Published in final edited form as:

Nat Nanotechnol. 2017 May ; 12(5): 447–452. doi:10.1038/nnano.2017.15.

Translation of Rod-Like Template Sequences into Homochiral Assemblies of Stacked Helical Oligomers

Quan Gan^{1,3}, Xiang Wang¹, Brice Kauffmann², Frédéric Rosu², Yann Ferrand¹, and Ivan Huc^{1,*}

¹Université de Bordeaux, CNRS, IPB, CBMN – UMR 5248, Institut Européen de Chimie et Biologie, 2 rue Robert Escarpit, 33600 Pessac, France

²Université de Bordeaux, CNRS, INSERM, IECB – UMS3033 – US001, Institut Européen de Chimie et Biologie, 2 rue Robert Escarpit, 33600 Pessac, France

Abstract

At the molecular level, translation refers to the production of a new entity according to a template that has a different chemical composition. In this way, chemical information may be translated from one molecule into another. The process is useful to synthesise structures and thus functions that might be difficult to create otherwise and reaches exquisite levels of efficiency in biological systems as illustrated by protein expression from mRNA templates^{1,2} or by the assembly of the tobacco mosaic virus capsid protein according to the length of its RNA³. In synthetic systems, examples of template-directed syntheses are numerous^{4–6}, but general and versatile schemes in which a non-natural sequence actually encodes the information necessary to produce a different sequence are few, and far from being optimised^{7–10}. Here we show high-fidelity enzyme-free translation of long rod-like alkylcarbamate oligomers into well-defined sequences of stacked helical aromatic oligoamides. Features present in the rods, including the number and distance between carbamate functions and stereogenic centres, template the self-assembly of complementary stacks of helices each having a defined right (*P*) or left (*M*) handedness, length, and single or double helicity. This process enables the production of very large (> 20 kDa) abiotic artificial folded architectures (foldamers¹¹) that may, for example, serve as scaffolds to organize

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to I.H. (i.huc@iecb.u-bordeaux.fr).

³Current address: School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, Wuhan 430074, P. R. China

Author Contributions. Q.G. and X.W. synthesised all new compounds and contributed equally to this work. Q.G. and X.W. carried out solution studies. B.K. collected X-ray data and solved the crystal structures. F.R. carried out ion mobility mass spectrometry measurements. I.H. and Y.F. designed the study. I.H. and Y.F. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing financial interests. The authors declare no competing financial interests.

Data availability statement. Crystallographic data for **2**⊃**16**, **(5)**₂⊃**17**, **(5)**₂⊃**21**, **(5)**₄⊃**22**, **(5)**₆⊃**23**, and **(3.6)**₂⊃**27** have been deposited with the Cambridge Crystallographic Data Centre under reference numbers CCDC-1482309, 1482270, 1482295, 1482290, 1482301, and 1482282 respectively. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk/data_request/cif). Other data that support the findings of this study (spectroscopic or mass spectrometric data) are available from the corresponding author upon reasonable request.

Additional Information. Reprints and permissions information is available online at www.nature.com/reprints.

appended functional features at positions in space defined with atomic precision across nanometric distances.

