Bottom-up design of peptide nanoshapes in water using oligomers of *N*-methyl-L/D-alanine

Jumpei Morimoto[1,*,†], Yota Shiratori[2,†], Marin Yokomine[1,†], Takumi Ueda[3], Takayuki Nakamuro[4], Kiyofumi Takaba[5], Saori Maki-Yonekura[5], Koji Umezawa[6,7], Koichiro Miyanishi[8,9], Yasuhiro Fukuda[1], Takumu Watanabe[1], Wataru Mizukami[9], Koh Takeuchi[3], Koji Yonekura[5,10,11], Eiichi Nakamura[4], Shinsuke Sando[1,2,*]

[†]These authors contributed equally. The names of the three authors are listed in alphabetical order.
[*]Corresponding authors

[1]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
[2]Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
[3]Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
[4]Department of Chemistry, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
[5]RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan.
[6]Department of Biomedical Engineering, Graduate School of Science and Technology, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano 399-4598, Japan.
[7]Institute for Biomedical Sciences, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University, Matsumoto, Nagano 390-8621, Japan.
[8]Division of Advanced Electronics and Optical Science, Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[9]Center for Quantum Information and Quantum Biology, Osaka University, 1-2 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[10]Advanced Electron Microscope Development Unit, RIKEN-JEOL Collaboration Center, RIKEN Baton Zone Program, Sayo, Hyogo 679-5148, Japan.
[11]Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai, Aoba-ku, Miyagi 980-8577, Japan.

Abstract

De novo design of peptide nanoshapes is of great interest in biomolecular science since the local peptide nanoshapes formed by a short peptide chain in the proteins are often key to the biological activities. Here, we show that the *de novo* design of peptide nanoshapes with sub-nanometer conformational control can be realized using peptides consisting of *N*-methyl-L-alanine and *N*-methyl-D-alanine residues as studied by NMR, X-ray and XFEL crystallographic and computational analyses as well as by direct imaging of the dynamics of the peptide's nanoshape using cinematographic electron microscopic technique. The conformation of *N*-methyl-L/D-alanine residue is largely fixed because of the restricted bond rotation, and hence can serve as a scaffold on which we can build a peptide into a designed nanoshape. The local shape control by per-residue conformational restriction by torsional strains starkly contrasts with the global shape stabilization of proteins based on many remote interactions. The oligomers allow the bottom-up design of diverse peptide nanoshapes with a small number of amino acid residues and would offer unique opportunities to realize the *de novo* design of biofunctional molecules, such as catalysts and drugs.

Introduction

Molecular functions depend on the three-dimensional structures of the molecules. Therefore, the establishment of the design principles of three-dimensional molecular nanoshapes in water is essential in molecular science for producing biofunctional molecules. The biological system utilizes proteins composed of nature's choice of suitable sequences of amino acids. Their diverse functions originate from local three-dimensional peptide nanoshapes formed by a short peptide chain constructed within the global protein structure. For example, an antibody is composed of 600–700 amino acid residues, but its essential function, i.e., antigen recognition, is majorly achieved by one of the local loop structures called complementarity determining region H3 consisting of 2–26 amino acid residues[1]. Each amino acid residue that constitutes a protein is intrinsically conformationally flexible, but the conformational freedom is restricted by remote interactions among amino acid residues, e.g., hydrogen bonds, *van der* Waals contacts, hydrophobic interactions, and electrostatic interactions, resulting in the construction of the tertiary structure of proteins (Figure 1a). Many interactions among amino acid residues are required to realize a hierarchically stabilized global protein structure and peptide nanoshapes therein[2–4], and previous studies suggest that approximately 50 amino acids are necessary for proteins to adopt their folds[5].

There are rare exceptions called miniature proteins, e.g., Trpzip[6], Trp-cage[7], chignolin[8,9], HP7[10], and a small protein containing a WSXWS motif[11], that achieve global stabilization of a three-dimensional structure with less than 20 amino acid residues (Figure 1a). The interplay of extensive inter-residual interactions can realize such a small protein. For example, the β-sheet structure of CLN025, a chignolin analog, is stabilized by hydrogen bonds and an aromatic–aromatic interaction[9]. The helix–loop–strand structure of the miniature protein containing a WSXWS motif is stabilized by an extensive cation–π interaction network[11]. However, ten or more amino acid residues are still required to achieve a stable global structure and local peptide nanoshapes.

Bottom-up peptide design using conformationally-fixed amino acid residues is a promising alternative approach to creating peptide nanoshapes with a minimal number of amino acid residues. Having a fixed local conformation, some amino acid residues strongly stabilize specific structural motifs per amino acid residue, and hence can be utilized for the bottom-up design of peptide nanoshapes (Figure 1b). Proline residue is a typical example. The ring constraint of proline renders its oligomer to form a rod-like shape called PPII helix stably[12]. An oxopiperazine residue is another example of a conformationally-constrained building block. Oligomers of oxopiperazine are conformationally constrained and have been utilized as molecular scaffolds[13–15]. *N*-methylalanine (NMA) residue has also been recently reported to be an amino acid residue with a fixed conformation. The conformation of the NMA residue in its oligomers is rigidified by pseudo-1,3-allylic strain, and the oligomer forms a stable extended shape[16–18]. These amino acid residues have hence served as attractive building blocks for the bottom-up design of peptide nanoshapes[19,20]. The bottom-up

approach enables the direct design of peptide nanoshapes with a minimal number of amino acid residues. However, the bottom-up design of stable peptide nanoshapes using a single building block has so far realized repetitive linear sequences with limited structural diversity.

The principal moment inertia (PMI) analysis[21] is a method to compare the three-dimensional structures of different classes of molecules quantitatively, and the PMI plots in Figure 1a, right and Figure 1b, right illustrate the large gap between the peptide nanoshape diversity realized in proteins and that realized by the bottom-up peptide design. The analysis extracts the three principal moments of inertia from the molecules and plots the information on a triangular diagram indicating the rod-, disc-, and sphere-likeness of the three-dimensional molecular structure. Here, for a fair comparison, the PMI analysis was conducted using the main chain atoms (N, C_α, and C=O) of six consecutive amino acid residues in each miniature protein and peptide. As represented by the PMI plot[21], a wide range of three-dimensional structures including rod-like, disc-like, and sphere-like are realized by the peptide nanoshapes within the hierarchically folded miniature proteins (Figure 1a). In contrast, the bottom-up peptide nanoshape designs have realized only rod-like shapes using conformationally-fixed amino acid residues (Figure 1b). Since the molecular shape visualized by the PMI plot is intimately linked to the biological activity of the molecules[21], expansion of the peptide nanoshape diversity realized by the bottom-up approach is important.

