Guidelines for designing peptoid structures: Insights from the Peptoid Data Bank

James R. B. Eastwood | Ethan I. Weisberg | Dana Katz | Ronald N. Zuckermann | Kent Kirshenbaum

Department of Chemistry, New York University, New York, New York, USA
Upper School, Packer Collegiate Institute, Brooklyn, New York, USA
Molecular Foundry, Lawrence Berkeley National Lab, Berkeley, California, USA
Correspondence
Kent Kirshenbaum, Department of Chemistry, New York University, New York, NY 10003, USA.
Email: kent@nyu.edu

Abstract
The number of structural studies of peptoids has grown dramatically over the past 20 years. To date, over 100 high-resolution structures have been reported for peptoids, which are typically defined as N-substituted glycine oligomers. We have collected these structures and standardized their sequence representations to facilitate structural analysis as the dataset continues to grow. These structures are presented online as The Peptoid Data Bank (databank.peptoids.org), which also provides persistent links to the published structural data. This review analyzes the present collection of structures and finds extensive support for grouping side chains by their chemistry at the position adjacent to the backbone nitrogen. Groups of side chains with similar chemistry at this position show similar influences on the conformational preferences of the backbone. We also observe a relationship between the side chain and backbone conformations for many monomers that has not previously attracted significant discussion: the values of the $\chi_1$ and $\phi$ dihedrals are correlated. We outline a general design strategy for attaining a specific backbone conformation based on the patterns seen in the collected structures.

KEYWORDS
foldamer, macrocycle, oligomer, peptidomimetic, secondary structure

INTRODUCTION

In 1992, Zuckermann et al. first reported the “submonomer” method for synthesis of N-substituted glycine oligomers, commonly known as peptoids[1] (Scheme 1). This advance led to the efficient generation of sequence-controlled peptoids incorporating an extraordinarily diverse set of side chain groups. The versatility of peptoids, along with their broad applications, has attracted researchers ranging from polymer scientists to pharmaceutical chemists to pursue their study and design. Families of other N-substituted amide oligomers have also been developed, including $\beta$-peptoids[2–4] (N-substituted $\beta$-alanine), N-substituted $\alpha$-alanines,[5] and other structurally related oligomers.[6–10] Although the promise of peptoids as conformationally ordered “foldamer” molecules was initially slow to develop, the first X-ray crystal structure of a peptoid macrocycle in 2008[11] subsequently led to an explosion of high-resolution structures. Peptoid chemists have now designed or discovered numerous stable backbone conformations of various sizes and topologies, some discovered serendipitously and some through careful selection of side chain sequences.

Concurrently over the past two decades, the field of de novo protein design has made enormous strides, and the corresponding design principles and protocols provide an inspiration to chemists developing folded biomimetic oligomers, known as foldamers. Much of the progress in the protein design field has been enabled by the collection and curation of high-resolution protein structures in the Protein Data Bank.[12] The PDB has illustrated sequence-structure-function relationships while also enabling revolutionary advances in structure prediction and de novo design methods through the statistical interpretation of vast quantities of structural information.[13–15]
Peptoid foldamer scientists seek a straightforward set of design rules for selecting a side chain sequence to stabilize a desired backbone fold. A central collection of high-resolution structures presented in a unified format would offer many benefits to the burgeoning peptoid field, and may even propel the development of suitable design guidelines.

In pursuit of these goals, we have assembled the published peptoid structures into a database with standardized metadata describing them, including their chemical structures. We have made this collection available as the Peptoid Data Bank (databank.peptoids.org). Analysis of the structures in this Data Bank yields insights that establish design principles, which have been understood, implicitly or explicitly, by some in the community but have not been fully articulated in the literature. These principles are notably distinct from the strategies used to design \textit{de novo} proteins. To date, peptoid designs rely less on non-local interactions or on stable secondary structure modules than protein designs. Instead, peptoid designs depend primarily upon the conformational preferences of individual monomers due to local side chain–backbone interactions. The comprehensive analysis of side chain–backbone relationships presented here will empower future peptoid chemists to produce a new generation of peptoid designs.

This review is intended to provide even novice peptoid chemists with a detailed understanding of design strategies used by some of our colleagues (and ourselves) who have been working in this field for over a decade. Recent reviews of peptoids have evaluated molecular structures\textsuperscript{16,17} and side chain influences on amide bond geometries.\textsuperscript{18} This work will provide an up-to-date look at some of these same topics, but is intended to be a prospective, rather than retrospective, look at the field of peptoid structure. It is focused on what we have found to be the simplest, most important chemical features governing sequence-structure relationships, and supports these observations with analysis of the Peptoid Data Bank. We also offer this analysis as a demonstration of the utility of the Peptoid Data Bank and a tutorial on its use.