Foldaxanes are helix-rod host-guest complexes that have been shown to form upon winding aromatic oligomers around *para*-phenylenes,^{12,13} poly-alkylammoniums,¹⁴ or dicarbamates derived from α,ω -diaminoalkanes.^{15,16} In the latter case, arylamide foldamer hosts¹⁷ may be either single¹⁵ or double¹⁶ helical once wound around the rod (Figure 1a, Supplementary Figure 8, Supplementary Table 1). Our set of building blocks comprises short oligomers **1**, **2** and **4** along with newly synthesised longer sequences **3**, **5**, **6** (Figure 1b, Supplementary Figure 1) and rods **7-29** (Figure 1b, Supplementary Figures 2-7). Single strand rods **7-14** served to assess the stability of host-guest complexes as a function of guest length. Association constants in CDCl₃ were systematically determined by ¹H NMR titrations (Supplementary Table 2, Supplementary Figures 9 to 27). In addition, a Van't Hoff plot allowed us to calculate thermodynamic parameters for the formation of (6)₂⊃**15** (the sign ⊃ stands for “include”, meaning **15** is included in (6)₂). The obtained values ($K_a = 590 \text{ L.mol}^{-1}$ at 293 K; $\Delta H = -41.3 \text{ kJ.mol}^{-1}$, $\Delta S = -89.0 \text{ J.K}^{-1}$, Supplementary Figure 28) reveal a strongly enthalpically driven process with a large entropic barrier. Complex formation involves hydrogen bonding between carbonyl groups of the guest and amide protons of 2,6-pyridinedicarboxamide units of the host, located at the two ends of the single helices, or at one end of each strand of the double helices. Consequently, a strict match is required between on one hand the distance between hydrogen bond donors on the helices (*i.e.* the number of helix turns) and on the other hand the length of the alkyl chains connecting hydrogen bond acceptors on the rods.¹⁵ In *single* helical complexes, shrinking the rod by a single CH₂ unit may result in a large or even complete loss of stability. In contrast, screw motions within the *double* helices allow them to adjust their length and to bind to rods differing by 3 to 4 CH₂ units with comparable affinities.¹⁶ This prior knowledge hinted at the possibility of loading numerous single and/or double helices, each having a defined length, on rods possessing the complementary binding stations arranged in a chosen sequence. The sequence would thus template the assembly of an organised multi-helical supramolecular polymer arrangement (Figure 1a). Templatation would be facilitated by the facts that: (i) complex formation does not have to involve a threading mechanism which would require the stepwise introduction of the helices in the order of their arrangement on the rod. Instead, it may also occur *via* an unfolding of the helix and its refolding around the rod, allowing error correction in the process. Thermodynamic products may thus form regardless of the order in which components are assembled; (ii) both single and double helices may slide along the rods and find their best binding station without dissociating.^{15,16}

However, a major unsolved problem was the control of helix handedness. In the absence of chiral information transfer from the rod into a preferred handedness of the helix hosts, and in the absence of end-to-end helix-helix handedness communication between contiguous helices on the rod, the loading of numerous helices would only yield complex mixtures of diastereomeric helix arrangements all having a different spatial organization. Thus, we first aimed to control helix handedness using stereogenic centres on the rod. A series of rods having one helix binding station and bearing one or two terminal chiral groups were

prepared (**16-20**, Supplementary Figure 2) and their host-guest complexes with single or double helical hosts of matching length were investigated (Supplementary Figures 29-34 and 36-37). As illustrated by the emergence of a major species in ^1H NMR spectra (Figure 2a), the two chiral phenethyl groups on guest **16** efficiently induced the handedness of **2** (diastereomeric excess *d.e.* = 93%). A crystal structure of the **2**⊂**16** complex (Figure 2e, Supplementary Figure 42, Supplementary Table 3) allowed us to unambiguously assign *M* helicity as being favoured by (*R,R*) chirality on the rod. This resulted in a strong induced negative circular dichroism (CD) band at 342 nm whilst the (*S,S*) enantiomer induced a positive band (Figure 2c). Diastereoselectivity was moderate when the rod possessed a single chiral group (**3**⊂**18**, see Figure 2d), and also when the host was a double helix ((**5**)₂⊂**17**, see: Figure 2f, Supplementary Figure 43, Supplementary Table 4). Using a chiral terminal naphthylethyl group (**3**⊂**19** and (**6**)₂⊂**20**) gave rise to similar handedness induction. Altogether, the five entries shown in Figure 2d indicate subtle variations of handedness induction as a function of the nature (phenethyl or naphthylethyl) and the number (one or two) of chiral groups, the length of the binding station and the single or double helicity of the host. Although a more thorough study would be required to decipher the interplay between all these parameters, the several cases we report already show efficient handedness induction.