Here, we report a non-hierarchical bottom-up design of diverse peptide nanoshapes that self-stand in water using a combination of *N*-methyl-L-alanine (L-NMA) and *N*-methyl-D-alanine (D-NMA) residues as building blocks (Figure 1c). We reasoned that a combinatorial usage of multiple different building blocks with fixed conformations would realize diverse three-dimensional molecular nanoshapes. In this study, we describe sub-nanometer conformational control of the NMA oligomers through a rational combination of multiple L-NMA and D-NMA residues, as studied by NMR, crystallographic and computational analyses as well as by direct imaging of the dynamics of the peptide's nanoshape using a cinematographic electron microscopic technique[22,23]. The diversity of the peptide nanoshapes achieved by the bottom-up design using oligo-L/D-NMA approaches the level realized by miniature proteins (Figure 1c, right).

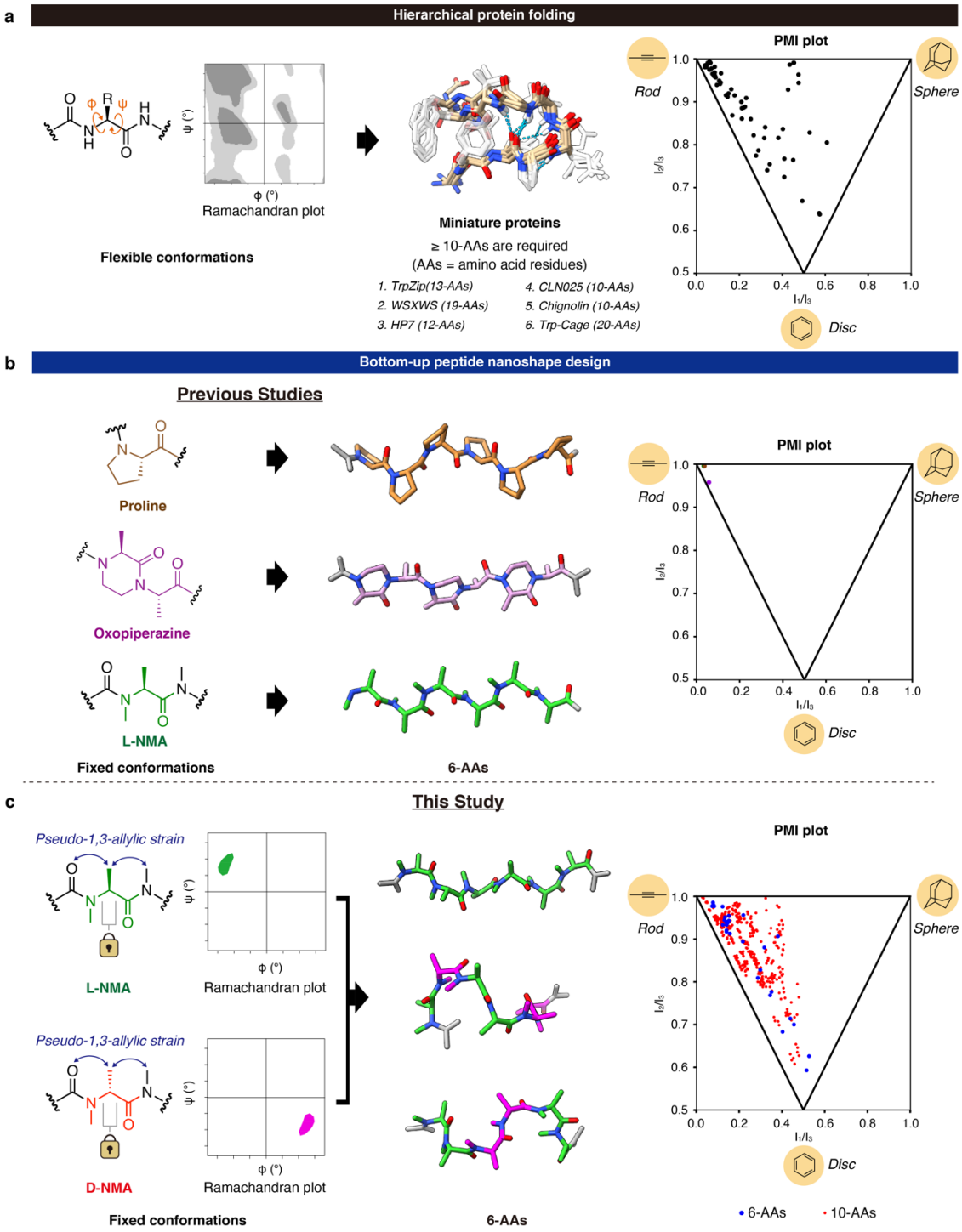


Figure 1 | The comparison of peptide nanoshape in the globally folded protein structures and the bottom-up formation of peptide nanoshape consisting of conformationally fixed amino acid residues. a. The conformation of each amino acid residue in proteins changes depending on the amino acid sequence. The Ramachandran plot was generated using PyRAMA (https://github.com/gerdos/PyRAMA) based on the previously reported ϕ and ψ preferences[24]. The three-dimensional structure of chignolin (PDB ID: 1UAO) is shown as an example of miniature

proteins. The peptide nanoshapes in the representative miniature proteins are plotted on the PMI plot. Plotted are the peptide nanoshapes of consecutive six amino acid residues (AAs) in the proteins. **b**. The peptides consisting of the previously reported amino acid residues with fixed conformations. Oligoproline (brown), oligooxopiperazine (purple), and oligo-NMA (green) are plotted on the PMI plot. (Oligoproline and oligo-NMA are overlapped on the plot.) The structures of oligoproline and oligo-L-NMA are produced using CCDC-1014542 and CCDC-278108, respectively. The oligooxopiperazine structure was produced using a 3D molecular modeling software (Avogadro) based on the previous report[13]. **c**. The peptides consisting of L-NMA and D-NMA residues that have fixed conformations. The peptide nanoshapes realized by oligo-NMA of 6-AAs and 10-AAs are plotted on the PMI plot.

Results and Discussions

Two conformationally constrained monomers as building blocks for the *de novo* design of the three-dimensional peptide nanoshapes. Previously, density-functional theory (DFT) calculations suggested that a minimal model of L-NMA oligomer stably forms conformations that avoid pseudo-1,3-allylic strains. To further confirm that the building block maintains the same stable conformation in water, more detailed calculations and experimental validations were conducted. The allowed conformational space of an amino acid residue in a peptide chain is commonly visualized using the Ramachandran plot[25]. Here, we recruited a similar plot on which the free energy of an amino acid residue is plotted against the two conformational determinants: dihedral angles ϕ and ψ. We first reproduced the Ramachandran-type plot of acetyl-*N*-methyl-L-alanine dimethylamide (**1**), a minimal model of L-NMA oligomers, by DFT calculations with minor modifications on the procedures from the previous report.[17] In these calculations, ϕ and ψ values of **1**, which are the two determinants of the conformation of the molecule, were combinatorically rotated from –180° to 180° by 15° and the 625 conformers were geometry optimized by DFT calculations at the B3LYP/6-31G(d) level of theory with implicit water model. The energy of each conformer was calculated at the same level and plotted. The generated Ramachandran-type plot of **1** was shown to have a narrow low-energy region around the lowest energy conformer with (ϕ, ψ) = (–135°, 75°), which is consistent with the previous report[17] (Figure 2a). To understand the stability of the lowest-energy conformer, the conformations at eight local minima including (ϕ, ψ) = (–135°, 75°) on the Ramachandran-type plot were further geometry optimized without fixing the ϕ and ψ values and the energies of the optimized conformations were calculated at the B3LYP/6-311G(d) level of theory. Based on the calculated energies, the Boltzmann weight of the lowest-energy conformer that has (ϕ, ψ) = (–130°, 73°) was 99% (Figure 2b and Table S1). This result suggests that the minimal model structure forms a fixed conformation in water.