2 | BACKGROUND

2.1 | A brief history of peptoid structural studies

Structural analysis of peptoids began with the simplest of peptoids: oligomers composed of sarcosine (N-methylglycine) units. This was a key focus of the research in the lab of Johannes Dale at the University of Oslo, starting in the late 1960s. Dale was drawn to these molecules as a facet of his interest in the subtleties of molecular conformation, perhaps an esoteric topic for some of his colleagues. Dale’s 1978 monograph \textit{Sterechemie und Konformationssanalyse}\textsuperscript{19} featured cyclic tetrasarcosine in the final chapter, on large ring systems, because it was a particularly fascinating and complex example of the features he was investigating. In a contemporary review of the monograph, Peter Karlson wrote, “The title of this book is misleading. Most of the text is really conformational analysis, and very little is stereochemistry. . . . One has the impression that the author has written up his hobby. This is generally not the best starting point for a book.”\textsuperscript{20} Yet Dale’s investigations served as a starting point for the current field of peptoid structural design, through the study of conformational preferences of oligosarcosine macrocycles.

Though the first peptoid heterooligomers were reported in 1992, relevant structural studies on sarcosine homooligomers had been carried out as early as 1969. The earliest such studies were the work of Kirsten Titlestad, who undertook studies of “Cyclic oligopeptides of sarcosine (N-methyl glycine)” under the direction of Johannes Dale.\textsuperscript{21} In collaboration with crystallographer Per Groth, they obtained X-ray crystal structures of every oligosarcosine macrocycle size from four to ten residues (Figure 1), with the exception of the nonamer, in conjunction with 1D $^1$H-NMR studies.\textsuperscript{22–28} In a pre-scientific example of the value of these sorts of collected structural studies, they reported the conformational landscape of sarcosine in Ramachandran $(\phi - \psi)$ space based on their crystal structures and came to conclusions that would be reinforced nearly half a century later:\textsuperscript{30} that these molecules prefer two symmetrical values for $\phi$ ($\pm 150^\circ$) and only a single value for $\psi$ (around $180^\circ$) except in the small macrocycles.\textsuperscript{31}

After this initial set of oligosarcosine structures, a few researchers continued to study oligomer sequences incorporating N-methyl or N-benzyl substituted glycines as proline analogs, and reporting X-ray structures of a few macrocycle crystals during the 1980s\textsuperscript{32,33} and into the modern peptoid oligomer era.\textsuperscript{34} This modern era began in 1992, with Zuckermann’s aforementioned solid phase submonomer synthesis strategy,\textsuperscript{35} but the blossoming of the peptoid structure field was nurtured by the concept of “foldamer” chemistry, which gained currency throughout that decade.\textsuperscript{33} This concept describes a design pattern in which an iterative synthesis protocol creates a sequence-specific heteropolymer that can then fold into three-dimensional
shapes of dramatically higher complexity. An early version of the foldamer concept was laid out by Ken Dill and co-workers based on a review of simple exact models of protein folding, where they proposed that “The encoding of a unique compact chain conformation may not require amino acids; it may require only the ability to synthesize specific monomer sequences.” [36]

The first experimental evidence that linear peptoids could fold into stable, well-defined structures came from Circular Dichroism (CD) spectroscopy of a homooligomer of S-N-(1-phenylethyl)glycine. [37] This experiment was followed by NMR determination of the solution structure of a closely related compound. [38] The observed dihedral angles corresponded closely to peptide Poly-Proline Type I (PPI) helices. A similar X-ray crystal structure (Figure 2) was obtained shortly thereafter [39] but not published for several years. [40] Subsequent studies have established this PPI-like helix as an important secondary structure archetype for peptoids.

The Nrch₅ helical pentamer remained the only oligomer crystallized by modern peptoid chemists until Yoo et al. reported head-to-tail cyclization of peptoids of various sizes (again highlighting the prescience of Titlestad and Dale), crystallizing both a hexamer and an octamer. [11] From those initial cyclic structures, the data depositions began to proliferate, engaging new researchers and representing more diverse topologies and sequences.

2.2 Spectroscopic studies at monomer level

Throughout the modern era of peptoid research, scientists studying homooligomers and “dipeptoids” (capped monomers) have used CD and NMR spectroscopy to study conformational preferences of individual residue types, often seeking to correlate experimental observables with simulations using Quantum Mechanical (QM) or Molecular Mechanical (MM) methods. For peptoid monomers with suitable capping groups, ratios of cis- to trans-amide bond conformers can be reliably determined via ¹H-NMR, [41] and helical propensity can be determined by CD spectroscopy of suitably-sized homooligomers (typically pentamers). [37] As we explain below, the conformational preferences revealed by these studies can be generalized to the same residues in diverse sequences, enabling highly localized sequence-structure design strategies.

2.2.1 Solid state X-ray crystal structures

By analogy to the study of proteins, X-ray crystal structures of peptoids are considered to provide definitive experimental evidence of folding properties. However, the peptoids crystallized thus far occupy a much different regime than globular proteins: because of their small size, their solid state is dominated by intermolecular, rather than
intramolecular contacts. Many of the peptoid crystal structures solved are composed of even numbers of residues with a repeating sequence. Upon cyclization, these molecules become compatible with multiple symmetry point groups. Structures of odd-numbered macrocycles, non-repeating sequences, and linear oligomers are still relatively rare. These observations suggest the possibility that the observed molecular conformations are stabilized by crystal-packing forces, and may not dominate in solution. Nevertheless, peptoid crystal structures have proved to be a valuable data set for understanding conformational preferences, and QM and MM studies have been correlated with these results.

