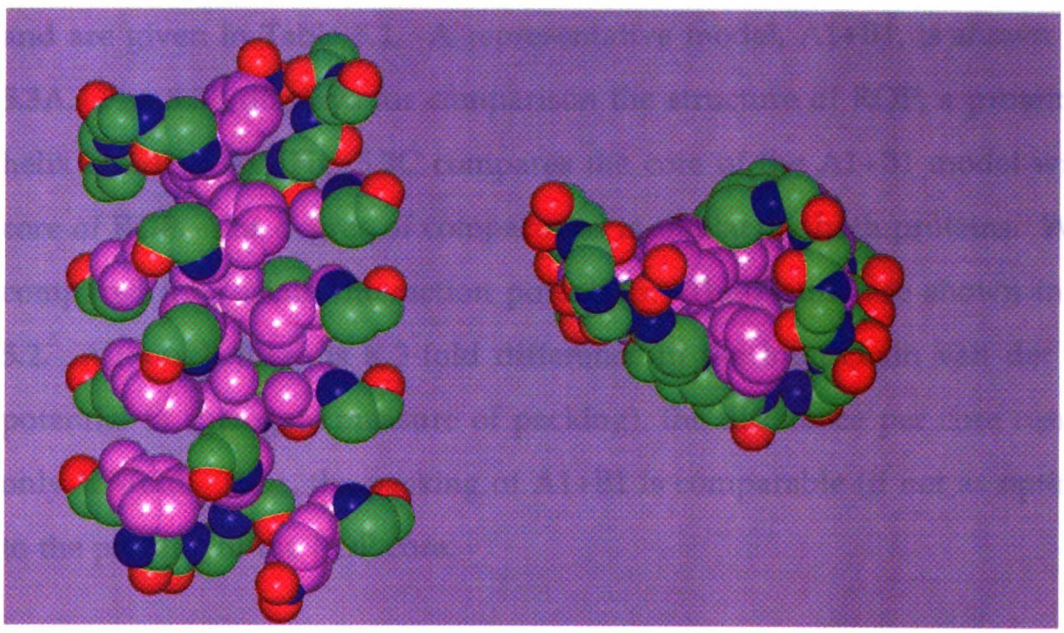
Name[1]	Sequence[2]	Syn[3]
Bundle1	ScSnSpSnScIbSnScSpScScSn(GLY)$_5$ScSnIbScScSpSnScSpSn(GLY)$_4$ SnScIbSnScSpScSnAe (GLY)$_4$ SnScSpScScSpScSnIbSnScSc	no
A1	SnScSnIpSnScIpScSnIpSnScIpScAeSc(GLY)$_4$SnScIbScSnIbSnScIpScSnIb	yes
A2	SnScSnIpSnScIpScSnIpSnScIbScAeSc(GLY)$_5$SnScIpScSnIbSnScBuScSnIb	yes
A3	SnScSnIpSnScIpScSnIpSnScIpScAeSc(GLY)$_5$SnScIpScSnIbSnScIpScSnIb	yes
A4	SnScSnIpSnScBuScSnBuSnScIbScAeSc(GLY)$_6$SnScIpScSnIpSnScIbScSnBu	yes
A5	SnScSnIpSnScIpScSnIpSnScIpScAeSc(GLY)$_6$SnScIpScSnIbSnScIbScSnIb	yes
A6	SnScSnIpSnScIpScSnIpSnScIpScAeSc(GLY)$_4$SnScIpScSnIbSnScBuScSnIb	yes
A7	SnScSnIpSnScBuScSnBuSnScIpScAeSc(GLY)$_5$SnScIpScSnIbSnScIbScSnBu	yes
A8	SnScSnIpSnScIbScSnIbSnScIbScAeSc(GLY)$_6$SnScIbScSnIbSnScIbScSnIb	no
A9	SnScSnIpSnScIpScSnSpSnScIpScAeSc(GLY)$_5$SnScIpScSnIbSnScIpScSnIb	yes
A10	SnScSnIpSnScIpScSnIpSnScIpScAeSc(GLY)$_6$SnScIpScSnIbSnScSpScSnIb	yes
B1	SnSnScIbScSnIbSnScBuScSnSc(GLY)$_4$SnAeIpScSnPeScScIbSnScIbScSn	yes
B2	SnSnScIpScSnBuSnScBuScSnSc(GLY)$_5$SnAeIpScSnPeScScIbSnScBuScSn	yes
B3	SnSnScIbScSnIpSnScIpScSnSc(GLY)$_5$SnAeIpScSnPeScScIpSnScIbScSn	yes
B4	SnSnScIbScSnIbSnScIpScSnSc(GLY)$_6$SnAeIpScSnPeScScIbSnScIbScSn	yes
B5	SnSnScIbScSnBuSnScBuScSnSc(GLY)$_6$SnAeIpScSnPeScScIbSnScIbScSn	yes
B6	SnSnScIbScSnIbSnScIbScSnSc(GLY)$_4$SnAeIpScSnPeScScIbSnScIbScSn	yes
B7	SnSnScIbScSnIpSnScIpScSnSc(GLY)$_5$SnAeIpScSnIpScScIbSnScIbScSn	no
B8	SnSnScIpScSnBuSnScBuScSnSc(GLY)$_6$SnAeIpScSnIbScScIbSnScBuScSn	yes
B9	SnSnScIbScSnIpSnScIpScSnSc(GLY)$_5$SnAeIpScSnIbScScIpSnScIbScSn	yes
B10	SnSnScIbScSnIpSnScIpScSnSc(GLY)$_6$SnAeIpScSnIpScScIbSnScIbScSn	no

[1]See text

[2]Letters refer to the following residues (for residue nomenclature see Table 3.1):
Sp=*N*spe, Pe=*N*pe, Sn=*N*snp, Ip=*N*ip, Ib=*N*ib, Bu=*N*bu, Ae=*N*ae, Sc=*N*sce,GLY=Glycine

[3]Syn: whether molecule was synthesized properly, according to analytical HPLC and MS analysis

Figure 5.2: Model of the hydrophobic core of Bundle1
Oxygen atoms are red, nitrogen atoms are blue, backbone carbon atoms are green, side chain carbons are pink. Hydrogen atoms have been omitted for clarity. *Left*, viewed perpendicular to the bundle's long axis. *Right*, viewed parallel to the bundle's long axis.



We therefore ran the method to get half-bundle sequences, that is, sequences spanning only two helices and one intervening loop. This would alleviate synthetic difficulties as well as allow combinatorial screening of all the possible pairs. Starting with the geometry obtained in Step 2, we picked two pairs of helices to join together, and proceeded with the remaining three steps. By using various parameters in Step 5 and combining some of the solutions, we generated 10 sequences for each half, with varying loop lengths. We also had to modify some of the sequences based on synthetic requirements. The resulting sequences are named A1 to A10 and B1 to B10, and are given in Table 5.1. A representative model, A1+B1, is shown in Fig. 5.3A. Fig. 5.3B displays for comparison the structure of ROP, a protein four-helix bundle (38). Fig. 5.3C compares the core of the A1+B1 model with the core of ROP. The degree of compactness is similar for both proteins. We also compared inter-chain interaction potentials. The results are shown in Table 5.2. Although there is a 3-fold difference in the inter-chain van der Waals potential (a surrogate measure of packing), the difference per core residue is only 1.7-fold. Thus, the packing of A1+B1 is comparable (if not as optimized) to the packing of the ROP core.