To experimentally validate that the minimal L-NMA model stably forms the lowest-energy conformer from the DFT calculations, we conducted NMR measurements in water and obtained nuclear Overhauser effect (NOE) signals that reflect the distances between the pairs of protons (Figure

S1–4 and Table S2–3). The consistency of the observed NOE signals with the proton proximities on the stable conformers was examined. On the stable conformers from the DFT calculations, the two proton pairs were found to come close to each other: intra-residual *N*-methyl protons and β-protons; α-proton and *N*-methyl protons of the C-terminal structure. This is reasonable considering the conformational restrictions by pseudo-1,3-allylic strains. Consistent with the proton proximities in the DFT model (Figure 2b), medium to strong NOE signals were observed for the two proton pairs (Figure S4 and Table S2). The NOE signals strongly suggest that the minimal L-NMA model stably forms the lowest energy conformer from the DFT calculations.

X-ray crystallographic studies were also conducted to obtain an atomic-resolution snapshot of the monomer structure. To generate single crystals, derivatives of **1** with different N/C-terminal groups were synthesized and subjected to crystallization. From several tested compounds, acetyl-*N*-methyl-L-alanine with a C-terminal 1-(*tert*-Butyloxycarbonyl)piperazine (Boc-piperazine) (**2**) provided single crystals with good X-ray diffraction. The solved structure was close to the lowest-energy conformer of the minimal monomer model **1** from the DFT calculations. The ϕ and ψ angles of the solved structure were –134.6° and 77.7°, respectively (Figure 2c), which is close to those of the most stable conformation predicted by the DFT calculation. These computational and experimental studies strongly suggest that the building block (L-NMA residue) forms a fixed conformation in water.

The Ramachandran-type plot of acetyl-*N*-methyl-D-alanine dimethylamide (**3**), which is a D-NMA minimal model, was also generated in the same procedures. As expected, the Ramachandran-type plot is point symmetric to the plot of the L-NMA minimal model **1** and has a narrow low-energy region around (ϕ, ψ) = (135°, –75°) (Figure 2d). These results indicate that L-NMA residue and D-NMA residue can be used as building blocks with a fixed conformation for a bottom-up peptide nanoshape design.

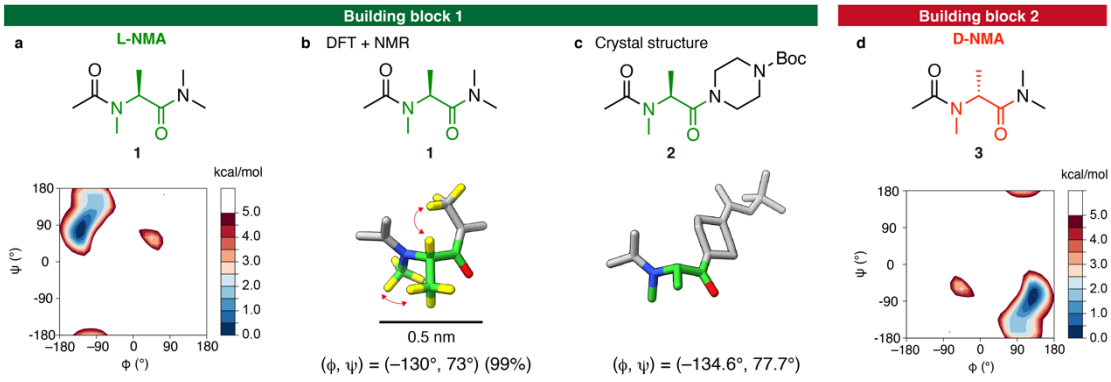


Figure 2 | Ramachandran-type plots of minimal model structures of oligo-*N*-methylalanine. a. Ramachandran-type plot of acetyl-*N*-methyl-L-alanine dimethylamide (a minimal model of L-NMA oligomers). **b.** The DFT-optimized structure of L-NMA. The lowest-energy conformer in the Ramachandran-type plot in Figure 1a was geometry optimized. Hydrogen pairs on non-neighboring carbons showing medium to strong NOE signals are connected using

red arrows. The hydrogens with NOE signals are shown in yellow. **c.** The crystal structure of acetyl-*N*-methyl-L-alanine with a C-terminal Boc-piperazine. Hydrogen atoms are omitted for clarity. For **b** and **c**, atoms and bonds other than the L-NMA residue are shown in gray. **d**. Ramachandran-type plot of acetyl-*N*-methyl-D-alanine dimethylamide (a minimal model of D-NMA oligomers).

A minimal model study of the bottom-up peptide nanoshape design using dimers consisting of L-NMA and D-NMA residues. We hypothesized that combinatorial usages of the two building blocks can realize the rational control of diverse peptide nanoshapes with sub-nanometer conformational control. Because the conformations of the building blocks are largely fixed, an oligomer is expected to form a predictable nanoshape that is a simple connection of the stable conformations of the building blocks.

