Showcasing research from Dr. Zuckermann’s laboratory, The Molecular Foundry, Lawrence Berkeley National Laboratory, California, USA.

A bio-inspired approach to ligand design: folding single-chain peptoids to chelate a multimetallic cluster

Multimetallic clusters play a critical role in enzyme-mediated chemical transformations such as the conversion of water into fuel by sunlight. There is great interest in creating stable mimics of these enzymes, but they are difficult to design and synthesize. Here, we used a bio-inspired approach, where sequence-defined peptoid oligomers were designed to fold into three-dimensional structures that chelate a Co₄O₄ cubane cluster. 2D NMR spectroscopy was used to determine the atomic structure of these metalloenzyme mimics. This modular synthetic approach is efficient and provides a new general route for multimetallic coordination chemistry.

As featured in:

See Ronald N. Zuckermann et al., Chem. Sci., 2018, 9, 8806.
A bio-inspired approach to ligand design: folding single-chain peptoids to chelate a multimetallic cluster†

Andy I. Nguyen, Ryan K. Spencer, Christopher L. Anderson and Ronald N. Zuckermann†

Synthesis of biomimetic multimetallic clusters is sought after for applications such as efficient storage of solar energy and utilization of greenhouse gases. However, synthetic efforts are hampered by a dearth of ligands that are developed for multimetallic clusters due to current limitations in rational design and organic synthesis. Peptoids, a synthetic sequence-defined oligomer, enable a biomimetic strategy to rapidly synthesize and optimize large, multifunctional ligands by structural design and combinatorial screening. Here we discover peptoid oligomers (≤7 residues) that fold into a single conformation to provide unprecedented tetra- and hexadentate chelation by carboxylates to a [Co₄O₄] cubane cluster. The structures of peptoid-bound cubanes were determined by 2D NMR spectroscopy, and their structures reveal key steric and side-chain-to-main chain interactions that work in concert to rigidify the peptoid ligand. This efficient ligand design strategy holds promise for creating new scaffolds for the abiotic synthesis and manipulation of multimetallic clusters.

Introduction

There is growing interest in multimetallic clusters found in enzymes that catalyze important reactions such as water splitting (photosystem II), nitrogen fixation (nitrogenase), and carbon dioxide reduction (carbon monoxide dehydrogenase). Successful mimicry of these enzymes requires synthesis of multimetallic clusters in a simplified chemical environment; however, approaches to synthesize chelating ligands needed to stabilize and control cluster formation are underdeveloped. Since a multimetallic cluster has a larger volume than a single metal, complicating both the design and synthesis.

Typical ligand designs use branching motifs with idiosyncratic synthetic routes that can take substantial amounts of time to optimize. The few known examples of synthetic ligands that chelate clusters depend on high symmetry, in order to simplify organic synthesis. An ideal ligand synthesis would permit full access to asymmetry, site-differentiation, and donor diversity, while also remaining easy to synthesize.

An alternative approach to ligand design is to mimic the way nature bind metals (Fig. 1). The linear and modular architecture of polypeptides is an efficient way to modulate size, shape, and function of suitable ligands. Individual metal binding moieties dispersed in a linear sequence are brought into the appropriate geometry through folding of the polypeptide chain. The folding of sequence-defined oligomers is complex, but the modular and efficient synthesis permits the use of combinatorial approaches to seek out properly folded or functional sequences. We aim to develop modular peptidomimetic systems to rapidly create and evaluate large organic ligands for metal cluster chelation. Recently, a short minimalist peptide comprised of natural and mirror-image amino acids was designed to chelate a Fe₃S₄ cubane, demonstrating an important first step toward non-natural metalloenzymes.