Next, we assessed helix-helix end-to-end handedness communication when multiple helices were loaded on a multistation rod. Initial attempts using the double helix (**4**)₂ as a host showed that its affinity for guests of matching length was too low to achieve quantitative binding of several stations on a single rod at low mM concentration (Supplementary Table 2). Longer oligomer **5** was thus prepared and shown to bind as a double helix to single station rod **21** with a K_a of 1700 M^{-1} in CDCl_3 at 298 K (Figures 3b, 4b, Supplementary Figures 40, 44, Supplementary Table 5). Rods **22-24** possess two, three and five binding stations identical to that of **21** and may in principle bind to two, three or five (**5**)₂ duplexes, respectively. The stations are separated by an ethylene glycol spacer which plays a critical role in the design. Too short a spacer may cause steric hindrance between duplexes (**5**)₂ bound to contiguous stations resulting in negative binding cooperativity and eventually in unoccupied stations. Too long a spacer and the absence of contacts between adjacent duplexes may result in the absence of helix handedness communication, giving rise to complex mixtures of diastereomeric aggregates. Indeed, [5]foldaxane (**5**)₄⊂**22** may exist as a pair of enantiomers *PP/MM* or as a *PM meso* species. Similarly, [7]foldaxane (**5**)₆⊂**23** may exist as three distinct pairs of enantiomers and [11]foldaxane (**5**)₁₀⊂**24** as ten pairs of enantiomers (Supplementary Figure 38). Upon mixing (**5**)₂ with **22**, **23** or **24**, ^1H NMR initially showed complex patterns which simplified over time, eventually resulting in the emergence of a major species. Integration of the rod and helix signals established that the final aggregate stoichiometry corresponds to binding of helices to all stations on each rod (Figures 4c-e, Supplementary Figure S39). In the case of **22** and **23**, the major species was unambiguously identified in the solid state by x-ray crystallography as being racemic homohelical (**5**)₄⊂**22** and (**5**)₆⊂**23**, in which all helices on a given rod have the same handedness (Figure 3d, 3e, Supplementary Figures 45, 46, Supplementary Tables 6, 7). Thus, contacts between helices having a “like” handedness are more favourable than contacts between helices having an “unlike” handedness, resulting in the translation of the

sequence of stations on the rods into a well-defined arrangement of helical aromatic oligoamides. It was assumed that the same rule holds true for $(\mathbf{5})_{10}\supset\mathbf{24}$ of which a molecular model was built showing a 9 nm long structure (Figure 3f, Supplementary Movie 1). Indeed, given the behaviour of $(\mathbf{5})_4\supset\mathbf{22}$ and $(\mathbf{5})_6\supset\mathbf{23}$, we find no reason to suggest the exclusive formation of a conformational isomer of $(\mathbf{5})_{10}\supset\mathbf{24}$ other than the all *P* or all *M*. Single crystals of this very large complex (22 kDa) could also be obtained, but diffraction intensity was too weak to resolve the structure. The crystal structures of $(\mathbf{5})_4\supset\mathbf{22}$ and $(\mathbf{5})_6\supset\mathbf{23}$ revealed direct face-to-face π - π contacts between the helices. At each station, the alkyl moiety of the rod is slightly bent which makes its two ends protrude from the helix at an angle (see Figure 3c). In contrast, the rods adopt a linear conformation at helix-helix junctions where a “like” handedness permits the same tilt angle of each helix with respect to the rod. This alternation of linear and bent segments of the rods gives rise to an undulated shape of the multi-helical aggregates.

The formation of homomeric multi-helical complexes was further studied in solution using double station rod **25** and double helix $(\mathbf{6})_2$ which forms the most stable complexes and was thus the least prone to dissociate upon heating (Supplementary Figure 35). Upon saturation of the two-station rod, two duplexes are bound to form $(\mathbf{6})_4\supset\mathbf{25}$. Aside from the major *PP/MM* species, minor signals assigned to a *PM* diastereomer allowed to calculate a *d.e.* value above 90%. This value did not change significantly with temperature suggesting that more stable *PP/MM* diastereomers are favoured by enthalpy, with negligible involvement of entropy.