Table 5.2: Comparison of inter-chain potentials

Potentials were calculated using the AMBER non-bonded potential

Molecule	# of core residues	vdWaals Total	Electrostatic Total	vdWaals per residue	Electrostatic per residue
A1+B1	15	-30.8	0.69	-2.1	-0.0
ROP	27	-93.16	-3.39	-3.5	-0.1

vdWaals=van der Waals potential

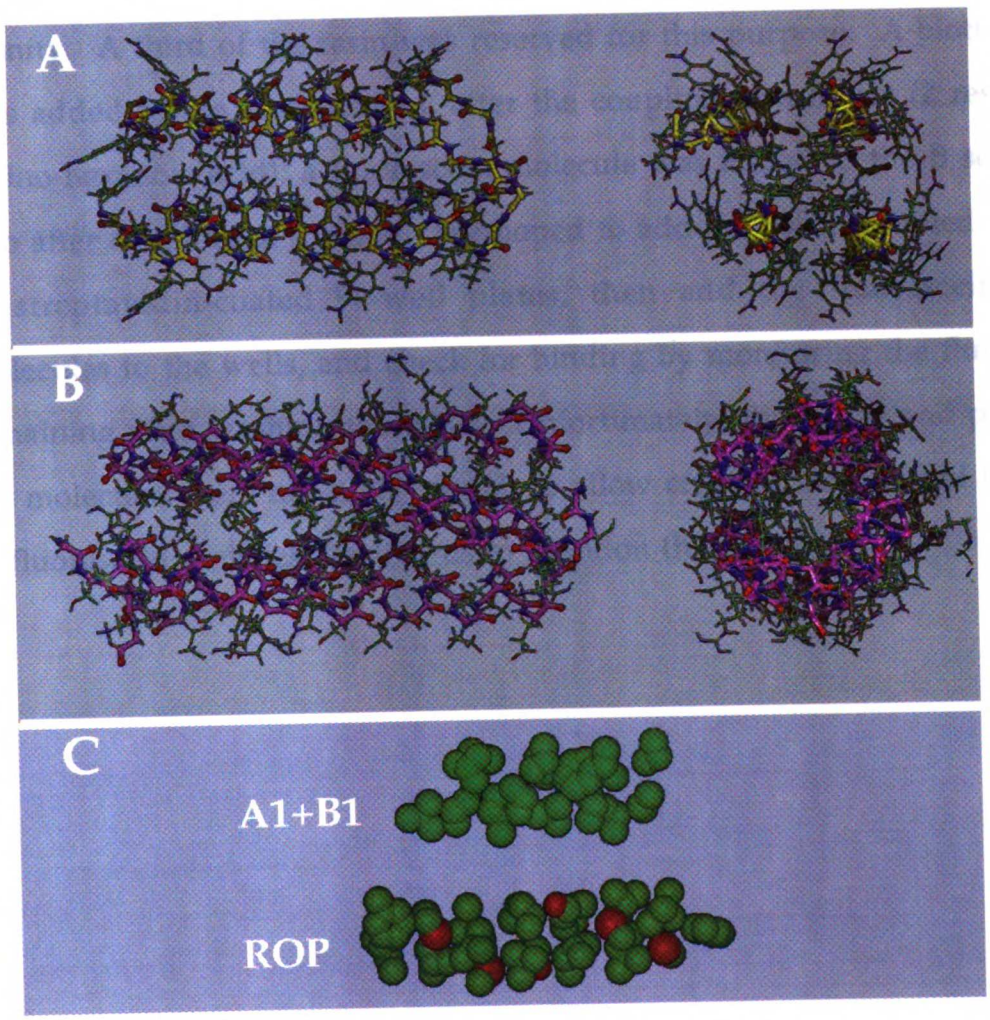
Figure 5.3: Comparison of A1+B1 bundle model and ROP

A. Model of A1+B1 bundle.

B. Structure of ROP (38).

Oxygen atoms are red, nitrogen atoms are blue, backbone carbon atoms are green, side chain carbons are yellow for A1+B1, pink for ROP. Hydrogen atoms have been omitted for clarity. *Left*, viewed perpendicular to the bundle's long axis. *Right*, viewed parallel to the bundle's long axis.

C. Comparison of the hydrophobic cores. Only the side chain atoms of the core residues are shown (carbon atoms in green, oxygen atoms in red).


A
B
C
A1+B1
ROP

Synthesis

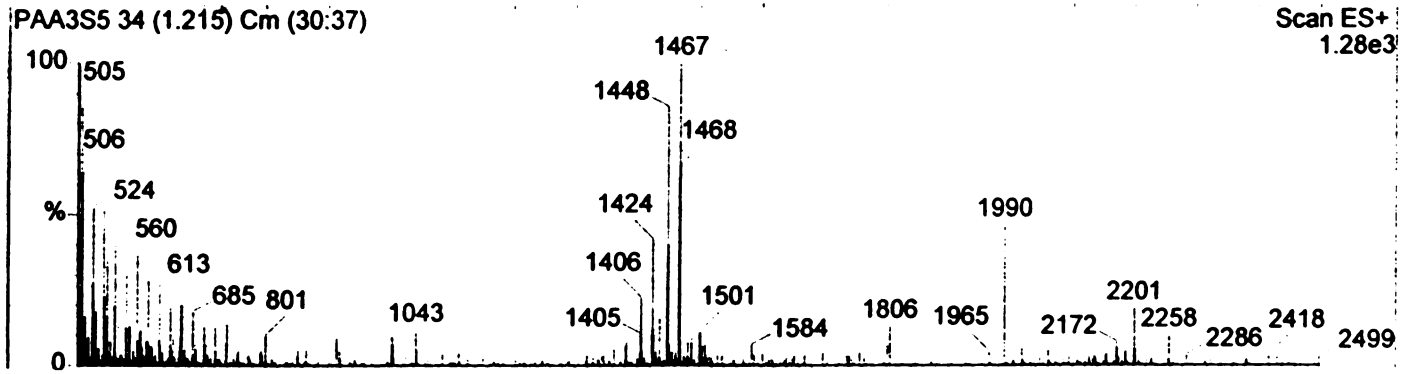
The twenty sequences shown in Table 5.1 (lines 2 to 21) were synthesized, and the products were analyzed by HPLC and electrospray mass spectrometry. Of the 10 A sequences, 9 were synthesized properly, as were 8 of the 10 B sequences. A representative analysis, of A3, is shown in Fig. 5.4.