N-Substituted glycine oligomers, also called “peptoids”, are a class of synthetic peptidomimetics with great potential for ligand design (Fig. 1). Like polypeptides, they are synthesized in an iterative fashion with their linear monomer sequence precisely defined at each elongation step. However, their synthesis via the solid-phase submonomer method is uniquely rapid and chemically diverse compared to other classes of sequence-defined polymers since the side-chain functional groups are readily introduced from easily obtainable primary amine synthons. Furthermore, the properties of peptoids can be efficiently evaluated by high-throughput synthesis and subsequent screening. Many laboratories have

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8sc04240c
engineered peptoid side-chains that locally influence the backbone dihedral angles to achieve robust secondary structures such as helices and sheets.\textsuperscript{16–23} In addition, peptoids can fold into a variety of other secondary structures including ribbons and square-helices.\textsuperscript{23–29} However, using peptoids as ligands for targeted metal coordination is still only recently gained momentum.\textsuperscript{30–33} With very few atomically-defined metallopeptoid structures. To date, most metallopeptoids have mainly been characterized by circular dichroism spectroscopy.\textsuperscript{34–36} Recently, the first examples of crystal structures of metallopeptoids were reported, however, their solution-phase characterization was limited.\textsuperscript{31} It is important to note that in cases like this where the molecules have multiple degrees of freedom and multiple pendant functional groups, crystal packing effects can yield solid-state peptoid conformations that do not correlate with their solution-phase structure.\textsuperscript{19,43} Here we address cluster chelation and peptoid folding in solution, through the design and discovery of conformationally stable peptoids that provide tetradentate to hexadentate chelation to a \([\text{Co}_3\text{O}_4]\) cluster \textit{via} several simple carboxylate moieties. The only previous example of peptoid–metalcluster interaction featured a lanthanide cluster capped with a peptoid monomer.\textsuperscript{44} NMR spectroscopy was used to elucidate the atomic structures of the peptoid cluster complexes in solution, and single-site mutation studies reveal several non-covalent interactions that contribute to stable peptoid folding.

The tetrametallic cobalt oxo cluster, \(\text{Co}_2\text{O}_4(\text{OAc})_4\text{py}_4\) (OAc = acetate, py = pyridine)\textsuperscript{45–46} (1) has garnered attention as a water splitting catalyst in the context of artificial photosynthesis and as a structural mimic for the oxygen-evolving center (OEC) in the active site of photosystem II (PSII).\textsuperscript{47–50} It also has convenient properties that make it well-suited for combinatorial and structural studies: kinetic stability, established ligand substitution chemistry,\textsuperscript{24} strong color (useful for colorimetric binding assay), and diamagnetism (useful for NMR spectroscopy). The cluster has a cubane geometry comprised of four cobalt(n) ions and four oxides. Designing a single peptoid chain that can encapsulate this cubane and determining its atomic structure should give general insight on the key elements required for metal cluster chelation, and provide inroads for synthesis of stable biomimetic “cutouts” of multietallic enzyme cofactors and their ligands outside of the native protein environment (Fig. 1).

Results and discussion

The crystal structure of PSII provides a good starting point for designing a peptoid ligand that can chelate a cubane.\textsuperscript{25} The cubane subcluster of the OEC is predominantly chelated by carboxylates at the interface of two different protein subunits, D1 and CP43. Examination of the D1 subunit reveals that it provides most of the carboxylate ligands to the cubane subcluster, three of which reside in a small 11-mer C-terminal domain (Fig. 2a). Two faces of the cubane are chelated by an aspartate side-chain (Asp342) and the alanine C-terminus (Ala344). A longer segment (9 residues) connects the second face to the third face \textit{via} Glu333. Here we set out to recreate the 3-dimensional orientation of the carboxylate moieties in the D1 fragment using a peptoid analogue of minimal length. To break down the complexity of the design, we first created peptoids that bound well to two faces of the cubane, and then built off that sequence to continue onto the third face. We therefore first synthesized the peptoid analogue of the C-terminal trimer fragment, Asp–Leu–Ala–OH, as Fm–Ncm–Nbn–NAla–NH\(_2\) (H\(_4\)A) (Fm = formyl, Ncm = N-carboxymethyl glycine, Nbn = N-benzyl glycine, NAla = N-carboxymethyl-L-alanine) (Fig. 2b) to target two-face binding of a cubane. Note that the isobutyl side-chain of Leu343 was changed to a benzyl moiety, but is not expected to have consequences on the binding behavior since it points away from the cluster.\textsuperscript{26} The peptoid and peptide have the same number of residues, but because the peptoid side-chains attach to N rather than C, the number of atoms between the carboxylates is longer in the peptoid (12 atoms) than in the peptide (11 atoms). Nonetheless, one equivalent of tripeptoid H\(_4\)A reacted with 1 to produce a complex, with \(m/z = 1169.12\) \((\text{Co}_2\text{O}_4(\text{OAc})_4\text{A}([\text{py}])_2\text{H}^+))\), consistent with binding of A to 1 and displacement of two acetic acid molecules (1-A) (Fig. S5†).\textsuperscript{5} H NMR spectroscopy of the purified complex in \(d_4\)-methanol\textsuperscript{14} showed four singlets \((1:1.5:0.7:1)\) ratio belonging to the formyl proton indicating that there are four conformational isomers (Fig. S6†).\textsuperscript{1} \(^{1}H–^{13}C\) HSQC showed two distinct classes of NAla methines \((1:1.3)\) (Fig. S8†). The chemical shifts of the methine, in addition to the nuclear Overhauser effect (NOE) crosspeaks, identify them as cis/trans amide isomers (\(\omega \sim 0\) or 180, respectively) of NAla with a \(K_{cis,trans} \approx 0.8\) (Fig. S10†). There is also cis/trans isomerization from the formyl cap, giving a total
of $^2$ conformational isomers. The $\omega$ of the second amide is undetermined, but it is either frozen in a single isomer or its conformation is coupled with that of the other residues. This distribution of isomers is consistent with the fact that in general, cis/trans isomerization of tertiary amides is nearly isoenergetic ($K_{cis/trans} \sim 1$).$^{23}$ The multiple overlapping resonances of these isomers precluded full assignment of the NMR spectrum for 1-A.