Having established rod-helix and helix-helix chiral communication independently, we challenged ourselves to integrate all types of information that a rod may contain to be translated into an organised sequence of helices: chiral groups to control absolute helix handedness, several stations to bind to several helices, different stations to bind to different helices, *e.g.* a single and a double helix (Figure 5e). Rods **26** and **27** possess two distinct and long stations to ensure a high thermodynamic stability of their complexes with the matching helices, separated by an ethylene glycol spacer to induce helix-helix end-to-end handedness communication. A terminal phenethyl group is placed next to a station complementary to the single helix of **3**, and a terminal naphthylethyl group is placed next to a station complementary to the duplex $(\mathbf{6})_2$. Rods **26** and **27** differ from the relative stereochemistry of their two stereogenic centres which may favour the binding of two hosts having opposite handedness in the case of **26** or the same handedness in the case of **27**, thus acting either antagonistically or synergistically with respect to helix-helix handedness communication. Indeed, upon mixing **3** and $(\mathbf{6})_2$ with either **26**-(*R,S*) or **27**-(*R,R*), very different outcomes resulted. With **26**-(*R,S*), a complex NMR spectrum formed corresponding to a mixture of four possible *PP*₂, *MM*₂, *PM*₂ and *MP*₂ ($\mathbf{3}(\mathbf{6})_2\supset\mathbf{26}$) complexes none of which had a strong prevalence (Figure 5c). In contrast, with **27**-(*R,R*), a sharp ¹H NMR spectrum indicated the presence of a dominant species (Figure 5d) whose structure was identified in the solid state as the favoured homohelical arrangement (Figure 5e,f, Supplementary Figure 47, Supplementary Table 8). The structure showed that the helix-helix end-to-end homohelical contact is similar for single helix-double helix communication to that found for double helix-double helix communication. Thus, the stereogenic centres at each end of the

rod as well as the helix-helix contact, although all remote from each other, 18 cooperatively contribute to the emergence of a complex multi-helical aggregate through the translation of information contained on a multistation guest.

The various helices **1-6** at our disposal selectively bind guests according to their length (Supplementary Table 2) in such a way that three host-guest complexes may form with negligible cross-association: $(\mathbf{6})_2\supset\mathbf{13}$, $\mathbf{3}\supset\mathbf{9}$, and $\mathbf{2}\supset\mathbf{7}$ where **13**, **9** and **7** contain undecylene, heptylene and pentylene chains, respectively. For these three complexes, K_a values in CDCl_3 are above $10^4 \text{ L}\cdot\text{mol}^{-1}$ and cross association is inferior to 0.2 %. Complexes between **1** and shorter guests would not cross-associate but their use would be hampered by lower binding. Creating additional orthogonal helix-rod association would entail the preparation of longer helices and longer guests, or the use of other helix-rod binding modes.^{12–14,19} With these complexes in hand, we tested the possibility of increasing information content during multifoldaxane formation. Rods **28** and **29** were prepared which possess three and five helix-binding stations, respectively, of the three kinds mentioned above. The formation of the complexes was monitored by electrospray ion mobility mass spectrometry and confirmed a perfect match (no outlier was detected) between the nature and number of stations on the rod, and the nature and number of helices assembled on it, which supports the conclusion that each helix locates on the station that it is meant to bind, thus resulting in a high fidelity translation of the sequence of binding stations of the rod. Thus, the sequential loading of **3**, $(\mathbf{6})_2$, and **2** yielded each expected intermediate up to final complex $\mathbf{3}\cdot\mathbf{6}_2\cdot\mathbf{2}\supset\mathbf{28}$ (Figure 6a–6c). Because of its large size (24 kDa), the observation of $\mathbf{3}\cdot\mathbf{6}_2\cdot\mathbf{2}\cdot\mathbf{6}_2\cdot\mathbf{3}\supset\mathbf{29}$ required to push instrumentation to its limits but was successful as well (Figure 6d). Provided the concentration is high enough to achieve quantitative binding, the formation of these sequences is quantitative when equilibrium is reached (see Methods section). In addition, ion mobility mass spectrometry gave access to the collision cross-sections of these objects in the gas phase. Calculated and measured values of foldaxanes from rods **21–24**, **28** and **29** are in excellent agreement (Figure 6e,f and Supplementary Figures 40–41). The linear relation between CSSs and molecular mass is consistent with a rigid cylinder model²⁰ and suggests a long persistence length of the multihelix-rod complexes.