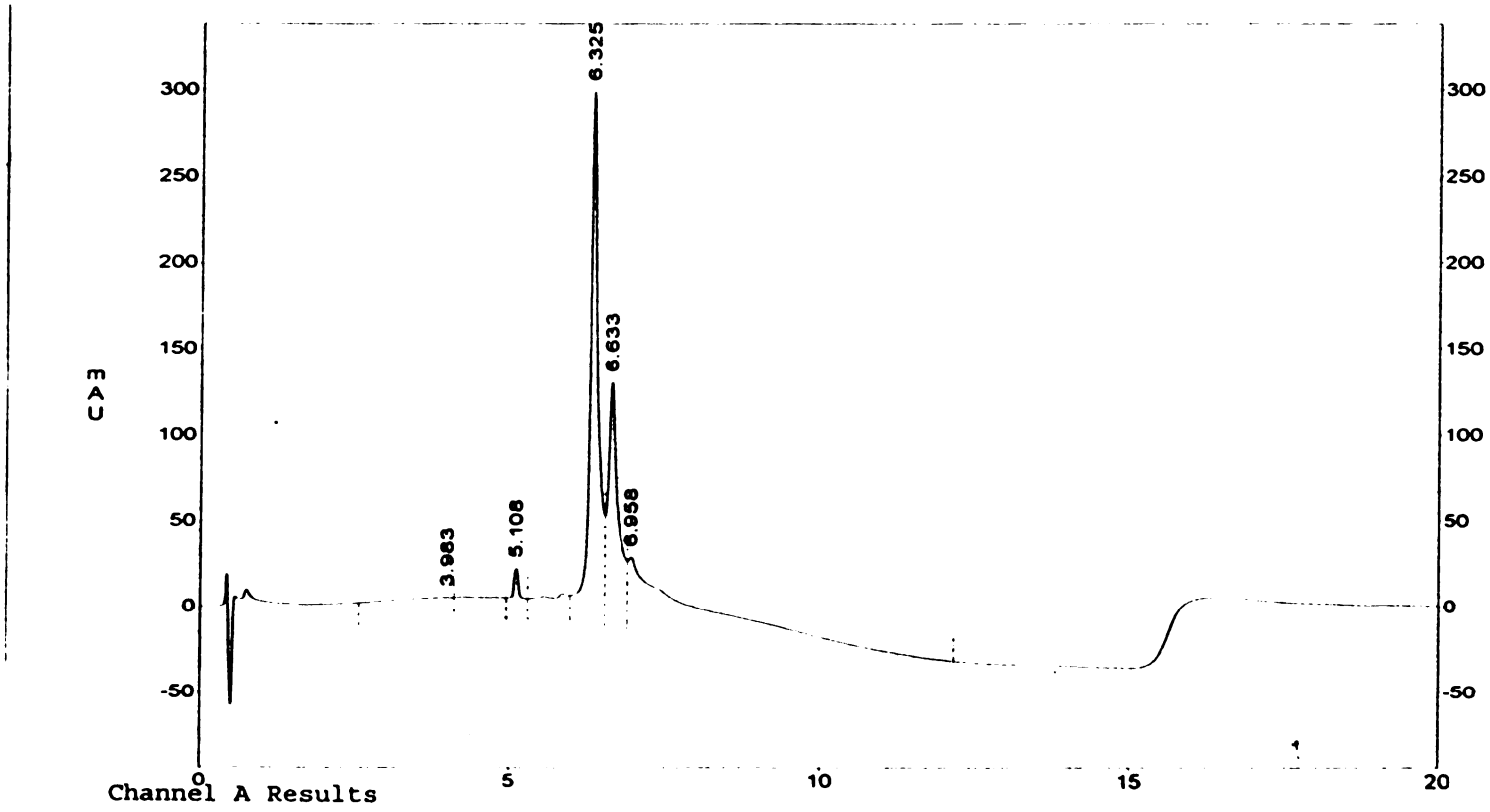
We attempted to couple labels on the molecule to allow a fast assay for affinity. A third of the resin was reserved for this purpose. A biotin moiety was added to the A sequences, after the coupling of a linker (2 residues of amino-hexanoic acid); a fluorescein molecule was added to the B sequences, also after coupling of a linker. We hoped to add the biotin-coupled molecule to streptavidin-coated 96-well plates, then add the fluorescein-labeled molecules to the wells, and check for binding by monitoring the fluorescence remaining after washing the wells. Unfortunately, the yields and purities of the molecules were not high enough to allow efficient coupling of the biotin or fluorescein moieties, leading us to abandon this approach.

Figure 5.4: Analysis of HPLC-purified A3

top, mass spectrometry results. The expected peaks are 2201 and 1467. Note the strong peak at 1448 which indicates contamination by molecules missing one of the loop's glycine residues (even after HPLC purification).

bottom, HPLC trace. The smaller, delayed peak likely represents the molecules that are missing a glycine residue.


PAA3S5 34 (1.215) Cm (30:37)
Scan ES+
1.28e3
100
%
0
505
506
524
560
613
685
801
1043
1405
1406
1424
1448
1467
1468
1501
1584
1806
1965
1990
2172
2201
2258
2286
2418
2499


UV3
300
250
200
150
100
50
0
-50
3.983
5.108
6.325
6.633
6.958
0
5
10
15
20
Channel A Results

Assays

Our first screen was based on size-exclusion chromatography (SEC). The molecules were all purified by HPLC, and combined in all possible A+B combinations (72 possible pairs in all). We ran all of the pairs on a size-exclusion chromatography column in phosphate-buffered saline (PBS), and compared the chromatogram of each pair with those of its component monomers. We expected that if any non-covalent aggregation occurred between the monomers, the chromatogram of the pair would contain a peak of higher hydrodynamic radius (lower retention time). The column was calibrated using protein standards of known molecular weights in PBS (Fig. 5.5A), to ensure that the column could at least separate protein species in the size range of interest. Fig. 5.5B shows the chromatograms of serial dilutions of one monomer, A9, with concentrations going from 1mM to 10μM. The major peak did not shift, down to the lowest detectable concentrations, indicating that this peak most likely represents a monomeric species (or that the oligomerization equilibrium has a Kd < 10 μM). Figs. 5.5C and 5.5D show results for two representative pairs, A7+B6 and A7+B5. In both cases, the major peak of the pair has the same retention volume as those of the monomers. Also, a shoulder is apparent at a smaller retention volume (higher hydrodynamic radius). However, the shoulder is present in the chromatogram of the monomers, and the relative size of the shoulder compared to that of the peak does not increase in the pair chromatogram, and does not decrease with dilution (Figs. 5.5C and 5.5D, insets). Hence this shoulder probably reflects either an impurity present in the solution of the monomers, or an alternatively folded conformation with a higher hydrodynamic radius. Fig. 5E shows all the pair chromatograms. In no case

Figure 5.5: Size-exclusion chromatography results

All injections were 10 µl, run at 0.5 ml/min in PBS.

A. Calibration with protein standards. The proteins were diluted in PBS to a concentration of 1 mg/ml. From left to right, the peaks are: albumin (MW≈60,000), ribonuclease A (MW≈13,000), insulin (MW≈7,000), insulin B chain (MW=3,000), angiotensin II (MW=1,720).

B. Serial dilutions of A9 monomer. A9 was run at concentrations (from largest to smallest peak) of 1mM, 0.5mM, 0.1mM, 0.05mM, and 0.01mM.