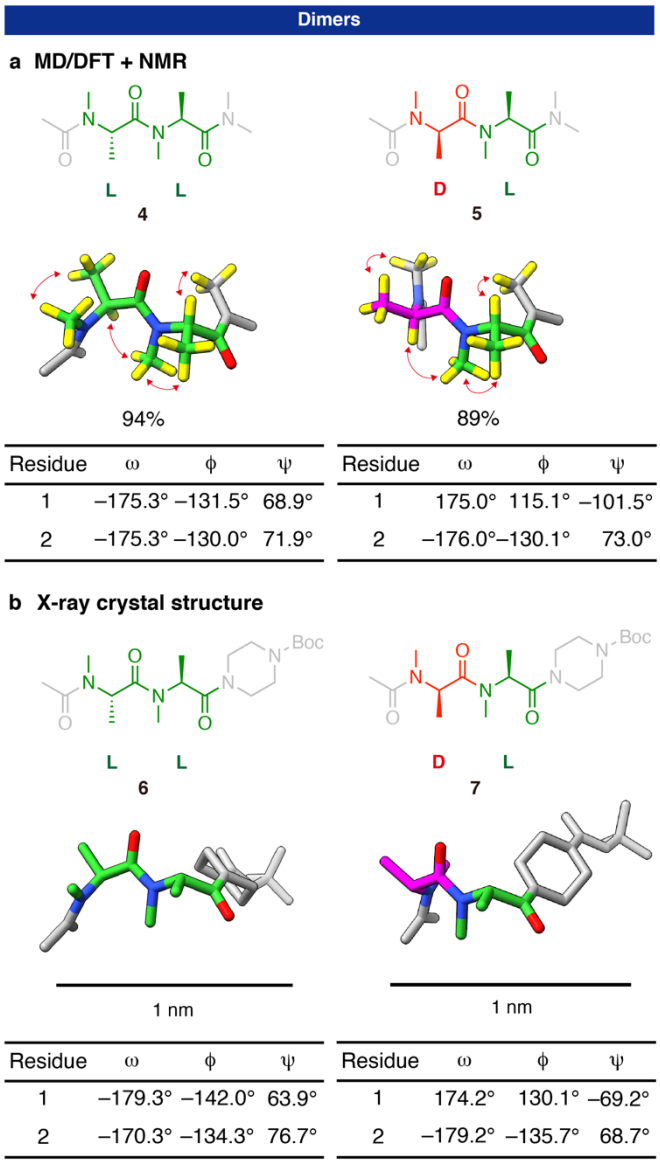
To examine the hypothesis, we first analyzed the three-dimensional structures of dimers, which are the shortest oligomers. LL dimer (**4**) and DL dimer (**5**) were studied (Figure 3). (L and D in the oligomer sequence denote L-NMA residue and D-NMA residue, respectively.) DD dimer and LD dimer were not examined since they are mirror images of the two dimers.

The stable structures of the two dimers were assessed using a combined computational and experimental approach. We first conducted multicanonical molecular dynamics (McMD) simulations[26] to comprehensively explore the allowed conformational space of the two dimers. The conformers at 300 K from the McMD simulations were clustered, and a representative conformer of each cluster was geometry optimized by DFT calculations at the B3LYP/6-311G(d) level of theory. The free energy of the geometry-optimized conformation was calculated by the DFT method with harmonic frequency analysis at the same level of theory. As a result, the most stable conformation of both the dimers was found to have dihedral angles (ϕ, ψ) around (–135°, 75°) for L-NMA residue and (135°, –75°) for D-NMA residues, which are the most stable points at the Ramachandran-type plots. In the most stable conformations, the two amide bonds in the dimers were found to be both in the *trans* configuration. The Boltzmann weight of the most stable conformer of **4** (LL) and **5** (DL) was calculated to be 94% and 89%, respectively (Figure 3a and Table S4). None of the other conformers were with a Boltzmann weight over 5%. These results suggest that the dimers stably form uniform three-dimensional structures that are predictable from the stable conformations of the monomers.

We conducted NMR measurements to experimentally validate the conformations of the dimers predicted by the computations (Figure S5–12 and Table S5–8). On the ^{1}H-NMR spectrum of the LL dimer **4**, there was a dominant set of peaks with two sets of minor peaks. According to the NOE signals, the dominant set of peaks corresponds to a conformational state with which the amide connecting the acetyl group and the 1st residue and the amide connecting the 1st and 2nd residues are both in the *trans* configuration. The two minor sets of peaks correspond to conformational states in which one of the two amides is in *trans,* and the other is in *cis*. More than 80% of both the amides

were in the *trans* configuration, which is evidenced by the existence of the inter-residual NOE between *N*-methyl protons and β-protons and the absence of the inter-residual NOE between α-proteins (Figure S8 and Table S5). This is consistent with the McMD result that the two amides in the dominant conformer are both in the *trans* configuration. As with the monomer structure, two kinds of hydrogen pairs are in close proximities on the stable conformer: 1. Intra-residual *N*-methyl protons and β-protons; 2. α-proton and *N*-methyl protons of the following NMA residue or C-terminal structure (Figure 3a, left). Consistently, medium to strong NOE signals were observed for these hydrogen pairs (Table S5). A similar result was obtained for the DL dimer **5**. 82% of the conformational ensemble was in a conformational state in which all the amides were in the *trans* configuration (Figure S12 and Table S7). In addition, the NOE signals expected from the proton proximities on the calculated stable conformer were observed (Figure 3b, right and Table S7). These results have shown that the dimers stably form the three-dimensional structures in which the most stable conformation of L-NMA residue and D-NMA residue are simply connected with a *trans* amide bond.

X-ray crystallographic study of the dimers further confirmed the dimer structures. Dimers with the *N*-terminal acetamide and C-terminal Boc-piperazine amide (**6** and **7**) gave X-ray quality single crystals. The conformers found in the crystals are close to the most stable conformers predicted by the calculations. All the dihedral angles in the crystal structures (Figure 3b) were within ±15° from the values of the most stable conformers of L-NMA residue ($\phi = -130°$, $\psi = 73°$) and D-NMA residue ($\phi = 130°$, $\psi = -73°$) (Figure 2b).



Residue	ω	ϕ	ψ
1	−175.3°	−131.5°	68.9°
2	−175.3°	−130.0°	71.9°

Residue	ω	ϕ	ψ
1	175.0°	115.1°	−101.5°
2	−176.0°	−130.1°	73.0°

Residue	ω	ϕ	ψ
1	−179.3°	−142.0°	63.9°
2	−170.3°	−134.3°	76.7°

Residue	ω	ϕ	ψ
1	174.2°	130.1°	−69.2°
2	−179.2°	−135.7°	68.7°

Figure 3 | Three-dimensional structure analysis of dimers consisting of *N*-methyl-L/D-alanine residues. a. Solution structures based on MD/DFT which were validated by NMR analysis, and **b.** X-ray crystallographic structures of an LL dimer (left) and DL dimer (right). L-NMA residue and D-NMA residue are colored green and magenta, respectively. Dihedral angles (ϕ, ψ) are described under each structure. In **a**, the Boltzmann weight of the conformer is shown under the most stable structure. Non-neighboring hydrogen pairs with medium–strong NOE signals are shown in yellow and the pairs were connected by red arrows. In **b**, the scale bar is shown under each structure.

Bottom-up design of peptide nanoshapes in water using the oligo-L/D-NMAs. To showcase the utility of L/D-NMA oligomers for the *de novo* design of diverse three-dimensional peptide nanoshapes, we examined hexameric oligomers and generated a catalog of the predicted stable conformations using DFT calculations (Figure 4a). The most stable conformations of L/D-NMA residues were connected,

and the geometry of the generated conformer was optimized by DFT calculations at the B3LYP/6-311G(d) level. The collection of the predicted structures is shown in Figure 4a. The nanoshapes were also plotted on the PMI plot to evaluate the shape diversity (Figure 1c, right, blue circles). The plot demonstrates that combining two amino acid residues, i.e., L- and D-NMA residues, realizes a large nanoshape diversity. The peptide nanoshape diversity approaches the level that is achieved by the same chain length (six amino acid residues) of peptides in miniature proteins (Figure 1a, right).