To remove the amide isomerization at each terminus, NAla was changed to Ala-OH (simulating cis-NAla since it places the 2-carboxyethyl functionality cis with respect to the carbonyl O atom) and formyl was replaced with p-toluenesulfonyl (Ts). Reaction of one and two equivalents of Ts-Ncm–Nbn–Ala–OH ($\text{H}_2\text{B}$) with 1 produced the mono-peptoid complex, $\text{Co}_4\text{O}_4(\text{Ts})_2(\text{B} \text{py})_4$ (1-B), and bis-peptoid complex, $\text{Co}_4\text{O}_4(\text{B} \text{py})_4$ (1-2B), respectively, demonstrating that termini modification did not significantly affect binding ability. The expected masses for 1-B and 1-2B were observed ($m/z = 1238.06$ and 1623.24, respectively; Fig. S11 and S13†), and their $^1$H NMR spectra in $d_1$-methanol showed only one conformer (Fig. S12 and S14†). In contrast, the free peptoid shows two rotamers due to isomerization about the amide bond between Nbn and Ncm (Fig. S2†). While the spectra for 1-B and 1-2B show similar features, the higher symmetry of 1-2B allowed a more facile assignment. The NMR spectrum of 1-2B is consistent with a $C_2$ symmetric molecule, evidenced by two sets of equally intense geminal protons and a singular set of peptoid resonances. This $C_2$ point group symmetry implies that the two chains wrap around the equatorial plane of the cubane in a "head-to-tail" or antiparallel fashion. In contrast, parallel alignment of the peptoid chains around the cubane would result in $C_1$ symmetry that would be distinguishable by NMR spectroscopy. Methylene protons of the peptoid backbone and side-chains are diastereotopic with geminal protons separated up to 1.5 ppm (Fig. S12 and S14†), suggestive of a rigid fold. In addition to mass spectrometry and NMR data, evidence that the $[\text{Co}_4\text{O}_4]$ core remains intact upon peptoid coordination comes from solution electrochemistry data, which is a known metric that is sensitive to the $[\text{Co}_4\text{O}_4]$ coordination environment.$^{24}$ For context, the $[\text{Co}_4\text{O}_4]$ core in compound 1 exhibits a characteristic reversible redox event in the cyclic voltammogram (CV) corresponding to the $[\text{Co}^\text{III}_4\text{O}_4]^+/\text{Co}^\text{IV}_4\text{Co}_4\text{O}_4]^2+$ couple (0.35 V vs. $\text{Fe}^+/\text{Fe}$ in methanol, Fig. S33†), and this redox couple has been previously shown to be also sensitive to the ligand environment around the cubane.$^{25}$ Compound 1-2B exhibits a nearly identical feature at 0.34 V in the CV (Fig. S33†), suggesting that the cubane core is preserved and surrounded in a highly similar ligand environment to that of 1.