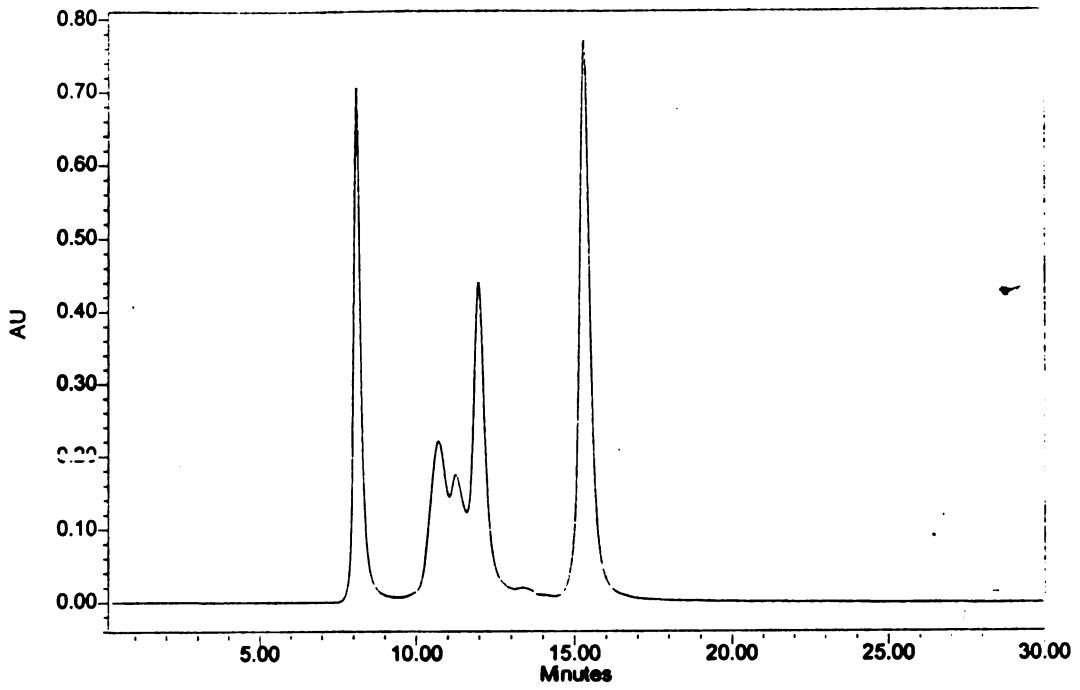
C. A7+B6 analysis. The traces, from largest to smallest peak, are: A7+B6 (~200µM each), A7+B6 diluted 1:2 (~100µM each), A7 (~200µM), and B6 (~200µM). *inset*, traces normalized by the size of their major peak. The largest shoulder belongs to the chromatogram of A7 alone.

D. A5+B6 analysis. The traces, from largest to smallest peak, are: A5+B6 (~200µM each), A5+B6 diluted 1:2 (~100µM each), B6 (~200µM), A5 (~200µM), and A5+B6 diluted 1:10 (~20µM each).

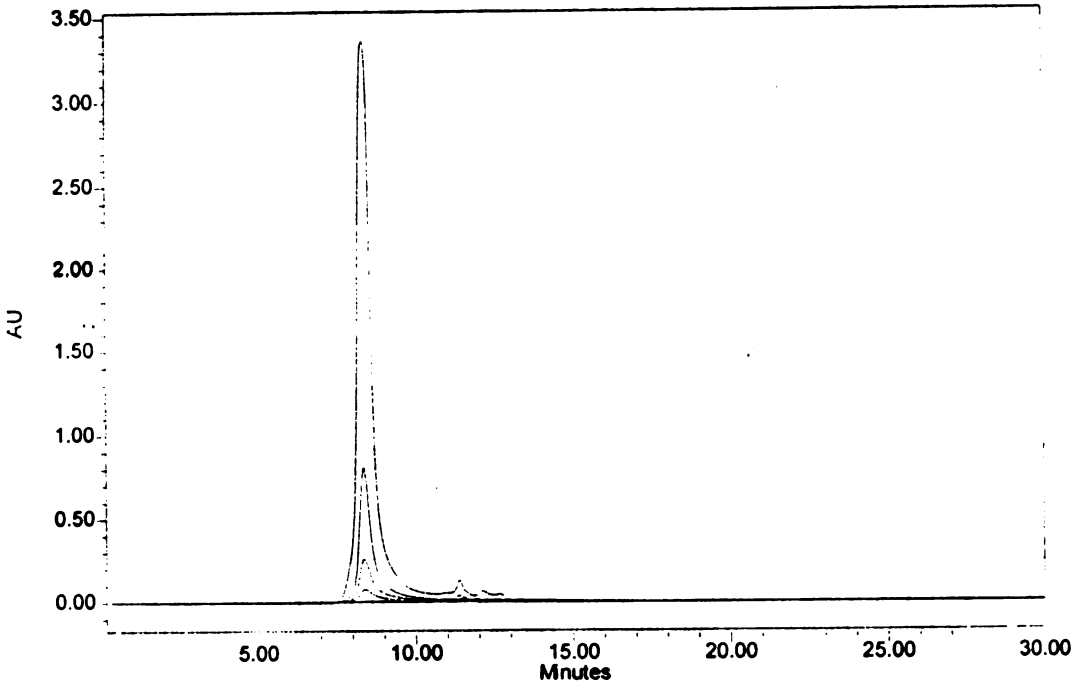
inset, traces normalized by the size of the major peak. The largest shoulder belongs to the chromatogram of A5 alone.

E. All pair chromatograms.

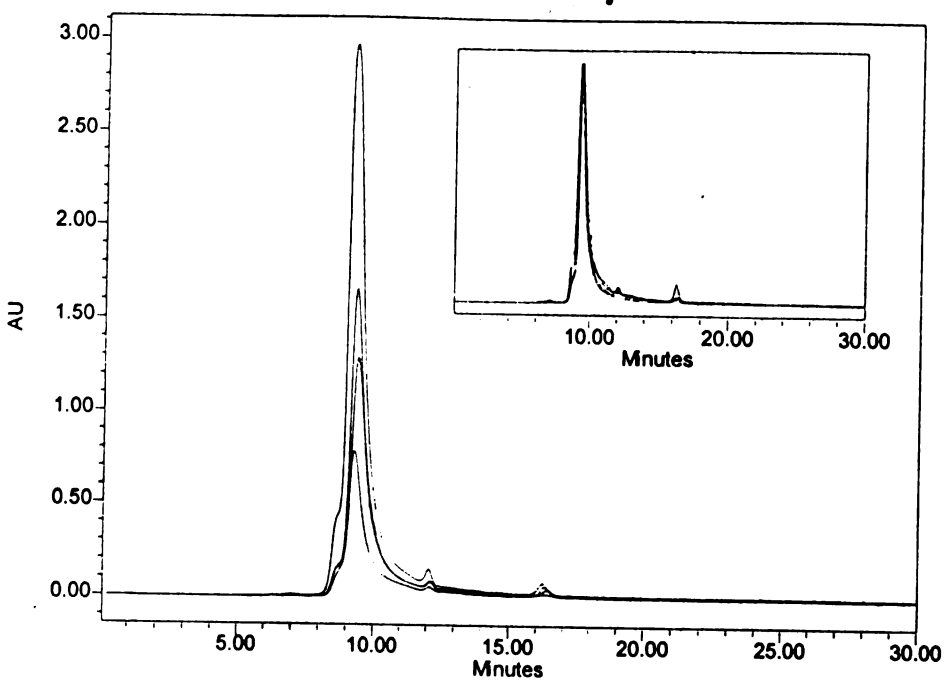
A. Column Calibration with Protein Standards