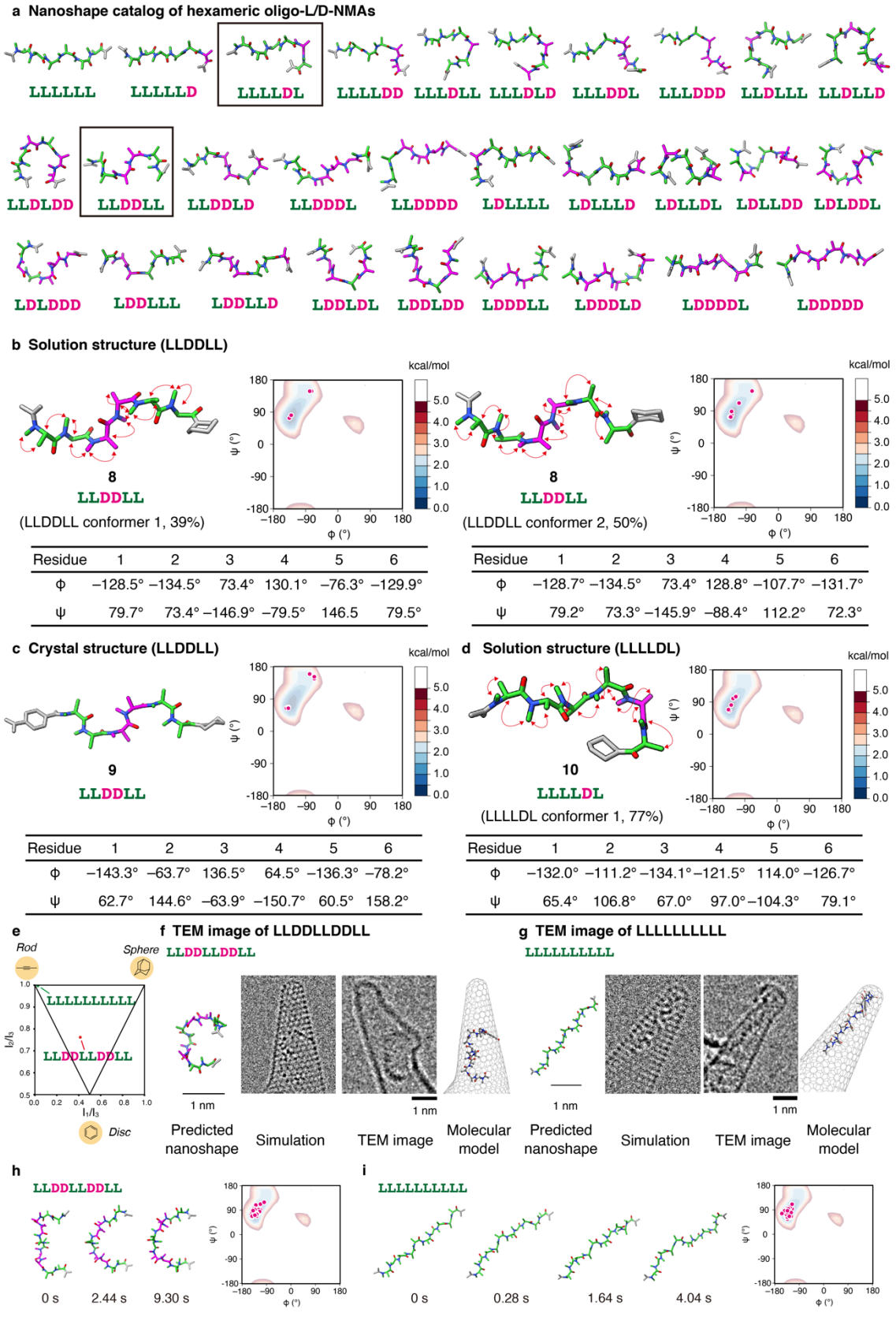

a Nanoshape catalog of hexameric oligo-L/D-NMAs
LLLLLL
LLLLLD
LLLLDL
LLLLDD
LLLDLL
LLLDLD
LLLDDL
LLLDDD
LLDLLL
LLDLLD
LLDLDD
LLDDLL
LLDDLD
LLDDDL
LLDDDD
LDLLLL
LDLLLD
LDLLDL
LDLLDD
LDLDDL
LDLDDD
LDDLLL
LDDLLD
LDDLDL
LDDLDD
LDDDLL
LDDDLD
LDDDDL
LDDDDD
b Solution structure (LLDDLL)
8
LLDDLL
(LLDDLL conformer 1, 39%)
kcal/mol
ψ (°)
Φ (°)
8
LLDDLL
(LLDDLL conformer 2, 50%)
Residue 1 2 3 4 5 6
Φ −128.5° −134.5° 73.4° 130.1° −76.3° −129.9°
ψ 79.7° 73.4° −146.9° −79.5° 146.5 79.5°
Residue 1 2 3 4 5 6
Φ −128.7° −134.5° 73.4° 128.8° −107.7° −131.7°
ψ 79.2° 73.3° −145.9° −88.4° 112.2° 72.3°
c Crystal structure (LLDDLL)
9
LLDDLL
d Solution structure (LLLLDL)
10
LLLLDL
(LLLLDL conformer 1, 77%)
Residue 1 2 3 4 5 6
Φ −143.3° −63.7° 136.5° 64.5° −136.3° −78.2°
ψ 62.7° 144.6° −63.9° −150.7° 60.5° 158.2°
Residue 1 2 3 4 5 6
Φ −132.0° −111.2° −134.1° −121.5° 114.0° −126.7°
ψ 65.4° 106.8° 67.0° 97.0° −104.3° 79.1°
e
Rod
Sphere
Disc
LLLLLLLLLL
LLDDLLDDLL
I2/I3
I1/I3
f TEM image of LLDDLLDDLL
g TEM image of LLLLLLLLLL
1 nm
Predicted nanoshape
Simulation
TEM image
Molecular model
h
LLDDLLDDLL
0 s
2.44 s
9.30 s
i
LLLLLLLLLL
0 s
0.28 s
1.64 s
4.04 s