2.2.2 | Solution phase NMR structures

Because of the limitations of solid state structures in elucidating solution conformational properties, solution phase NMR structures have been highly desired by researchers, and some of the earliest high resolution structures were solved by NMR techniques. However, several properties of the peptoid backbone make NMR data challenging to interpret. First, the lack of an amide NH group frustrates assignment strategies familiar to peptide chemists. Second, isomerization of the amide bond is slow on the NMR time scale, leading to multiple sets of peaks. For these reasons, X-ray structures currently outnumber NMR structures ten to one.

The utility of the submonomer method in producing diverse peptoid sequences lends itself to a distinct NMR experimental design: several key positions in the sequence can be designed to include uniquely substituted aromatic rings. The electronic properties of the substituent alter the chemical shift of nearby nuclei, facilitating assignment of peaks that would otherwise overlap. Of nine NMR structures solved thus far, five are designed following this strategy (Figure 3). The largest linear structure solved to date is a decamer designed according to this scheme, which adopts a “ribbon” secondary structure of alternating cis- and trans-amide bonds.

2.3 | Unique features of peptoid backbones

Many peptoid chemists study these particular molecules because they offer immense promise as foldamers. Realization of this promise will require a sophisticated understanding of the influence of side chain properties on backbone conformation, also described as a sequence-structure relationship. The structural studies outlined above provide ample experimental evidence from which to develop this understanding. The structural properties of peptoids are distinct from peptides in several ways that merit careful attention.

Analogy to peptide design suggests that a Ramachandran-like analysis of the allowed regions of dihedral space would inform design efforts. Prior review of the available high-resolution peptoid structures showed that peptoids essentially prefer only four distinct
backbone conformations.\textsuperscript{30} The dihedral angles that distinguish these conformations are not \( \phi \) and \( \psi \), as in peptides, but \( \omega \) and \( \psi \) (Scheme 1; detailed comparisons of the folding landscape of peptoids to peptides have been previously reported\textsuperscript{30,43}). As observed in sarcosine macrocycles, \( \psi \) falls into only one conformational cluster (close to 180°) in almost all peptoid structures. The allowed regions of \( \omega \) versus \( \psi \) space are well-separated, with high energy barriers between them.

Several types of side chains have been shown to exert a strong influence over the conformational landscape of the backbone at a local level, in a manner distinct from peptides. The first structure-inducing side chain types identified were chiral N-1-phenylethyl substituents, which were shown to induce a helical conformation in the backbone, with the handedness of the helix dependent on the chirality of the side chain.\textsuperscript{57} Further work has determined this helix-induction effect to be the sum of two distinct properties: branching at the C\( \text{\textsubscript{ω}} \) position promotes a cis-amide bond at the junction with the preceding residue (controlling \( \omega \)), and chirality at that same position selects one helical sense over the other (controlling \( \psi \)). Similar control over \( \omega \) can be achieved electrostatically by some positively charged (or protonatable) side chains.\textsuperscript{49,50}

Symmetrical branched side chains (e.g., tert-butyl) still exert control over \( \omega \) but cannot enforce a helical sense in the absence of chiral side chains at nearby positions.\textsuperscript{51} Other side chain types, including N-aryl,\textsuperscript{52} N-hydroxy,\textsuperscript{53} N-alkoxy,\textsuperscript{54} and N-imino (or N-alkylamino)\textsuperscript{55} side chains, have been shown to strongly favor a trans-amide bond. For a thorough review of strategies to control isomerization of the amide bond, see Kalita et al.\textsuperscript{18} These general chemical properties that allow a side chain to restrict its backbone from four to two or even one conformation are an advantage of peptoids and lend themselves to the distinct design strategies we will describe below.

Both the backbone and side chains of peptoids adopt discrete rotameric states that can be overlaid on theoretical energy landscapes from various levels of theory.\textsuperscript{30,43,56} Side chain rotamer preferences of proteinogenic amino acids have been a critical component in protein structure prediction and design.\textsuperscript{57–59} Approximating peptoids with similar models will facilitate prediction and design of not only backbone conformations but complete, three-dimensional arrangements of all their chemical features.\textsuperscript{156}

2.4 | Challenges to structural analysis

Without a large collection of high quality experimental structures, researchers have had limited ability to understand the full range of peptoid conformations and dynamics. Efforts to analyze the conformations of all peptoid structures have faced several challenges. Identifying which chemical species are of interest is the first challenge. Some researchers have studied structures of any peptide oligomers with at least one N-substituted glycine,\textsuperscript{60} while others have restricted analysis to oligomers containing only tertiary amides.\textsuperscript{30,43}

The next challenge is locating the relevant data files. Nearly all X-ray structures reported by modern peptoid researchers have been deposited in the Cambridge Structural Database, but we have identified exceptions. Some legacy structures are available in the CSD but are not easily discoverable due to incomplete metadata. NMR structures are not deposited in any database, and are only accessible online as Supporting Information from the publisher of the associated research report, where they are not readily searchable. A few of these structures were never published as SI, and have been available only from the original author upon request.