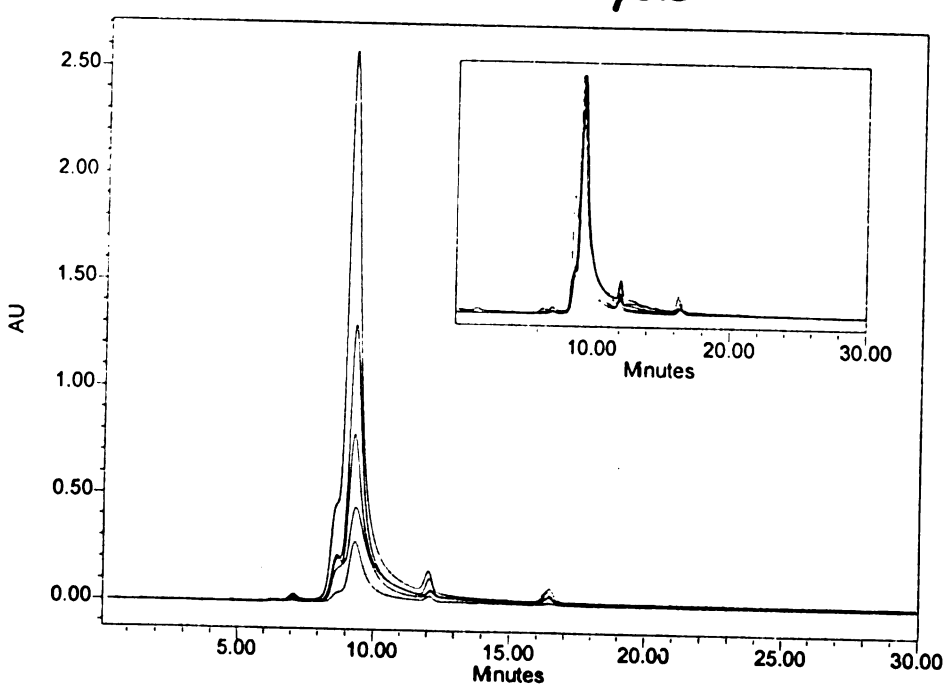
B. Serial Dilutions of A9 monomer



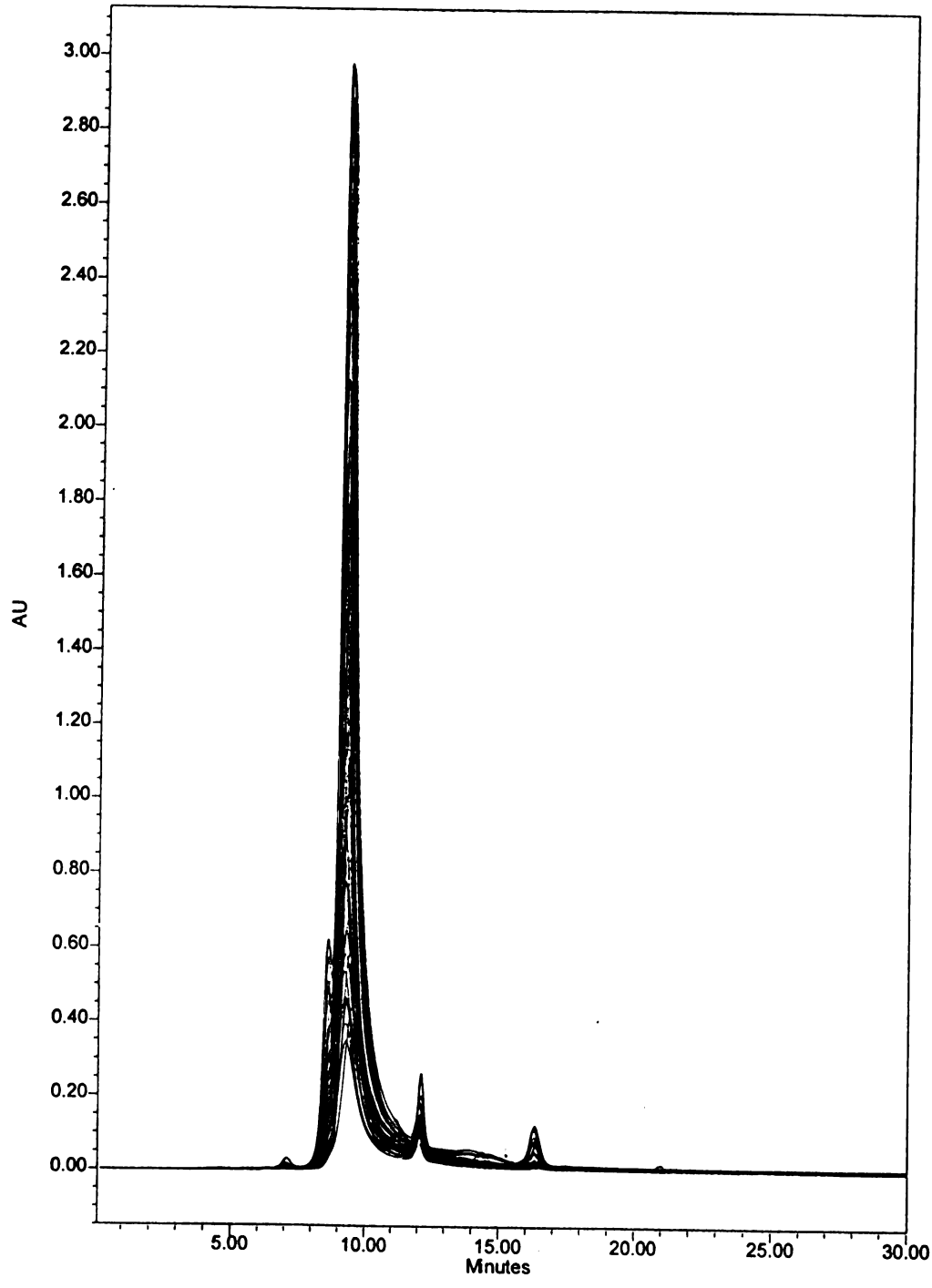
C. A7+B6 Analysis



D. A5+B6 Analysis



E. All Pair Chromatograms



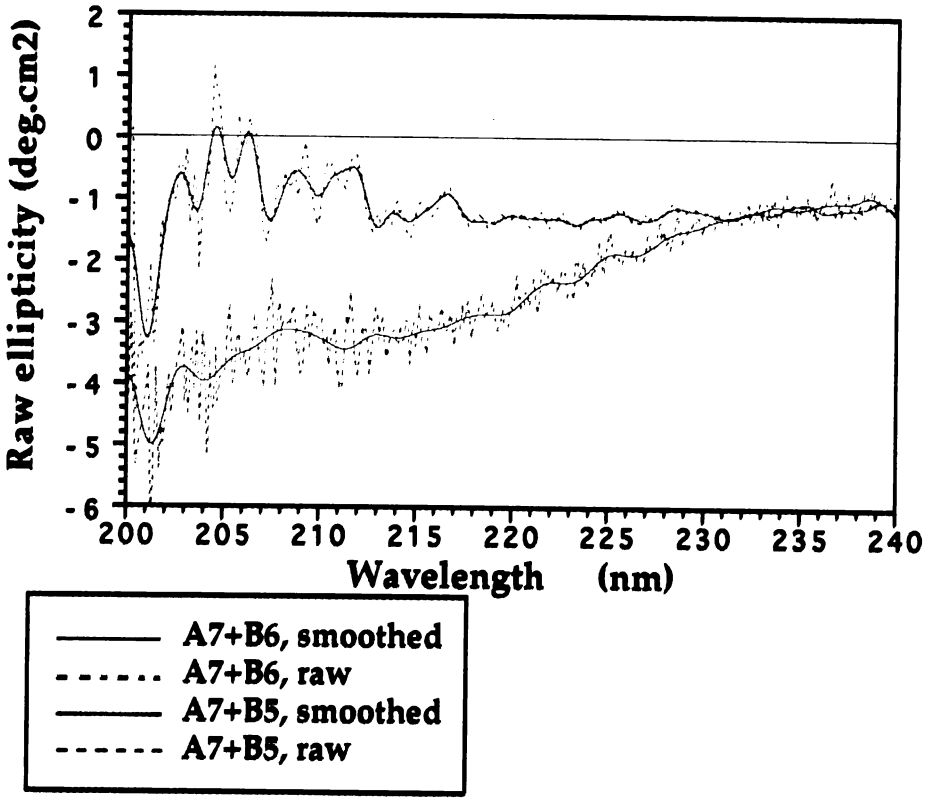
In conclusion, we were unable to show the presence of any bundle formation by this assay. We are therefore currently performing a CD-based assay, by obtaining CD spectra of all of the pairs and, in each case, subtracting from the spectrum of the pair the sum of the spectra of the pair's component monomers. Some of the preliminary results are shown in Fig. 5.6. In some, but not all of the cases, the subtraction spectrum shows evidence of structure induction, and possibly a double-minimum suggesting induction of helical structure (compare Fig. 5.6, dark blue trace, to Fig. 1.4).

Figure 5.6: Preliminary CD results

The black trace shows the subtraction CD spectrum of A7+B6: it was obtained by subtracting the sum of the CD spectrum of A7 (~10 μM) and that of B6 (~10 μM), from the CD spectrum of A7+B6 (~10 μM each). This trace was mathematically smoothed to give the dark blue trace.

The green trace shows the subtraction CD spectrum of A7+B5. This trace was mathematically smoothed to give the light blue trace.

All molecules were diluted in PBS.


Raw ellipticity (deg.cm2)
2
1
0
-1
-2
-3
-4
-5
-6
200
205
210
215
220
225
230
235
240
Wavelength (nm)
A7+B6, smoothed
A7+B6, raw
A7+B5, smoothed
A7+B5, raw

Discussion

In the hope of creating peptoid tertiary structure, we have developed a method to design a water-soluble peptoid four-helix bundle. This method relies on a computational design that uses sequence-structure minimization to generate hydrophobic, compact cores. The computational design proceeds along hierarchical lines similar to the experimental hierarchy used by DeGrado and co-workers (2) to create protein four-helix bundles. First, a suitable geometry is sought for the four helices, without any loop constraints. This step is meant to mimic the experimental search for non-covalent aggregates of helices, except that it allows the helices to differ in sequence. Once a geometry has been found, loops are added with various lengths and register, and the best loops are chosen based on their compatibility to the geometry found in the previous steps. Finally, the models are further refined, to further improve the packing and hydrophobicity of the cores with the added constraints imposed by the loops. This step is meant to mimic the experimental refinement of the cores to improve the bundles' stability (10,11).

It was our hope, grounded in previous results (Chapters 1 to 3) that peptoids would lend themselves to such a complex design. Indeed, peptoids containing *N*spe-based side chains with as few as five side chains seem to readily adopt helical conformations (Chapter 2). We thought that the designed helices would also form easily, and that no alternative conformation would strongly compete with the helical one; this would leave the helices ready to assemble to bury their hydrophobic faces. Moreover, we were encouraged by the success of molecular mechanics modeling in predicting the salient features of the conformation of a pentamer of *N*spe-like side chains (Chapters 1 and 3). Inferring from this success that sterics play an