Figure 4. Diverse peptide nanoshapes realized by the oligo-L/D-NMAs. **a.** A peptide nanoshape catalog of hexameric L/D-NMA oligomers. The stable conformations of L-NMA and D-NMA residues were connected with *trans* amides, and the resulting structures were optimized by DFT calculations to generate the model structures. Among the possible 64 diastereomers, six stereoisomers that have intramolecular steric crashes are omitted. Considering that the remained 58 stereoisomers are 29 enantiomer pairs, only 29 diastereomers that are not enantiomers with each other are shown in the figure. The hexamers that were experimentally examined are highlighted by black squares. **b.** Solution structures of a hexameric oligomer LLDDLL. The two most stable conformers from McMD/DFT simulations are shown. **c.** A crystal structure of LLDDLL obtained by XFEL crystallography[27]. **d.** A solution structure of LLLLDL. A representative conformation in the highest population cluster from the McMD simulations is shown. For **b** and **d**, the non-neighboring hydrogen pairs with significant NOE signals are indicated by red arrows. For **b**–**d**, hydrogens are omitted for clarity. For **b**–**d**, the dihedral angle (ϕ, ψ) pair of each NMA residue is plotted on the Ramachandran-type plot of L-NMA residue on the right using a magenta circle. To discuss L- and D-NMA residues on the same plot, the sign of each dihedral angle value of the D-NMA residues was reserved, and then the values were plotted. **e.** The PMI plot of the predicted nanoshape of LLDDLLDDLL (red circle) and LLLLLLLLLL (green circle). **f**–**g.** A SMART-EM analysis taken at 50 fps and an electron dose rate of $3.8-4.9 \times 10^6$ e^- nm^{-2} s^{-1} of the LLDDLLDDLL decamer (**f**) and the LLLLLLLLLL decamer (**g**). A stable conformation predicted from the shapes of the L/D-NMA residue is shown (left). A simulated TEM image (second from left). The TEM image of the oligomer on an amino-CNT (second from right). A molecular model (right). **h**–**i.** The conformations of LLDDLLDDLL (**h**) and LLLLLLLLLL (**i**) from MD simulations that match the observed TEM images. The time from the beginning of the observation is shown below the conformation. Dihedral angles (ϕ, ψ) of each L/D-NMA residue of the observed conformers are plotted on the Ramachandran-type plot of L-NMA as magenta circles. To discuss L- and D-NMA residues on the same plot, the sign of each dihedral angle value of the D-NMA residues was reserved, and then the values were plotted.

To validate the predicted nanoshapes, stable conformations of one of the oligomers (LLDDLL) were evaluated by a combination of computation and experiment. We first conducted McMD of the oligomer (**8**) to systematically explore the allowed conformational space. The generated conformers at 300 K were clustered and the representative conformer of each cluster was reoptimized by DFT calculations (Table S9). The dihedral angles (ϕ, ψ) of all the NMA residues in the conformer with the highest population in the McMD simulation were within 2.0 kcal/mol on the Ramachandran-type plots (Figure 4b). The Boltzmann weight of the conformer that represents the highest population cluster (LLDDLL conformer 1) was 39%. Based on the DFT calculations, the Boltzmann weight of another conformer (LLDDLL conformer 2) was 50%. Most of the dihedral angles of the L/D-NMA residues in the two conformers were close to the lowest energy point on the Ramachandran-type plot. The dihedral angle (ϕ, ψ) of the 3rd and 5th NMA residues of conformer 1 and 3rd NMA residue of conformer 2 are slightly off from the lowest energy point on the Ramachandran plot, but still the energies are within 2 kcal/mol from the lowest energy point. The two conformers together, the Boltzmann weight is 89%. This result suggests that the NMA hexamer majorly forms the conformation determined by the low-energy conformations of the building blocks and the nanoshape has a small conformational freedom, i.e., each NMA residue can change the conformation within the narrow low-energy region of the Ramachandran-type plot around the lowest-energy point.

To experimentally validate the simulation result, we have conducted NMR measurements of the oligomers in water (Figure S13–15 and Table S10–11). Similar to the dimers, the dominant set of peaks corresponds to a conformational state with which all the amides are in the *trans* configuration according to the NOE signals (Figure S15 and Table S10). This is consistent with the McMD result. As with the monomer and dimers, two kinds of hydrogen pairs are in close proximities on the stable conformer: 1. intra-residual *N*-methyl protons and β-protons; 2. α-proton and *N*-methyl protons of the following NMA residue or C-terminal structure. These NOE signals are consistent with the two conformers (LLDDLL conformers 1 and 2) (Figure 4b and Table S10).

We also conducted X-ray crystallographic studies to obtain atomic resolution snapshots of the oligomer structures. The LLDDLL oligomers with an *N*-terminal *p*-nitrobenzoyl group (**9**) yielded only tiny crystals that were too small for conventional single-crystal X-ray diffraction. Instead, still diffraction patterns were collected using high-intensity X-ray free electron lasers (XFEL) and the crystal structure was determined from the diffraction data by a serial XFEL crystallographic technique for smaller molecules recently introduced[27] (Figure 4c and Table S12). The overall nanoshape of the oligomer is close to the nanoshape in the catalog and the solution structure. The dihedral angles (ϕ, ψ) of 1st, 3rd, and 5th NMA residues are close to the lowest energy point on the Ramachandran-type plot, and 2nd, 4th, and 6th NMA residues are slightly off from the lowest energy point. The dihedral angles of the 2nd, 4th, and 6th NMA residues in the crystal structure resemble the dihedral angles of the 3rd NMA residue of the LLDDLL conformer 2 from the McMD simulations. This result combined with

the McMD/DFT calculations and NMR measurements implies the time-included depiction of the conformational state of the oligomer: the oligomer majorly forms the conformational state consisting of the lowest-energy conformations of the building blocks and the nanoshape has a small conformational freedom around the low-energy area on the Ramachandran-type plot.

To demonstrate the nanoshape diversity realized by the oligo-L/D-NMA, another diastereomer, LLLLDL (**10**), was synthesized and analyzed by McMD/DFT calculations and NMR (Figure 4d). The molecule is predicted to form a linear structure with a bent structure at the C-terminus (Figure 4a). The oligomer was first analyzed by McMD simulations to obtain the allowed conformational space of the oligomer. The generated conformers at 300 K were clustered and the representative conformer of each cluster was reoptimized by DFT calculations (Table S9). As a result of the calculations, the Boltzmann weight of the conformer that represents the highest population cluster was 77%, which suggests the oligomer predominantly forms a single conformer in water. The dominant conformer was found to have the dihedral angles (ϕ, ψ) of around (–135°, 75°) for L monomer and (135°, –75°) for D monomer. Then, the simulation result was validated by NMR measurements (Figure S16–18 and Table S13–14). NOE signals of the oligo-NMA were consistent with the major conformer from the calculations (Table S13). This result further confirmed the conformational rigidity of the oligo-L/D-NMA. The results strongly support that the three-dimensional nanoshapes of the oligopeptides can be designed with subnanometer conformational control by simply connecting the stable conformations of the building blocks.