Even with a complete collection of data in hand, the peptoid analyst still faces several hurdles. They must still make some (occasionally subjective) determinations about residue definitions, as many published structures contain capping groups or other subunits that do not neatly conform to the N-substituted glycine pattern. Many structural studies must begin with identifying and aligning common substructures (e.g., for consistent designation of bond dihedrals). Without a consistent residue pattern defined, this process is difficult to automate, and performing it by hand is tedious, error-prone, and not easily verified. Reporting structures by their IUPAC names offers little immediate insight into the underlying residue sequence. General-purpose cheminformatics databases, formats, and tools are poorly suited to representing and processing these structures. To address these limitations in the data depositions, to encourage more sophisticated cheminformatic analyses of peptoids, and to lay a stable foundation for future growth of the field, we have curated metadata describing high resolution structures of molecules of interest to understanding peptoid conformational preferences, and present it as the Peptoid Data Bank.

3 | THE PEPTOID DATA BANK

The Peptoid Data Bank is an online resource providing access to peptoid structural data along with a consistent representation of peptoids as a sequence of discrete residues. The data represented in the Peptoid Data Bank include all high resolution solved structures of peptoids, which we defined for purposes of curation as oligomers consisting of at least two N-substituted amino acids not bearing side chains on the backbone carbon atoms. This definition excludes the many structures of N-methylated peptides, and also many oligoproline structures, as these offer little insight into the conformational preferences of the oligomers potentially accessible via the submonomer synthesis method.

3.1 | Contents of the Peptoid Data Bank

For each structure included, we have collected and curated metadata that describe the publication history of the structure, report the structural properties needed to interpret it as a sequence of residues, and provide stable access to the data itself. We have also assigned a unique Identifier Code to each structure by joining the two-digit publication year to a two-letter index of peptoid publications for the year, followed by a one-digit index of structures included in a publication.
This code is reported in the structure's name in the Data Bank along with two descriptive fields – an integer indicating the number of residues in the structure, and a character indicating the topology (as of this publication, represented topologies include acyclic “A”, cyclic “C”, and multicyclic “M”). As an example, a peptoid octamer macrocycle is named “07AA2-8-C.” The first field of the name is the unique identifier code, in this case assigned because this is the second structure included in the first paper which published peptoid structures in 2007.\cite{11}

Analysis of this metadata provides an overview of progress in the peptoid field (Figure 4).
The curated collection begins with the aforementioned X-ray crystal structures of cyclic sarcosine oligomers. A small number of cyclic structures deposited in the CSD during the subsequent decade have been included as well, all of which include at least one Proline residue. They are included both for historical interest and for the structural relevance of fully-N-substituted macrocycles to modern peptoid chemistry.

Two peptoid structures are curated here which were reported in early publications but for which coordinate files have not been readily available: A nonamer solved by NMR and a hexamer with phenyl side chains solved by X-ray crystallography. These files have been recovered from members of the community and are presented as Supporting Information with this publication (06AA1-9-A.pdb, 008AC-6-C.cif).

We have included structures of β-peptoids and of oligomers of mixed α- and β-peptoids. We have included the small number of structures of macrocycles with mixed peptoid-peptide backbone, as these hybrid macrocycles show great promise in biomedical applications.

3.2 Technical implementation of the Peptoid Data Bank

We have chosen the OpenSMILES string format to represent the chemical structure of the residues. As with any linearization of a graph, this representation can take many valid forms; we have chosen a consistent residue SMILES format that makes string manipulation of residues to generate polymer SMILES as facile as possible. Each residue is written with the backbone nitrogen first, and the backbone carbonyl last (with the carbonyl oxygen written as a branch). The side chain is written as a branch off of the backbone nitrogen.

We built three interfaces to the database of peptoid structures: a website and two different Application Programming Interfaces. The web interface offers a visual presentation of peptoid structures, and allows the user to search by any of the metadata fields, such as experimental method or topology, in addition to linking structures to their authors and the residues included. The RESTful API mimics the structure of the website, but returns raw formatted data instead of HTML web pages. The GraphQL API enables users to access the database as a graph model consisting of nodes and edges, which permits selective access to metadata/properties of interest. Example scripts demonstrating access to these interfaces are available as Supporting Information. For more information, the source code can be found on GitHub: https://github.com/Kirshenbaum-lab/peptoid-db-website

The data curated in the Peptoid Data Bank have not been standardized. The X-ray crystal structures, deposited in the Cambridge Structural Database, follow the standards of that repository. The NMR structures curated are all available in a PDB format readable by standard structure viewing software, but residue names, atom names, and some other features such as TER records are inconsistent between different research groups.

3.3 The future of the Peptoid Data Bank and its value

The database structure we have used to build the website is intended to be robust to future developments in the peptoid field. The structure of the data recorded has been carefully considered to permit representation of many plausible features that have not yet been included in peptoid structures, such as complex topologies (e.g., lariats, knots, catenanes) or alternative backbone chemistries (e.g., gamma peptoids). The specification of residue sequence in each structure will permit the use of bioinformatic and text- and language-processing algorithms to analyze and manipulate peptoid sequences, while the SMILES representation corresponding to each residue will allow translation of sequences into molecular representations for use with conventional cheminformatic software.