important part in the conformational preferences of peptoids, we reasoned that molecular mechanics calculations, which effectively model steric interactions, would be a good guide in designing the association of the helices. There are precedents for successful core designs (30,32) using similar approaches. We have chosen a robust minimization method (simulated annealing), and have searched a wide space by allowing both the backbone dihedral angles and the helix orientations to change. This search space is larger than comparable, previously published computational designs (30,31,32), which keep the backbones fixed and work on single chains.

Our design approach was implemented and a first set of candidate sequences was generated (Table 5.1, lines 2 to 21). As expected, we were able to generate well-packed hydrophobic cores (Table 5.2 and Fig. 5.3). These sequences were synthesized and assayed by SEC. So far we have not found any evidence of bundle formation by this method. However, preliminary CD results are encouraging, suggesting that in some cases, mixing two monomers together results in the induction of (possibly helical) structure. The molecules with such properties are candidates for re-synthesis and further structural studies.

At this juncture, it is therefore important to re-evaluate our approach, to define what the major stumbling blocks could be, and how they could be circumvented in future design efforts.

The first set of possible problems comes from the computational strategy. We have not performed control calculations to check the validity of our parametric choices. There are experimental data set available that have been used as test cases for core packing algorithms (32,39), and which could

provide a test case of our evaluation function. Alternatively, parameters could be checked by comparing predicted core side chain dihedral torsion angles to their experimentally determined values.

Another possible weakness in the calculation is the lack of a search for alternative folds. Unfortunately, the search space of all possible conformations is far too large to adequately sample; moreover, we know little about alternatives to the helical conformation in peptoids.

The second set of problems are inherent to peptoids. The most obvious difficulty in designing a complex peptoid structure is the fact that some of the carbonyl oxygens must be buried unpaired in the core, which is energetically unfavorable. We hoped that the formation of a compact hydrophobic core would outweigh this desolvation cost. It is difficult to assign a numerical value to the energetic cost of carbonyl burial. Yang and Honig (40) quote a value of ~3 kcal/mol for this burial. By comparison, the burial of the main chain methylene group and an isopropyl side chain (the smallest hydrophobic side chain that we use) should provide an energetic benefit in the neighborhood of 4 kcal/mol (assuming a transfer energy of 0.9 to 1.4 kcal/mol/methylene group (40)), potentially outweighing the desolvation cost of the carbonyl. Larger side chains would provide an even greater energetic benefit upon burial. Thus it is not unreasonable to expect the association of the helices, while tempering this expectation by the knowledge that the numbers just quoted are controversial.