Cinematographic microscopic imaging of structural dynamics of decamers. The foregoing experiments on oligopeptides consisting of two and six L/D-NMA residues provided strong enough evidence of the validity of our *de novo* nanoshape-design strategy. However, neither the NMR, crystallographic, nor computational method can provide solid structural evidence on larger, flexible, and non-crystalline oligomers with huge conformational possibilities. Microscopic methods capable of sub-angstrom and sub-millisecond imaging over a long period time should serve the purpose of probing this challenging issue. Thus, we examined the potential of the single-molecule atomic-resolution time-resolved electron microscopy (SMART-EM), which recently elucidated the dynamic structures of the oligomers of a lipopeptide antibiotic, daptomycin[28]. The key feature of the method includes the chemical ligation of a peptide onto aminated conical carbon nanotubes (amino-CNTs), and cinematographic recording of the dynamic motions of peptide chains over minutes — the timeframe unavailable to molecular dynamics simulation.

We examined the structures of oligomers consisting of ten L- and D-NMA residues, which is equal to the minimal amino acid residues for miniature proteins. The PMI plot of the oligomers indicates the shape diversity is further expanded and covers a wider area (Figure 1c, right, red circles). We chose two oligomers, LLDDLLDDLL and LLLLLLLLLL, that locate at different regions on the PMI plot.

The model structure of LLDDLLDDLL combines rod-, disc-, and sphere-like nanoshape and locates around the center of the PMI plot (Figure 4e, red circle) while the model structure of LLLLLLLLLL is a rod-like nanoshape and locates at the top-left of the plot (Figure 4e, green circle). DFT calculations suggest a nanoshape of the alphabet "C" as the most stable conformation of the LLDDLLDDLL decamer, and a linear nanoshape for the LLLLLLLLLL decamer (Figure 4f, "Predicted nanoshape"). The oligomer was synthesized with a C-terminal amide and conjugated to amino-CNTs with the aid of potassium acyltrifluoroborate (KAT) chemistry as previously described[28]. Since the TEM image contrast strongly depends on the atomic number of the atoms being imaged[29], we can visualize selectively the main chain of the peptide oligomers, especially, by using a slightly larger defocus value (see SI). As shown in Figure 4f "TEM image", we indeed found that the LLDDLLDDLL decamer persistently took the nanoshape of "C", which matches a simulated image based on DFT calculation (Figure 4f, "Simulation" and "Molecular model"). In sharp contrast, the LLLLLLLLLL decamer persistently took a linear nanoshape, which matches a simulated image based on DFT calculation (Figure 4g, "Simulation" and "Molecular model").

Cinematographic atomic resolution imaging provides further structural information unavailable from the still pictures shown in Figures 4f and 4g. During the SMART-EM observation for 10 s at 298 K at a frame rate of 50 frames per second (fps, 20.0 ms frame^{-1}), the "C" or the linear nanoshape of the oligomers was maintained most of the time, but the nanoshape fluctuated to some degree (Figure S19) indicating that minor conformations are populated occasionally. To assess whether the observed conformational freedom of the oligomer can be explained by the conformational freedom of the L/D-NMA residues on the Ramachandran-type plots, we conducted MD simulations *in vacuo* at 200 and 300 K for 500 ns starting from the DFT-derived stable conformation as the initial conformation. As found by the EM imaging, the MD data also indicated that alphabet "C"-like structures for the LLDDLLDDLL decamer and the linear structures for the LLLLLLLLLL decamer are the predominant conformers. The minor conformer results from the conformational fluctuation of L/D-NMA residues reflected in the low-energy region on the Ramachandran-type plots. We found a variety of conformers of the two decamers interchanging at the beginning of the SMART-EM observation, and could assign the structure to most of them through comparison with the conformers found in the MD trajectories (Figure 4h–i, and Figure S19–20). This agreement suggests that the EM observed conformational fluctuation falls into the range expected from the conformational fluctuation of the monomeric NMA residue at the low-energy region around (ϕ, ψ) = (–135°, 75°) on the Ramachandran-type plots. The dihedral angles (ϕ, ψ) of the observed conformers were restricted in a small region on the Ramachandran-type plot of the L/D-NMA monomer that is within 2.0 kcal/mol from the lowest energy point (Figure 4h–i, right). The result is also consistent with the NMR analysis of the solution structure of the hexamers that the overall nanoshape of an oligo-L/D-NMA is maintained in solution with a small degree of rotational freedom as assessed from the low-energy area on the Ramachandran-type

plots. The EM observation has also given evidence that the combinatorial usage of the L- and D-NMA realizes the expansion of the peptide nanoshape diversity that is produced from the bottom-up designs using conformationally-fixed amino acid residues (Figure 1c and Figure 4e).

Conclusions

In this study, we have shown that a rational combination of L-NMA residues and D-NMA residues allows us to *de novo* design a wide range of three-dimensional peptide nanoshapes with sub-nanometer conformational control. We found that a single NMA residue predominantly exists as a single stable conformer in water and tends to maintain this conformation even when incorporated into an oligomeric sequence. This observation led us to propose a per-residue bottom-up design of peptide nanoshapes by choice of appropriate combinations of L- and D-NMA residues incorporated into an oligomeric sequence.

The combinatorial usage of the two NMA residues with fixed conformations, i.e., L-NMA and D-NMA residues, has enabled the bottom-up design of diverse peptide nanoshapes in water (Figure 4). The *n*-mer oligopeptides consisting of L/D-NMAs can potentially realize 2^n different nanoshapes. As demonstrated by the PMI plot, the bottom-up design realizes a large nanoshape diversity (Figure 1c), which approaches the diversity of the peptide nanoshapes realized by miniature proteins. The bottom-up approach for the peptide nanoshape design is more direct and straightforward compared to the design approach in proteins and foldamers[30–32] which are globally stabilized by a network of inter-residual interactions. The introduction of additional building blocks, such as a proline residue and *N*-substituted β-amino acid residues[33], would further increase the peptide nanoshape diversity in the future. It would realize the peptide nanoshape diversity equal to or even larger than that realized by proteins.

In addition to the facile nanoshape design in a subnanometer scale, oligomers of *N*-methylalanine residues have another advantage in that various functional groups can be easily installed both at the amide nitrogen and the α-carbon of each building block of the oligomers using a peptide-like solid-phase synthetic method[16,34,35]. The facile design of functionalized peptide nanoshapes with sub-nanometer precision in water is potentially useful for the *de novo* design of biofunctional materials, such as catalysts and drugs.