By enabling the correlation of sequence representation with structural information across a large number of structures, the Peptoid Data Bank will enable new approaches to designing peptoids. Peptoids are already being used to antagonize protein-protein interactions by targeting protein surfaces, improving these designs will require a detailed understanding of their conformational preferences. Our vision is that access to the Peptoid Data Bank will inspire peptoid chemists to design new structures, which will be added to the Peptoid Data Bank to guide further structure designs in a positive feedback loop.

4 STRUCTURAL ANALYSIS OF THE CONTENTS OF THE PEPTOID DATA BANK

The conformational preferences of peptoids can be revealed by analysis of the structures in the Peptoid Data Bank. Peptoid chemists have studied these conformational preferences by examining two-dimensional landscapes of pairs of dihedral angles, analogous to the Ramachandran maps of peptides. This analysis can be extended to all possible pairs of the five dihedral angles common to most α-peptoids (Scheme 1). The β-peptoids and other structurally related oligomers are not yet represented well enough in the Data Bank to permit this level of analysis. The results of this extended analysis are shown in Figure 5, and illustrate a few key points about the conformational preferences of α-peptoids.

The individual dihedral angle histograms (show along the diagonal in Figure 5) reveal each of these dihedral angles to have clear preferred values. The asymmetry in the histograms of φ and χ1, indicating more structures displaying negative values of these dihedrals than positive, reveals the historical influence of the early studies of (S)-N-(1-phenylethyl)glycine and derivatives. By reporting the behavior of one arbitrary enantiomer, these early studies established a bias in the field for side chains with (S) stereochemistry. This side chain stereochemistry influences the conformation of not only the side chain but also the backbone, in a manner evident by studying pairwise plots of the relevant dihedral angles.
The pairwise plots of each of the five dihedral angles reveal interesting conformational features across a few dimensions of the landscape. As described previously, the two dominant mirror-image clusters in $\phi - \psi$ space contain nearly all data points. Consequently, $\phi - \omega$ space contains most of the structural information describing the peptoid backbone. In $\phi - \chi_1$ space, four clusters are well-defined, but two of these are much more heavily populated. This effect is clearly evident when contour lines are added to the plot based on a kernel density estimate, as in the lower left panels of Figure 5. The causes and consequences of this effect, along with its relationship to side chain stereochemistry, will be explored below. Finally, $\chi_2$ exhibits some intriguing clustering and correlations with both $\chi_1$ and $\psi$. Since the preferences of $\chi_2$ are dependent on the chemical features of the side chain, these effects will best be studied on separate classes of side chain and are outside the scope of this review.
4.1 Secondary structures in the Peptoid Data Bank

Much of the research into peptoid structures has been focused on identifying stable backbone conformations, along with rules for inducing them. Here, we interpret “secondary structures” to be defined by regular patterns of backbone bond dihedrals. In peptides, the two dominant secondary structures (α-helices and β-sheets) are stabilized by backbone hydrogen bonds, and the forces that lead to the formation of one of these secondary structures over another are subtle and often depend on long-range interactions (as measured by number of residues). In peptoids, no such dominant secondary structures have been identified that are strongly stabilized only by repetitive long-range backbone interactions. Virtually all peptoid structures rely on side chain – backbone interactions, interactions with metal ions, or discrete long-range constraints. The known peptoid secondary structures are stabilized by an interplay of interactions that can be grouped into three categories:

- Weak backbone interactions
- Long-range constraints (covalent or non-covalent)
- Side chain – backbone interactions

This last category is likely to dominate in linear peptoids.

A typical example of a secondary structure stabilized primarily by the first category are the Poly-Proline II-like peptoid helices, which align successive carbonyl dipoles to form energetically favorable interactions. This interaction has been studied as an electron delocalization effect, n → π* orbital overlap, but can also be modeled (if not properly quantified) as a classical electrostatic attraction. Weak hydrogen bonds between methylene hydrogen and carbonyl oxygen can also be invoked to explain some of the same conformations. Though these stabilizing interactions are neither as strong nor as long-range as the hydrogen bonds that stabilize the peptide α-helix, the result is a cooperative effect directing the folding of a peptoid helix. Still, the formation of a peptoid helix depends on additional stabilization from stronger sources, such as local side chain – backbone interactions.

Longer-range interactions have been shown capable of stabilizing alternative peptoid structures. Early studies of peptoid helices discovered that nonamers formed an alternative non-helical structure. NMR experiments showed this structure to be a “threaded loop,” stabilized by hydrogen bonding between the C-terminal unsubstituted amide and carbonyl oxygens near the N-terminus.

Long range covalent constraints can stabilize a peptoid conformation, with backbone-cyclized peptoids being the typical example. Many of these cyclic peptoids readily form stable patterns of amide bond isomers due to the constraint imposed by cyclization. In solution, these molecules (absent other structure-inducing elements) still populate a conformational ensemble, but their conformational space is restricted enough that they are readily crystallized, as evidenced by the surfeit of cyclic structures in the Peptoid Data Bank. Pierri et al. have recently published several 12-residue peptoid macrocycles (the longest single-chain peptoids crystallized to date, discussed in more detail below) that crystallized in the same conformation, demonstrating the power of this simple covalent constraint to enforce conformational order in a large oligomer.