Multiple NOE crosspeaks were used as distance restraints for structural determination (Fig. S34, S35, S42†) by replica-exchange molecular dynamics (REMD) to obtain low energy structures consistent with the NMR data (Fig. 2c). In particular, a strong NOE crosspeak between the backbone methylene protons of Ncm and Nbn indicates a less than 3Å separation, which is diagnostic of a cis-amide linkage between those two residues. The determined NMR structure is consistent with expectations, with the two $\mu_2$-carboxylate donors per peptoid interacting with three cobalt ions, for an overall tetradentate chelation per peptoid. Analysis of the backbone dihedral angles, $\phi$ (ϕ) and $\psi$ (ψ), gives insight into the validity of the structure (Fig. 3a). Satisfyingly, dihedral combinations (ϕ, ψ) of the backbone reside in the low energy regions of the cis-peptoid Ramachandran plot (Fig. 3b).$^{27,54}$

It is noteworthy that Ncm adopts a cis-amide, whereas PSII’s Asp–Leu–Ala–OH sequence is all trans; perhaps the shorter Cα–Cz distance in a cis-amide backbone is needed to account for the longer linkage (12 atoms) between the two carboxylate groups in the peptoid (vide supra). Titration of a monodentate

![Image](https://example.com/image1.png)

**Fig. 2** (a) Selected amino acid residues from the D1 subunit (blue) that bind to the Mn$_4$Ca oxygen-evolving complex (OEC). (b) Design of the peptoid analogues based on D1. (c) Ensemble of the five lowest energy NMR structures of 1-2B (pyridine ligands omitted for clarity). (d) The lowest energy structure of 1-2B.
Competitor, d₄-acetic acid into 1-2B, monitored by NMR spectroscopy, gave stepwise stability constants of \( \log(K_{a1}) = 2.0 \pm 0.6 \) and \( \log(K_{a2}) = 0.9 \pm 0.1 \), and overall stability \( \log(\beta) = 2.9 \pm 0.7 \), demonstrating a stabilization from the chelate effect (Fig. S30–S32†).

Next, we attempted to extend the sequence of 1-2B to allow carboxylate binding to the third face within a single ligand. Similar to the tripeptide fragment that binds the OEC, the N-terminus of [Co₄O₄]-bound B is pointed outward and slightly away from the third face of the cluster such that more than one peptoid monomer would be needed to connect to the third face in a \( \mu_2 \)-fashion. Preliminary geometry-optimized 3-dimensional computer models suggested that a three peptoid monomer chain should be long enough to link the Ncm–Nbn–Ala–OH segment to a carboxylate residue, Nce (N-2-carboxyethyl glycine), bound on the third face. It was unclear whether there would be a preference for cis or trans amides in this loop (as either appeared reasonable), and concern that a loop that is too flexible would diminish binding.

Thus, we employed a combinatorial approach for discovering classes of loop sequences that would result in high-affinity cluster-binding peptoids. The “split-and-pool” strategy," was used to synthesize a “one-bead-one-compound” library of 125 resin-bound sequences varied at the three loop positions (Scheme S2†). In this method, the ensemble of beads contains all possible sequence combinations, but a single bead contains only a homogeneous population of a unique sequence. The peptoids were covalently attached to a TentaGel macروبead support by a labile methionine linkage that can be cleaved using cyanogen bromide, which is orthogonal to the synthesis and screening conditions (see ESI† for details). The monomer basis set comprised of two hydrogen bond donors, Nae (N-2-aminoethyl glycine) and Nhe (N-2-hydroxyethyl glycine), a chiral Nrpe (N-(R)-1-phenylethyl glycine), a strictly trans-enforcing residue Nanis (N-4-methoxyphenyl glycine), and a neutral residue, sarcosine (N-methyl glycine). The peptoid library was synthesized by solid-phase synthesis, and side-chains were deprotected (without cleavage from the resin) by trifluoroacetic acid. The [Co₄O₄] unit was complexed onto the resin-bound peptoids by ligand exchange with 1.