Acknowledgment

We thank Dr. Takuma Kato for his help in solving the crystal structures, and Drs. Ichiro Inoue and Kensuke Tono for setting up the XFEL data collection system. We also thank Prof. Koji Harano for the helpful discussion in the early stages of this work. The authors acknowledge the One-stop Sharing Facility Center for Future Drug Discoveries at the University of Tokyo for the use of micrOTOF II. A part of the computation was performed using Research Center for Computational Science, Okazaki, Japan (Project: 23-IMS-C071). Another part of the computation was carried out using the facilities of the Supercomputer Center at the Institute for Solid State Physics, University of Tokyo. The XFEL experiments were performed at the BL2 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2022A8025). This work was supported by CREST, Japan Science and Technology Agency, Grant Numbers JPMJCR21N5 (to S.S.), PRESTO, Japan Science and Technology Agency, Grant Number JPMJPR21AF (to J.M.), JST-Mirai Program (Grant Number JPMJMI20G5 to K.Y.), Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP22ama121006 (to K.Y., S. M.-Y. and K.T.), JSPS KAKENHI (JP19H05459 to E.N., JP22K14704 to T.N.), a Grant-in-Aid for Scientific Research on Innovative Areas "Materials Science of Meso-Hierarchy" from Japan Society for the Promotion of Science (JSPS) JP23H04874, JSPS KAKENHI (JP20K21494, JP20H03378, and JP22K18374) (to K.T.), and MEXT/JSPS KAKENHI (JP23H02618, JP21H05509, and JP20H03375) (to T.U.). The NMR experiments were performed at NMR Platform of The University of Tokyo supported by MEXT, Japan.

Conflicts of interests

The authors declare the following competing financial interest: The authors (J.M., Y.F., T.W., S.S.) have filed a patent application (PCT/JP2020/27010).

References

1. Wu, T. Te, Johnson, G. & Kabat, E. A. Length distribution of CDRH3 in antibodies. *Proteins Struct. Funct. Bioinforma.* **16**, 1–7 (1993).
2. DeGrado, W. F. & Korendovych, I. V. De novo protein design, a retrospective. *Q. Rev. Biophys.* **53**, e3 (2020).
3. Woolfson, D. N. A Brief History of De Novo Protein Design: Minimal, Rational, and Computational: De novo protein design. *J. Mol. Biol.* **433**, 167160 (2021).
4. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
5. Privalov, P. L. Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 47–69 (1989).
6. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5578–5583 (2001).
7. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9**, 425–430 (2002).
8. Honda, S., Yamasaki, K., Sawada, Y. & Morii, H. 10 Residue folded peptide designed by segment statistics. *Structure* **12**, 1507–1518 (2004).
9. Honda, S. *et al.* Crystal structure of a ten-amino acid protein. *J. Am. Chem. Soc.* **130**, 15327–15331 (2008).
10. Andersen, N. H. *et al.* Minimization and optimization of designed β-hairpin folds. *J. Am. Chem. Soc.* **128**, 6101–6110 (2006).
11. Craven, T. W., Cho, M.-K., Traaseth, N. J., Bonneau, R. & Kirshenbaum, K. A Miniature Protein Stabilized by a Cation−π Interaction Network. *J. Am. Chem. Soc.* **138**, 1543–1550 (2016).
12. Wilhelm, P., Lewandowski, B., Trapp, N. & Wennemers, H. A crystal structure of an oligoproline PPII-helix, at last. *J. Am. Chem. Soc.* **136**, 15829–32 (2014).
13. Tošovská, P. & Arora, P. S. Oligooxopiperazines as nonpeptidic α-helix mimetics. *Org. Lett.* **12**, 1588–1591 (2010).
14. Lao, B. B. *et al.* Rational design of topographical helix mimics as potent inhibitors of protein-protein interactions. *J. Am. Chem. Soc.* **136**, 7877–88 (2014).
15. Lao, B. B. *et al.* In vivo modulation of hypoxia-inducible signaling by topographical helix mimetics. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7531–6 (2014).
16. Morimoto, J. *et al.* A Peptoid with Extended Shape in Water. *J. Am. Chem. Soc.* **141**, 14612–14623 (2019).
17. Yokomine, M., Morimoto, J., Fukuda, Y., Shiratori, Y. & Kuroda, D. Oligo(N-methylalanine) as a Peptide-Based Molecular Scaffold with a Minimal Structure and High Density of Functionalizable Sites Angewandte. *Angew. Chem. Int. Ed.* **61**, e202200119 (2022).

18. Zhang, S., Prabpai, S., Kongsaeree, P. & Arvidsson, P. I. Poly-N-methylated alpha-peptides: synthesis and X-ray structure determination of beta-strand forming foldamers. *Chem. Commun.* 497–499 (2006).
19. Dobitz, S., Aronoff, M. R. & Wennemers, H. Oligoprolines as Molecular Entities for Controlling Distance in Biological and Material Sciences. *Acc. Chem. Res.* **50**, 2420–2428 (2017).
20. Morimoto, J. & Sando, S. Peptoids with substituents on the backbone carbons as conformationally constrained synthetic oligoamides. *J. Synth. Org. Chem.* **78**, 1076–1084 (2020).
21. Sauer, W. H. B. & Schwarz, M. K. Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* **43**, 987–1003 (2003).
22. Koshino, M. *et al.* Imaging of single organic molecules in motion. *Science* **316**, 853 (2007).
23. Nakamura, E. Atomic-Resolution Transmission Electron Microscopic Movies for Study of Organic Molecules, Assemblies, and Reactions: The First 10 Years of Development. *Acc. Chem. Res.* **50**, 1281–1292 (2017).
24. Lovell, S. C. *et al.* Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins* **50**, 437–450 (2003).
25. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
26. Ikebe, J. *et al.* Theory for Trivial Trajectory Parallelization of Multicanonical Molecular Dynamics and Application to a Polypeptide in Water. *J. Comput. Chem.* **32**, 1286–1297 (2011).
27. Takaba, K. *et al.* Structural resolution of a small organic molecule by serial X-ray free-electron laser and electron crystallography. *Nat. Chem.* **15**, 491–497 (2023).
28. Nakamuro, T. *et al.* Time-Resolved Atomistic Imaging and Statistical Analysis of Daptomycin Oligomers with and without Calcium Ions. *J. Am. Chem. Soc.* **144**, 13612–13622 (2022).
29. Xing, J. *et al.* Atomic-number (*Z*)-correlated atomic sizes for deciphering electron microscopic molecular images. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2114432119 (2022).
30. Gellman, S. H. Foldamers: A Manifesto. *Acc. Chem. Res.* **31**, 173–180 (1998).
31. Horne, W. S. & Gellman, S. H. Foldamers with heterogeneous backbones. *Acc. Chem. Res.* **41**, 1399–1408 (2008).
32. Le Bailly, B. A. F. & Clayden, J. Dynamic foldamer chemistry. *Chem. Commun.* **52**, 4852–4863 (2016).
33. Kim, J. *et al.* Residue-based program of a β-peptoid twisted strand shape via a cyclopentane constraint. *Org. Biomol. Chem.* **20**, 6994–7000 (2022).
34. Fukuda, Y. *et al.* Peptoid-based reprogrammable template for cell-permeable inhibitors of protein–protein interactions. *Chem. Sci.* **12**, 13292–13300 (2021).
35. Gao, Y. & Kodadek, T. Synthesis and Screening of Stereochemically Diverse Combinatorial Libraries of Peptide Tertiary Amides. *Chem. Biol.* **20**, 360–369 (2013).