Side chain covalent constraints have also been used to restrict peptoid conformational space, though presently the only such structures that have been characterized at high resolution are highly constrained multi-cyclic molecules.

Lastly, careful exploitation of local side chain – backbone interactions has generated several secondary structures that appear to have no stabilizing force within the backbone, but are stabilized entirely by local side chain – backbone interactions. The ribbon secondary structure, with its alternating cis- and trans-amide bonds, is the typical example of this category. The selection of side chains to induce such a pattern is the subject of the next section.

4.2 Side chains in the Peptoid Data Bank

In the absence of both chirality and hydrogen bonding within the backbone, peptoid structures rely heavily on the influence of side chains to induce structure. Several classes of peptoid side chain have been identified which impose strong constraints on the backbone conformation of the residue. The two important degrees of freedom in the peptoid backbone, ω and φ, flank the nitrogen, at which the side chain is attached, so steric interactions between side chain and backbone can restrict both of these degrees of freedom. The nature of these steric interactions depends almost entirely on the chemistry of the first atom in the side chain (denoted C₉N). Polarg groups distal from the backbone can have further influences on the backbone, but too few examples have been characterized to draw broad conclusions about these effects. Enough is known about the effect of the C₉N position that we can group the known peptoid side chains according to this feature, and identify four side chain groups with distinct properties:

1. Unbranched N-alkyl
2. Branched N-alkyl
3. N-aryl
4. N-X (N-alkoxy, N-imino, N-alkylamino)

The relationship between these groups of side chains and the backbone are apparent in Figure 6.

4.2.1 Unbranched N-Alkyl

Alkyl side chains without a branch at the C₉N position are in many ways the easiest side chains to consider as design elements. They are often synthetically tractable, and can be rapidly installed in high yields using the corresponding primary amines. They do not impose significant additional steric constraints on the backbone (compared to sarcosine), though some charged side chains have been shown to stabilize either a cis- or a trans-amide bond, mostly by electrostatic interactions or CH-O hydrogen bonds with the carbonyl oxygen. With the
caveat that polar functional groups may introduce such additional interactions, unbranched N-alkyl side chains can be considered “flexible” residues.

4.2.2 | Branched N-Alkyl

Alkyl side chains that branch at the C\textsubscript{\textalpha N} position modify the interactions between the side chain and backbone of residue i-1, resulting in a preference for cis-amide bonds by steric repulsion\cite{71}. Bulkier side chains confer a stronger amide bond preference, with N-(tert-butyl)glycine a notable example with an overwhelming preference for the cis-amide\cite{72}.

4.2.3 | Chiral N-alkyl

When the substitution pattern at the C\textsubscript{\textalpha N} position is chiral (i.e., C\textsubscript{\textalpha N} is a stereocenter), the chirality breaks the symmetry of the side chain \chi angle landscape, favoring a single value for this dihedral. In an effect described in more detail below, \phi typically adopts the opposite sign of \chi, so that this class of side chain controls both degrees of freedom of the peptoid residue.

4.2.4 | N-Aryl

An aromatic ring attached to the backbone nitrogen imparts a strong preference for a trans-amide bond through a complex stereoelectronic effect\cite{52}. NMR studies of these residues measure a $K_{\text{cis/trans}}$ between 1/10 and 1/20. N-aryl side chains with substitutions in the ortho position of the aromatic ring have been shown to form atropisomers due to the high energy barrier to rotation about \chi\cite{70,71}.

4.2.5 | Exotic functional groups

Several peptoid monomers have been characterized that contain a side chain appended to the backbone via some bond other than N-C. For example, N-OH and N-O-R glycine residues have been characterized as components of larger structures\cite{53,54} and are discussed here. These “exotic” side chains have subtle effects on peptoid conformations: it is not clear how rotameric the side chain conformations might be, and the backbone nitrogen can become substantially pyramidalized\cite{54}. Though they are not yet included in the Peptoid Data Bank, N-imino and N-alkylamino glycine residues have been studied as monomers, where they adopt a backbone conformation rarely seen in unconstrained peptoids\cite{55}.

4.3 | Local side chain - backbone relationships

Peptoid structural studies have made clear that the chemical features of the side chain can have a strong influence on the behavior of the \omega dihedral angle. Comprehensive analysis of the structures in the Peptoid Data Bank agrees with prior knowledge accumulated in the field: \alpha-branched side chains induce a strong preference for cis-amide bonds, and N-aryl side chains induce an overwhelming preference for trans-amide bonds. The cis/trans ratios observed for these residue types in the Peptoid Data Bank are consistent with values measured by NMR (Table 1).

The data from the Peptoid Data Bank make clear a correlation between $X_1$ and $\omega$ (Figure 5). This result can be explained by simple steric repulsion: the side chain and the backbone adopt orientations that avoid steric clashes between them. This relationship between backbone and side chain can predict the side chain direction based solely on the backbone conformation. Likewise, it can be exploited to control backbone conformation by controlling the side chain.