In summary, we have established a robust and versatile scheme to produce well-defined homochiral arrangements of helical oligomers wound around multistation rod-like template guest sequences (Supplementary Movies 1–3). Templates having up to ten urethane functions were used to form the largest (up to 24 kDa) abiotic folded architectures known to date. Efficient helix-helix end-to-end handedness communication in discrete aggregates is conceptually different from earlier complexes between polymers and helical¹⁴ or macrocyclic^{21–25} hosts in which no communication between the hosts takes place. It can be related to the piling up of helices into columns with preferred “like”^{26–28} or “unlike”²⁹ contacts as observed in the solid state. As other helix-rod recognition patterns are being identified,¹⁹ template driven helical assemblies may be further expanded and allow the organization in space of various functional groups attached to each helical component. Another extension would consist in optimizing the covalent capture^{30–31} of these non-covalent assemblies to convert the stack of helices into a single molecular polymeric chains.

Methods section

Preparation and purification of homo- and heteromeric stacks of aromatic amide oligomers on oligo-urethane rods

Conditions differ depending on whether or not the rod possesses two bulky stoppers. Rods devoid of at least one stopper may directly thread themselves into the helix cavity through a much faster process than unfolding-refolding around dumbbell-shaped rods. Conversely, the kinetic stability of complexes between helices and dumbbell-shaped rods allow their purification using gel permeation chromatography. CDCl_3 was the preferred medium for easy monitoring by ^1H NMR. Prior to use, CDCl_3 was made free from water and traces of HCl to optimize foldaxane formation, by filtration through a pad of activated basic alumina and distillation from calcium hydride.

Rods devoid of at least one stopper—Aliquots of a stock solution of the urethane guest were added to an NMR tube containing the helical host(s) ($[\text{host}] > 4 \text{ mM}$) in the same solvent. The sample tube was shaken carefully after each addition and ^1H NMR spectra were recorded after ten minutes of incubation. Fast threading of the helix on the rod allowed the immediate observation of the corresponding foldaxane. Upon the addition of a stoichiometric amount of guest, the observed yield is quantitative (yield $> 99\%$) as far as NMR can detect) for affinity constants larger than 10^4 L.mol^{-1} . Note that for multiple station guests and/or chiral guests, a longer time (*i.e.* several hours) is required to reach stereochemical equilibria (*P* vs. *M* helicity).

Rods equipped with two stoppers—In an NMR tube, a stoichiometric amount of helices and dumbbell-shaped guest were dissolved in a minimum of chloroform. Typically a volume of $500 \mu\text{L}$ was used to dissolve the powder and was subsequently reduced to $50 \mu\text{L}$ to reach a concentration above 50 mM . High concentrations enhance the kinetics of binding *via* unfolding-refolding of the helices around the rod. Under these conditions, mass spectrometry shows that each binding station is occupied by the helix that binds it best within one day, but not always with the preferred arrangement of helix handedness, resulting in complicated NMR spectra due to the presence of multiple diastereomers. To reach stereochemical equilibrium, incubation at $>50 \text{ mM}$ for at least five days at 35°C is required for deca-urethanes. The process may be interrupted, and the sample cooled to room temperature and diluted to monitor advancement by ^1H NMR, and resumed if necessary. As for rods devoid of at least one stopper, the yield is the outcome of thermodynamic equilibrium and may be quantitative for affinity constants larger than 10^4 L.mol^{-1} . The final product may be subjected to purification by recycling gel permeation chromatography (GPC), for example to eliminate an excess amount of helices. For those foldaxanes that yield crystals (sometimes directly in the NMR tube by liquid-liquid diffusion after layering with hexane as a precipitant), crystallization also proved to be efficient at purifying the final product from excess helices.