Since the [Co₄O₄] unit is a dark green color, direct colorimetric analysis was used to identify hit sequences. Subjecting the resin-bound clusters to increasing concentrations of acetic acid at 50 °C for 20 h and selecting for the remaining green-colored beads allowed identification high-affinity sequences. As a benchmark, the resin-bound tripeptoid, Ts-Ncm–Nbn–NALa, showed complete loss of green color at 1.0 M acetic acid (Fig. S3†). While most beads in the heptameric library lost color,

Fig. 3 (a) The \( \phi \) and \( \psi \) dihedrals of a peptoid. (b) Backbone dihedral combinations (\( \phi, \psi \)) of the lowest energy NMR structure for 1-2B (triangles) overlaid on the cis-peptoid Ramachandran free energy plot.27

Fig. 4 (a) Primary sequence of peptoid \( \text{H}_{3}\text{C} \). (b) Overlay of the five lowest energy NMR structures, with pyridine and acetate removed for clarity. (c) Peptoid backbone view only. (d) A selected low energy NMR structure.
Peptoid H₂C reacts with 1 in a 1 : 1 ratio to form the expected complex 1-C, with expected m/z = 1669.27 (Co₄O₄(OAc)(C)(py)₄⁺) and 835.14 (Co₄O₄(OAc)(C)(py)⁴H₂⁺) (Fig. S17†). The ³H NMR spectrum of purified 1-C in d₄-methanol exhibits ~51% of a single conformation species having sharp resonances, and the remaining less folded conformer exhibiting very broad resonances (Fig. S18†). Backbone methylene protons of the single conformation species are diastereotopic with large separation consistent with rigid folding. This species also exhibits a nearly identical redox couple to that of 1 and 1-2B at 0.36 V that suggests the [Co₄O₄] core remains intact (Fig. S33†).

The NMR resonances were fully assigned (Fig. S21†), and many NOE crosspeaks were resolved for structure determination (Fig. S36–S38†). Notably, the appearance of medium strength NOE crosspeaks between i, i + 1 and i, i − 1 backbone methylene protons correspond to cis amides for all peptoid residues. The side-chain methylene protons of Nce and Nae are also diastereotopic, and show a J-coupling pattern consistent with a gauche relationship of the non-hydrogen substituents (Fig. S39†). While the side-chain of Nce bound to the cluster should adopt a rigid rotational state, the rigidity of the Nae side-chain is more surprising and suggests that it may be involved in a strong non-covalent interaction. Several crosspeaks between the loop residues and the pyridine ligands of the cluster are also observed. NMR-constrained REMD was used to elucidate the structure of 1-C (Fig. 4 and S43†). All backbone dihedral angles fall into the stable regions of the peptoid Ramachandran plot (Fig. 5). Travelling from the C to N terminus, the chain begins to dip far below the equatorial plane of the cubane a[...]

~25% of the beads retained color up to 3.0 M acetic acid (Fig. S3†). The peptoids belonging to the colored beads were cleaved and sequenced with tandem MS/MS (see ESI† for details). A hit sequence, Ts–Nce–Nhe–Nrpe–Ncm–Nbn–NAla, appeared multiple times and was chosen for further analysis. The hit was resynthesized as the C-terminal acid derivative, Ts–Nce–Nhe–Nrpe–Ncm–Nbn–Ala–OH (H₃C), in order to prevent amide isomerization at the termini (vide supra). Note that due to trifluoroacetic acid used in HPLC purification procedures, Nae is in the ammonium form with a trifluoroacetate counterion.

Fig. 5 Backbone dihedral combinations (φ, ψ) of the lowest energy NMR structure for 1-C (triangles) overlaid on the cis-peptoid Ramachandran free energy plot. The residue numbers are indicated.