Another assumption that we have made is that the stretches designed to form helices would indeed form helices. This is based on our previous results (Chapters 1 and 2), but may be erroneous. Our NMR study of a pentamer of *N*spe-like residues indicated that at most 60% of the molecules adopted the right-handed fully helical conformation. The designed bundles,

in which only every third residue is *N*spe-like, may have an even smaller population of the helical conformation. In our design, there can only be one *N*spe-like residue per turn, since we do not have any polar *N*spe-like residues, and must therefore use *N*sce residues for solubility. To remedy this potential difficulty, we have been trying to develop new monomers (*(S)-N*-(1-(*p*-aminophenyl)ethyl)glycine, and *(S)-N*-(1-(*p*-carboxyphenyl)ethyl)glycine). These would have the chiral center and aromatic group apparently required to induce helix formation, but would also have a substituent on the aromatic group polar enough to make the molecule soluble in water. Hence two such residues could be used in each helical turn. So far, we have been unable to incorporate either of these two residues (the *p*-amino residue is not stable once incorporated, while the *p*-carboxyl residue is still synthetically inaccessible).

A final set of problems comes from our experimental approach. First, it proved difficult to synthesize the putative bundles in adequate purity and yield. Some of the monomers that we have chosen (particularly *N*snp) do not incorporate as efficiently as other monomers (such as *N*spe or *N*sce) which were used previously in the successful synthesis of long peptoids. Moreover, the couplings of the glycine residues at the end of 16-residue chains were not efficient either. These synthetic difficulties combined to decrease the yield and purity of the final products. This limited us in our assays, by forcing us to use low concentrations of peptoids (and therefore requiring a higher affinity between the bundle halves to be detectable). It also prevented us from coupling other reporter groups (*e.g.*, biotin or fluorescein groups) to the ends of our chains, which limited our possible choices of assay. In the future, loops could be synthesized using amino-hexanoic acid instead of glycine, or using a coupling agent stronger than HOBt.

Furthermore, our SEC assay may not have been sensitive enough to detect the formation of a bundle. While the column could distinguish between folded proteins of molecular weights 3700, 7000, and 13000, it is unclear whether it could distinguish between folded peptoids of MW 9000 and unfolded peptoids of MW 4500. We have synthesized a control molecule by dimerizing two monomers covalently; it appeared that the column could separate the dimer from the starting monomers, but we could not purify the dimer enough to get a definite answer (data not shown). Our CD assay is probably more sensitive than the SEC assay, but is unlikely to provide definitive evidence of four-helix bundle formation.

Finally, it should be noted that we have written the software for this work so that it would be applicable to other problems with only parametric changes. Possible uses include protein design, protein side chain conformational prediction, or ligand design (with protein, peptoid, or hybrid ligands).

Materials and Methods

Computational Design

We have written a C program, Oneiryx, that allows interactive display and manipulation of the model (sequence and structure), and can perform the minimizations. Calculations were done on SGI IndigoII workstations and on an SGI Origin200 machine; Figs. 5.2 and 5.3 were generated using the Insight® II 97.0 molecular modeling system (Biosym/MSI, San Diego, USA).

Oneiryx uses an internal angle representation for the models. Minimizations are performed with a simulated annealing protocol (41) using the Metropolis criterion (42). The annealing schedule relies on the standard deviation of the energy to determine the cooling increment (43). The number of steps at each temperature, as well as the δ cooling factor, are given as parameters (usual values are 5000 steps per temperature and δ=0.15). The simulation stops when the standard deviation of the energy falls below a threshold value at three successive temperature steps.

Steps can be either a point mutation in the sequence, a change in the orientation of one of the helices, or a change in one of the torsion angles of a helix. User-defined probabilities in a parameter file are used to determine the type of change at each step. For point mutations, all mutable residues are given in a parameter file, along with all the possible mutations at each position. In general, we have only left the internal residues mutable, and kept the external residues fixed. When a residue is changed, the side chain dihedrals are copied onto the new residues when possible. For helix geometries, there are six degrees of freedom (three translational, three rotational in reference to the helix' long axis). Boundaries are given in the parameter file. For torsion angles, movable torsions are specified in the

parameter file, along with the boundary values. We have restrained backbone and side chain dihedral angle values to ±10° and ±30° respectively around their equilibrium values (see Chapter 1 for those values).

The evaluation function consists of the following terms:

a) the non-bonded components (van der Waals + electrostatic) of the AMBER force field (36), with the charges determined by similarity to peptides, and a distance-dependent dielectric constant. A (user-defined) distance cutoff was used for this component, and the values are corrected for the cutoff values to remove the energy step associated with the cutoff. The non-bonded terms were calculated using a fast-lookup table (44).

b) a solvation potential, defined as the sum for each atom of the change in solvent-accessible surface area between the unfolded state and the current conformation, multiplied by an atomic solvation parameter (ASP). The unfolded state is modeled as an extended tri-peptoid centered on the residue of interest. ASPs were taken from the water-to-octanol free energy of transfer of model tri-peptides (45). This choice was based on the results in (46). Solvent-accessible surface areas were estimated using a 0.5Å cubic grid mimicking water occupancy sites. The area values obtained with the grid method had a > 0.95 correlation coefficient with the values obtained from a numeric accessible area calculation (data not shown).

c) for some of the simulations, a torsional potential was added (36), again with force constants estimated by similarity to peptides.

d) for some of the simulations, a compactness restraint was added as a set of quadratic distance restraints between the helices.

The relative weighing of the different terms is user-defined (except the torsional potential which come from the same force field as the non-bonded potential, and hence shares its weight). We have run the simulations with

different weights (with a ratio of solvation to non-bonded weight between 0.0001 and 0.005) to get a variety of candidate sequences. The maximal weight of the solvation term was the largest weight that would not result in steric clashes > 1000 kcal/mol.

To model the loops once a suitable geometry has been found for the helices (Fig. 5.1C), models are built with the four helices and intervening poly-glycine loops of variable lengths that start and end at various points along the helix (since the helices have a periodicity of three, there are three possible starting points and three possible ending points, hence nine possibilities for each helix). The molecules are then quickly minimized with only the backbone torsions movable in an energy field consisting of the non-bonded terms and a root-mean-square distance term that tends to superimpose the helices on the original loop-less model. The best loop is selected among the nine models by its favorable energy.

Once the simulations are finished, we perform a manual optimization of the sequence. Outside residues are modified to minimize the use of the costly *N*snp monomers, while adding solubility (through *N*sce and *N*ae side chains), and retaining at least one aromatic side chain per turn (since it seems that the aromatic, *N*spe-based side chains are required to induce helix formation). Electrostatic repulsion between external residues is also minimized. Internal residues are modified if their burial is inadequate (in which case we look for an alternative residue that could bury some hydrophobic surface while presenting polar atoms to solvent).