Nuclear Magnetic Resonance

NMR spectra were recorded on 3 different NMR spectrometers: (1) an Avance II NMR spectrometer (Bruker Biospin) with a vertical 7,05T narrow-bore/ultrashield magnet

operating at 300 MHz for ^1H observation and 75 MHz for ^{13}C observation by means of a 5-mm direct BBO H/X probe with Z gradient capabilities; (2) an Avance 400 NMR spectrometer (Bruker Biospin) with a vertical 9.4T narrow-bore/ultrashield magnet operating at 400 MHz for ^1H observation by means of a 5-mm direct QNP $^1\text{H}/^{13}\text{C}/^{31}\text{P}/^{19}\text{F}$ probe with gradient capabilities; (3) an Avance III NMR spectrometer (Bruker Biospin) with a vertical 16.45T narrow-bore/ultrashield magnet operating at 700 MHz for ^1H observation by means of a 5-mm TXI $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ probe with Z gradient capabilities. Chemical shifts are reported in parts per million (ppm, δ) relative to the ^1H residual signal of the deuterated solvent used. ^1H NMR splitting patterns with observed first-order coupling are designated as singlet (s), doublet (d), triplet (t), or quartet (q). Coupling constants (J) are reported in hertz. Data processing was performed with Topspin 2.0 software. Samples were not degassed. CDCl_3 from Eurisotop was used after filtration through an alumina pad followed by a distillation over calcium hydride.

Circular Dichroism

Circular dichroism studies were carried out in distilled chloroform using 2 or 10 mm pathlength cell. Homogenization and sample equilibration were performed after each addition of guest and CD spectra were recorded on a Jasco J-815 spectropolarimeter at 298 K.

Mass Spectrometry

Electrospray ion mobility mass spectrometry (ESI-IMMS) experiments were performed on a Agilent 6560 DTIMS-Q-TOF spectrometer (Agilent Technologies) equipped with a dual-ESI source operated in positive ion mode. The oligomers were analysed at a concentration of 150 μM in CHCl_3 . The syringe pump flow rate was 240 $\mu\text{l hr}^{-1}$. Ion mobility experiments were performed in helium ($P = 3.89$ torr). The trap entrance grid delta was set to 1 V to improve the softness and the funnels radio frequency amplitudes was set to 200 V to improve the transmission of high m/z ions. The data were analysed using Agilent MassHunter software (B.07) and IM-MS Browser B.07.01. Ion mobility give us access to the collisional cross section which is the momentum transfer collision integral averaged over all possible ion collision geometries with the helium gas. Collisional cross sections in helium ($^{\text{DT}}\text{CCS}_{\text{He}}$) of the ions have been obtained using stepped field experiment (5 electric fields).

Molecular modeling and theoretical collision cross-section (CCS calculation)

Starting structures of $(5)_2\text{D}21$, $(5)_4\text{D}22$ and $(5)_6\text{D}23$ were built from their X-ray structures, respectively. Structures of $(5)_{10}\text{D}24$, $3.6_2.2\text{D}28$ and $3_2.6_4.2\text{D}29$ were energy minimised using the MMFFs force field implemented in MacroModel version 8.6 via Maestro version 6.5 (Schrödinger Inc.). All simulations have been carried out using Hypercube 8.0.10 software (hypercube inc.) and the AMBER force field (parm99) has been used. The structures are relaxed and optimised using Polak-Ribiere algorithm until a RMS energy gradient of 0.05 kcal/ Å.mol is obtained. The RMSD on all atoms compared to the crystal structure obtained for $3\text{D}28$ and $3.6_2\text{D}28$ were lower than 0.9 Å which indicate that this force field can perform satisfactorily. The structures are then submitted to unrestrained