Fig. 6 (a) Close-up view of the loop region of 1-C. The loop residues are highlighted with green carbon atoms and labeled. The dashed line indicates the H-bond donor–acceptor distance. (b) LCMS of the crude reaction mixture between 1 and the mutated loop derivatives of H₂C (both 0.5 mM) in methanol after 20 h at 50 °C. The asterisk (*) indicates residual 1, and the red circle indicates peaks with masses corresponding to the hexadentate chelated cluster. Mobile phase: 5–95% gradient of acetonitrile/water with 10 mM NH₄OAc/HOAc (pH ~ 5.2) buffer.
adopts similar dihedrals as Nrpe4, beginning a partial helical stretch that exits back up to the cubane equatorial plane at Nae2. The chirality of the backbone is reversed at Nae2, creating a ∼90° bend that positions the chain parallel with the cubane face, allowing Nce1 to coordinate the cubane. The backbone can sample both positive and negative stable (ϕ, ψ) regions due to the prochiral nature of peptoids (Fig. 5). This flexibility seems advantageous in allowing such a short oligomer to “freely” mold into the most suitable fold. The cis-amides of the backbone are reinforced in two ways – (1) Nae forms a 7-membered H-bond to the i − 1 carbonyl (average N=–O distance of 3.2 ± 0.3 Å, average N–H–O angle of 122.3° ± 13.6°), and (2) the steric bulk of Nrpe favors a cis-amide in order to minimize allyllic strain (Fig. 6a).

Single mutation of each of the three loop residues to sarcosine,* analogous to alanine scanning*9 in peptides, was used to probe the structure–function relationship of each residue (Fig. 6b). Liquid chromatography mass spectrometry (LCMS) of the crude reaction mixture of 1 with H4C exhibited a large major peak corresponding to the desired complex, 1-C. Removing the cationic H-bond donor via a Nae2Sar mutation was highly detrimental, producing significant amounts of side products, including partially chelated and a dimeric cluster complex. Even a conservative mutation to an isosteric, weaker H-bond donor, Nae2Nhe, yielded similar side products. However, chelation ability is unaffected by an Nhe3Sar mutation. NMR characterization of the purified cluster-bound Nhe3Sar mutant (1-C-Nhe3Sar) shows a conformation with many characteristic NOE features of 1-C, indicating an identical fold (Fig. S24, S27, S40, S41†). This suggests that the third residue may be useful as a site for future substitution or conjugation. An Nrpe4Sar mutation does not produce as many side products, but gives two identical-mass species having different retention times, suggestive of isomers (possibly isomers about the ω or φ dihedral angle between residues 3–4). The results of this sarcosine scan are in line with the NMR structure of 1-C – the H-bond from Nae and steric constraints of Nrpe are critical in decreasing backbone flexibility leading to better binding.

Conclusions

We show that the synthetic efficiency of peptoids enables rapid strategies to overcome the highly challenging design of chelating ligands for multimetallic clusters. By exploiting the versatile solid-phase syntheses of sequence-defined oligomers, cluster-binding peptoids were discovered and optimized using both structure-based design and combinatorial screening. The better binding properties of the optimized sequences, which exhibit less aggregation and more quantitative chelation, result from stronger non-covalent interactions within the peptoid chain that decreases conformational entropy. The peptoid allows for binding to the metal cluster in an anisotropic fashion, where every residue of the peptoid is modifiable, allowing for site-differentiation. Looking forward, the techniques established in this work can be extended to target the chelation of the fourth acetate face, or even the pyridine faces. Preliminary studies with an non-optimized 11-mer peptoid, H4D, made by extending peptoid C’s sequence with another

three-residue loop (based of the loop discovered for H4C) and an Nce3 residue, exhibits the correct m/z value for octadentate binding by displacement of all four acetates from 1 to form Co4O4(D)2(py)4 (Fig. 529†). Although this species is predominant in the mass spectrum, a noticeable amount of hexadentate species (partially chelated) is present. Nonetheless, it demonstrates that the peptoid sequence can be systematically and easily extended to increase denticity. Also notable is that these compounds are the first conformationally stable metallopeptoids to be structurally characterized in solution phase, an important milestone in understanding metallopeptoid dynamics to create a new class of stable, protein-mimetic metal catalysts.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Dr John Edison for providing the raw data for the Ramachandran plot.

Notes and references


40 M. Baskin and G. Maayan, Chiral Cu(II), Co(II) and Ni(II) complexes based on 2,2′-bipyridine modified peptoids, *Dalton Trans.*, 2018, **47**, 10767–10774.


53 d$_1$-methanol was chosen for its ability to dissolve the peptoid-cluster complexes in high concentrations.