4.3.1 | Implications for design

The structural trends evident in the Peptoid Data Bank provide support for a design strategy that our lab and others (notably...
Zuckermann, Gorske) have used: identify a target backbone conformation, select side chains at several positions that restrict the local backbone to a conformation consistent with the target, and add long-range constraints as appropriate. This design strategy takes advantage of a key feature of peptoids: the ability to install side chains that restrict the conformational landscape of their backbone. The only analog in peptide design is proline, with its restricted $\phi$ value. In fact, Hosseinzadeh et al. have established a design strategy for well-ordered peptides that relies on identifying positions in the target backbone that contain $\phi$ values consistent with L- or D-proline and installing the indicated proline at those positions.\(^{73}\)

Peptoid chemists can implement a much richer adaptation of this strategy, since peptoids can readily incorporate several chemical features into one side chain at the designer’s discretion. Furthermore, the degrees of freedom left unconstrained by many classes of peptoid residue offer unexplored opportunities for designing flexible, switchable molecules.\(^{74,75}\)

### 4.3.2 How to design a peptoid structure

Researchers without extensive prior experience working with peptoids may benefit from a fundamental description of the design approach used by some veterans of the field. A general recipe for peptoid structure design is given below (Figure 7), illustrated by two examples: a cyclic hexamer\(^{52}\) (code 08AC1) and a linear decamer\(^{49}\) (code 19AF1).

1. **Choose a peptoid backbone secondary structure.**
   Backbone structures may be chosen from prior experiment (i.e., from the Peptoid Data Bank), from simulation, or from a hypothesized structural motif. Both of the example designs were based on prior experiment: the cyclic hexamer was based on a previously characterized macrocycle (code 07AA1), while the linear decamer was based on the ribbon pentamer (code 17AB1) first solved by Gorske et al (which was, in turn, based solely on a structural hypothesis).

2. **Provide covalent constraints if appropriate.**
   Many of the structures in the Peptoid Data Bank make use of backbone cyclization to constrain the conformation. The example cyclic hexamer was cyclized head-to-tail with a native amide bond. Side chain crosslinking can also be used, as in 12AB1 and 12AB2.\(^{69}\)

3. **Place structure-inducing side chains in appropriate positions.**
   Design a peptoid sequence that includes branched side chains at positions with cis-amide bonds, and N-aryl (or N-X) side chains at positions with trans-amides. For a review of other side chain types that may be chosen to achieve a similar effect, see Kalita et al.\(^{18}\)

In the example of the cyclic hexamer, a repeating sequence of

![Figure 7](https://example.com/figure7.png)

**Figure 7** Residue-based design strategy illustrated with two peptoids from the Peptoid Data Bank. A backbone conformation is chosen from prior experiment, and then side chain types are chosen by C$_{\alpha}$N$_{\text{chemistry}}$ to respect the side chain-backbone relationships illustrated in Figure 6. (a) Design process for 08AC1 (cyclic hexamer with trans-inducing N-aryl side chains). (b) Design process for 19AF1 (Ribbon decamer with cis-inducing tertiary amine side chains). (c) Chemical structure diagrams for final compounds in (a) and (b).
Design side chains for the desired application
In any remaining positions, choose side chain functional groups relevant to the intended application. Existing designs have incorporated free radical catalytic sites, metal-binding moieties, and protein surface-targeting motifs among other functions.

### 4.4 | Emerging strategies

Until this point, we have focused on local interactions, leading to a residue-based design approach. It is possible, however, to contemplate further developments in which designs integrate multiple secondary structures as ordered “modules” within a larger construct. One example of this approach is the multicyclic 16-mer by Vollrath et al. (code 12AB2). The largest structure in the Peptoid Data Bank, this molecule was formed by symmetric dimerization of two octamer macrocycles via the conjugation of azide and alkyne side chain functional groups. The resulting 16-mer structure retains the backbone conformations of the cyclic octamer modules and secures them in a specific, stable arrangement. Another, more recent, set of examples are the dodecamers of Pierri et al. This remarkable family of structures are the largest single-chain molecules in the Data Bank, and they adopt a conformation that can be analyzed as four distinct modules: a right-handed helix, a turn, a left-handed helix, and another turn. Each helical module is three residues long, corresponding to one full turn of a PPI helix. The turn modules each contain three residues, which closely match the motif found within the familiar cyclic octamer structure (Figure 8).

The fields of protein structure prediction and design have recently been upended by breakthroughs in machine learning technology. Peptide chemists, too, are discovering applications of these technologies that promise to unlock new possibilities in that field. The peptoid field has neither the wealth of experimental data nor the consistency in machine-readable structural information necessary to train these models effectively, but we hope that our presentation of the Peptoid Data Bank may open up new possibilities.

### 4.5 | Future directions

Peptoid researchers seek to establish a robust and predictable relationship between oligomer sequence, structure, and function. The structures in the Peptoid Data Bank provide a wealth of information on how monomer sequence relates to the molecular conformation. Furthermore, the geometry of the molecular packing in the lattices of the X-ray crystal structures similarly sheds light on preferred chain-chain interactions, which will aid in the design of peptide-based nano-materials. There are decades of additional studies of bioactive and catalytic peptides, but the molecular structure of most of these compounds are not yet known. Thus far, the Peptoid Data Bank lacks compelling examples of molecules that both adopt well-defined structures and perform valuable functions. Our hope in presenting this guide is that researchers will be inspired to bridge these two paths of inquiry and contribute high-resolution structures of functional peptoids to the Peptoid Data Bank.