molecular dynamics *in vacuo*. Gas phase MD simulations were run in canonical ensemble (at $T = 296$ K) for 1000 ps using an integration time-step of 1 fs. Structures were collected every 25 ps to calculate their CCS. Theoretical CCS calculation have been obtained using the exact hard sphere scattering model (EHSSrot)³² with the Siu's atom parameters³. 840000 trajectories are computed per ion structure and the average value is used for comparison with the experimental CCS.

Crystallography

Crystallographic data for host-guest complexes **2**⊃**16**, **(5)₂**⊃**21** and **(5)₄**⊃**22** were collected at the IECB X-ray facility (UMS CNRS 3033 – INSERM US001) on a Rigaku MM007 HF rotating anode (0.8 kW). Data were collected at the copper $\text{K}\alpha$ wavelength with a partial chi goniometer (AFC11). All data collection strategies were based on Omega scans. The X-ray source is equipped with high flux Osmic Varimax mirrors and a RAPID SPIDER semi-cylindrical image plate detector. The Rigaku CrystalClear suite version 1.36 and 2.1 were used to index, integrate and scale the data with a multi-scan absorption correction.

Crystallographic data for host-guest complexes **(5)₂**⊃**17**, **(5)₆**⊃**23** and **(3.(6)₂)**⊃**27** were collected at the synchrotron ID29 (ESRF) beamline at 0.8 Å. ID29 is a fully automated macromolecular crystallography beamline, intended for high-energy-resolution anomalous dispersion phasing experiments and for high-resolution X-ray diffraction experiments. The beamline is equipped with a MicroDiffractometer (MD2) - which allows to tailor the beam sizes down to 75 (full beam), 50, 30, 20 and 10 microns in diameter. Diffraction data are recorded with a fast readout Pilatus 6M pixel detector (Dectris LTD) that allows data collection with a maximum frame rate of 12 images/s. Data were collected with 0.2° increments (180° or 360° in total) and processed with the XDS package³⁴.

For all six crystal structures the resolution of the experimental data allowed phasing by direct methods or charge flipping. SUPERFLIP³⁵, SHELXT³⁶ and SHELXD³⁷ were run in parallel on all dataset and the best solutions were chosen as a starting point for the modelling and refinement processes. The program that gave the best starting model was found to be SHELXT for **(5)₂**⊃**21**, SHELXD for **2**⊃**16** and **(5)₄**⊃**22**, and SUPERFLIP for **(5)₂**⊃**17**, **(5)₆**⊃**23** and **(3.(6)₂)**⊃**27**.

The starting model of the host-guest molecule **(5)₆**⊃**23** was really poor. To accelerate the initial refinement process a “molecular replacement” approach was combined. Superimposition of a refined model of **(5)₂** in the orthorhombic cell on top of the phasing model helped a lot in getting a full model ready to refine (the superimposition was made “by hand” using the coot software³⁸ and Shelxle software³⁹). All structures were refined by full-matrix least-squares method on F^2 with SHELXL-2014³⁷. As a rule of thumb, every time where disorder could be modelled with partial occupation, it was so. However, more often than not, no simple models of the side chains and solvents can be obtained because multiple positions exists and because complexity is introduced by the swapping of side chains and solvent molecules at some positions, leading to unstable refinement. Introducing too many restraints on the side chains destabilizes the main chain and thus also has to be ruled out. Under such circumstances, an approach that many crystallographers would follow would be to squeeze out the disordered areas that cannot be modelled. However, we did not use this