Synthesis

The peptoid addition cycles were performed on 50 mg of resin as described in Chapter 2. Glycine residues were coupled as Fmoc-protected

glycine (Bachem), added to the resin at 0.4M in NMP with 0.4M HOBt and 0.44M DIC. Each addition was done with >3 couplings of 1 hr each at 35°. The Fmoc group was removed in 20% piperidine in DMF for 30 min. at 35°.

HPLC and MS analysis was done as described in Chapter 2.

Size-exclusion chromatography

Peptoids and protein standards were run on a YMC 60A 300mm column (6mm diameter), using a Waters 2690 Separations Module instrument. The molecules were dissolved in phosphate-buffered saline, and run at 0.5ml/min. Absorbance was monitored at 214nm.

Circular Dichroism

CD was performed as described in Chapter 2 with peptoid concentrations between 5 and 20μM (depending on the amount of peptoid available) in PBS.

References

1. Betz S. F., Raleigh D. P., DeGrado W. F. (1993) *Cur. Opin. Struc.t. Bio.* **3**:601

2. Bryson J. W., Betz S. F., Lu H. S., Suich D. J., Zhou H. X., O'Neil K. T., DeGrado W. F. (1995) *Science* **270**:935

3. Betz S. F., Bryson J. W., DeGrado W. F. (1995) *Curr. Opin. Struct. Bio.* **5**:457

4. Eisenberg D., Wilcox W., Eshita S. M., Prycial P. M., Ho S. P., DeGrado W. F. (1986) *Proteins: Struct., Funct., and Gen.* **1**:16

5. Ho S. P., DeGrado W. F. (1987) *J. Am. Chem. Soc.* **109**:6751

6. Hill C. P., Anderson D. H., Wesson L., DeGrado W. F., Eisenberg D. (1990) *Science* **249**:543

7. Osterhout J. J. Jr, Handel T., Na G., Toumadje A., Long R. C., Connolly P. J., Hoch J. C., Curtis Johnson W. Jr, Live D., DeGrado W. F. (1992) *J. Am. Chem. Soc.* **114**:331

8. Ciesla D. J., Gilbert D. E., Feigon J. (1991) *J. Am. Chem. Soc.* **113**:3957

9. Regan L., DeGrado W. F. (1990) *Science* **241**:976

10. Raleigh D. P., DeGrado W. F. (1992) *J. Am. Chem. Soc.* **114**:10079

11. Raleigh D. P., Betz S. F., DeGrado W. F. (1995) *J. Am. Chem. Soc.* **117**:7558

12. Schafmeister C. E., Miercke L. J. W., Stroud R. M. (1993) *Science* **262**:734

13. Hecht M. H., Richardson J. S., Richardson D. C., Ogden R. C. (1990) *Science* **249**:884

14. Kuroda Y., Nakai T., Ohkubo T. (1994) *J. Mol. Biol.* **236**:862

15. Kuroda Y. (1995) *Prot. Eng.* **8**:97

16. Betz S. F., DeGrado W. F. (1996) *Biochemistry* **35**:6955

17. Mutter M., Tuchscherer G. G., Miller C., Altmann K.-H., Carey R. I., Wyss D. F., Labhardt A. M., Rivier J. E. (1992) *J. Am. Chem. Soc.* **114**:1463

18. Akerfeldt K. S., Kim R. M., Camac D., Groves J. T., Lear J. D., DeGrado W. F. (1992) *J. Am. Chem. Soc.* **114**:9656

19. Sasaki T., Kaiser E. T. (1989) *J. Am. Chem. Soc.* **111**:380

20. Handel T., DeGrado W. F. (1990) *J. Am. Chem. Soc.* **112**:6710

21. Regan L., Clarke N. D. (1990) *Biochemistry* **29**:10878

22. Lieberman M., Sasaki T. (1991) *J. Am. Chem. Soc.* **113**:1470

23. Ghadiri M. R., Soares C., Choi C. (1992) *J. Am. Chem. Soc.* **114**:825

24. Choma C. T., Lear J. D., Nelson M. J., Dutton P. P., Robertson D. E., DeGrado W. F. (1994) *J. Am. Chem. Soc.* **116**:856

25. Robertson D. E., Farid R. S., Moser C. C., Urbauer J. L., Mulholland S. E., Pidikiti R., Lear J. D., Wand A. J., DeGrado W. F., Dutton P. L. (1994) *Nature* **368**:425

26. Handel T. M., Williams S. A., DeGrado W. F. (1993) *Science* **261**:879

27. Kamtekar S., Schiffer J. M., Xiong H., Babik J. M., Hecht M. H. (1993) *Science* **262**:1680

28. Desjarlais J. R., Handel T. M. (1995) *Curr. Opin. Biotech.* **6**:460

29. Dahiyat B. I., Mayo S. L. (1997) *Science* **278**:82

30. Desjarlais J. R., Handel T. M. (1995) *Prot. Sci.* **4**:2006

31. Hellinga H. W., Richards F. M. (1994) *Proc. Natl. Acad. Sci. USA* **91**:5803

32. Hurley J. H., Baase W. A., Matthews B. W. (1992) *J. Mol. Biol.* **224**:1143

33. Hellinga H. W., Richards F. M. (1991) *J. Mol. Biol.* **222**:763

34. Hellinga H. W., Caradonna J. P., Richards F. M. (1991) *J. Mol. Biol.* **222**:787

35. Wilson C., Mace J. E., Agard D. A. (1991) *J. Mol. Biol.* **220**:495

36. Cornell W. D., Ciepak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179-5197

37. Eisenberg D., McLachlan A. D. (1986) *Nature* **319**:199

38. Predki P. F., Agrawal V., Brunger A. T., Regan L. (1996) *Nat. Struct. Biol.* **3**:54

39. Lim W. A., Sauer R. T. (1991) *J. Mol. Biol.* **219**:359

40. Honig B., Yang A.-S. (1995) *Adv. Prot. Chem.* **46**:27

41. Kirkpatrick S., Gelatt C. D. Jr, Vecchi M. (1983). *Science* **220**, 671-680

42. Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H. (1953) *J. Chem. Phys.* **21**:1087

43. Hunt N. G., Gregoret L. M., Cohen F. E. (1994) *J. Mol. Biol.* **241**:214

44. Hunt N. G., Cohen F. E. (1996) *J. Comp. Chem.* **17**:1857

45. Kim A. (1990) Ph. D., Dissertation, UCSF Dept. of Pharm. Chem.

46. Schiffer C. A., Caldwell J. W., Kollman P. A., Stroud R. M. (1993) *Mol. Sim.* **10**:121

Conclusion

Nα-chiral peptoids form a new class of synthetically accessible, sequence-specific heteropolymer with distinct structural properties. They can form stable helices in a variety of solvents, without the use of hydrogen bonds. Moreover, the general conformational features of these helices could be anticipated by computational modeling. We hope that these molecules will find broad applications in fields ranging throughout chemistry and biology.