Future growth of the Peptoid Data Bank, both in its size and in its utility, will depend on two areas of development: first, on improvements in peptoid informatics; second, on establishing structural control over peptoid residues and using this control to design larger and more complex structures. Further developments in the informatics of peptoids, particularly for machine learning applications, will almost certainly require standardized residue naming conventions, and may also require a standardized file format to unify the few NMR structures with the many X-ray structures.

More work is needed to investigate the conformational influences of side chains containing both structure-inducing steric features and polar functional groups. As suggested above, electrostatic or electronic interactions between side chain and backbone may produce opposite effects to steric interactions, which dominate in the structure-inducing side chains reviewed here. Side chains that combine such polar groups with various steric features comprise an ongoing investigation in our lab.

While chiral N-alkyl side chains can establish control over ϕ while favoring a cis-amide bond, there are currently no peptoid side chains known to control ϕ while favoring a trans-amide. This blank space in our palette imposes a real limitation on the design strategy presented here. It should serve as a challenge to researchers in this field to discover such side chains, or to develop reliable strategies for accomplishing this level of control from higher-order interactions.
5 | CONCLUSIONS

The articles reviewed here report high-resolution experimental structures of peptoids, which we have collected into a resource we call the Peptoid Data Bank. In presenting the Peptoid Data Bank in its current form (databank.peptoids.org), we seek to make peptoid structural information readily accessible and standardized as a necessary condition to enable sophisticated sequence design protocols. Curated data collections like the Peptoid Data Bank are more important than ever in this new era of machine learning and artificial intelligence. We hope that the existence of this resource will inspire researchers with such expertise to apply these powerful tools to the problems of peptoid structure prediction and design.

This review supports an understanding that has led some peptoid researchers to a general design strategy: peptoid residues can be grouped by the chemical features at the Cα position of the side chain. Side chains from these groups can be used to control backbone conformation in a predictable, local manner. Now that this strategy can be illustrated conclusively, we anticipate that it may lead researchers to design novel peptoid structures that further enrich the Peptoid Data Bank. We hope that the residue-based design strategy described in this review may become a foundation upon which more sophisticated techniques may be built.

Peptoids will come to fruition as foldamers when we can reliably design stable tertiary structures and characterize them at high resolution. Without the benefit of peptoid tertiary structures in natural systems, we cannot be certain how this goal can be achieved. As the Peptoid Data Bank continues to grow, will the set of new structures be similar in size and topology to the first 100 we evaluate here, or will they continuously expand in complexity? Will the secondary structure modules identified thus far constitute building blocks for future tertiary structures? The answers to these questions will come from the community of peptoid researchers contemplating the Peptoid Data Bank and turning structural insights into design strategies.

ACKNOWLEDGMENTS

This work was supported by award CHE-2002890 from the National Science Foundation (NSF). The authors thank Miguel Grinberg for his freely available web development materials (miguelgrinberg.com).

CONFLICT OF INTEREST STATEMENT

Kent Kirshenbaum serves as Chief Scientific Officer for Maxwell Biosciences, Austin, Texas. Maxwell is developing commercial applications for peptoid oligomers as anti-infective agents.

Correspondingly, Kent Kirshenbaum has material financial interests in peptoid designs as applied towards anti-infective therapeutics.

DATA AVAILABILITY STATEMENT

The data that support the findings of this review are openly available in The Peptoid Data Bank at https://databank.peptoids.org.

Much of these data were derived from the Cambridge Structural Database at https://www.ccdc.cam.ac.uk/structures/.

Some additional data that supports the findings of this review are available in the supplementary material of this article.

The source code for the Peptoid Data Bank is available on github: https://github.com/Kirshenbaum-lab/peptoid-db-website.

ORCID

James R. B. Eastwood https://orcid.org/0000-0003-3895-5227
Ronald N. Zuckermann https://orcid.org/0000-0002-3055-8860

REFERENCES


AUTHOR BIOGRAPHIES

James R. Baggs Eastwood is a Ph.D. candidate in the chemistry department at New York University (New York, NY, U.S.A.). He uses a combination of experimental and theoretical techniques to study the conformational properties of peptoids, and develops tools for computational design of these molecules. He holds a B.S. in Nanosystems Engineering from Louisiana Tech University (Ruston, LA, U.S.A.) and is a native of New Orleans, LA (U.S.A.).

Ethan I. Weisberg conducted this work as a high school student at Packer Collegiate Institute (Brooklyn, NY, U.S.A.) and is now an undergraduate at the University of Pennsylvania (Philadelphia, PA, U.S.A.), majoring in Computer Science. He was named a Top 300 Scholar in the Society for Science’s 2021 Regeneron Science Talent Search. He is a native of Manhattan, NY (U.S.A.).

Dana Katz is a Ph.D. student in the chemistry department at New York University New York, (U.S.A.). She develops computational tools to design and evaluate peptoid macrocycles as inhibitors of protein-protein interactions. She holds a B.S. in Materials Science & Engineering from the University of Wisconsin-Madison, (Madison, WI, U.S.A.). She is a native of New York, NY (U.S.A.).

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.