approach which amounts to removing information from our data and which does affect the main chain displacement parameters. Instead, the side chains are modelled using the EADP constraints (equal atomic displacement parameters) and in most cases refined with isotropic atomic displacement parameters. The outcome is a view of the side chains in an average position which, in our view represents a real information contained in the data. Such modelling is efficient but due to radiation induced damage the position of some atoms from side chains could not be established (only in structure **23D(5)₆**). For all non-hydrogen atoms attempts to introduce anisotropic displacement parameters were made. However, whenever the ellipsoids have adopted unrealistic shape isotropic model was employed. FVAR instruction was used to force some of the isotropic temperature parameters to be equal. In order to model anisotropic displacement parameters ISOR, DELU and RIGU instructions were used. If it was necessary geometry of molecules was improved using DFIX, FRAG, FLAT or AFIX66 and AFIX116 SHELX instructions. Hydrogen atoms were positioned in idealised positions and refined with a riding model, with Uiso constrained to 1.2 Ueq value of the parent atom (1.5 Ueq when CH3). The positions and isotropic displacement parameters of the remaining hydrogen atoms were refined freely.

The final cif file was checked using IUCR's checkcif algorithm. A - Level and B - level alerts were detected. These alerts are unavoidable and inseparably connected with the data quality (weak intensities, moderate resolution), crystal composition (big size of the molecules, large, heavily disordered parts, large solvent content, etc.) and decisions made during data refinement (*i.e.* isotropic displacement parameters for non-H atoms, SQUEEZE40 procedure used only for solvent molecules (and after all attempts of modelling the disordered side chains) and do not reflect errors. All the A and B alerts (except for data resolution) concern the disordered solvent molecules and side chains of the helices but not the main chains or rod-shaped (guests) molecules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Conseil Régional d'Aquitaine, the China Scholarship Council and by the European Research Council under the European Union's Seventh Framework Programme (Grant Agreement No. ERC-2012-AdG-320892). The authors thank Christoph Mueller-Dieckmann (ESRF beamline ID29) for providing beamtime and help during data collection.

References

1. Moore PB, Steitz TA. The structural basis of large ribosomal subunit function. *Annu Rev Biochem.* 2003; 72:813–850. [PubMed: 14527328]
2. Ogle JM, Ramakrishnan V. Structural insights into translational fidelity. *Annu Rev Biochem.* 2005; 74:129–177. [PubMed: 15952884]
3. Namba K, Stubbs G. Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science.* 1986; 231:1401–1406. [PubMed: 3952490]
4. Petitjean A, Nierengarten H, van Dorsselaer A, Lehn J-M. Self-organization of oligomeric helical stacks controlled by substrate binding in a tobacco mosaic virus like self-assembly process. *Angew Chem Int Ed.* 2004; 43:3695–3699.

5. Suzuki K, Sato S, Fujita M. Template synthesis of precisely monodisperse silica nanoparticles within self-assembled organometallic spheres. *Nat Chem.* 2010; 2:25–29. [PubMed: 21124376]
6. Kondratuk DV, et al. Supramolecular nesting of cyclic polymers. *Nat Chem.* 2015; 7:317–322. [PubMed: 25970878]
7. Lewandowski B, et al. Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science.* 2013; 339:189–193. [PubMed: 23307739]
8. He Y, Liu DR. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nat Nanotechnol.* 2010; 5:778–782. [PubMed: 20935654]
9. McKee ML, et al. Multistep DNA-templated reactions for the synthesis of functional sequence controlled oligomers. *Angew Chem Int Ed.* 2010; 49:7948–7951.
10. Meng W, et al. An autonomous molecular assembler for programmable chemical synthesis. *Nat Chem.* 2016; 8:542–548. [PubMed: 27219697]
11. Guichard G, Huc I. Synthetic foldamers. *Chem Commun.* 2011; 47:5933–5941.
12. Tanatani A, Mio MJ, Moore JS. The size-selective synthesis of folded oligomers by dynamic templation. *J Am Chem Soc.* 2002; 124:5934–5935. [PubMed: 12022814]
13. Nishinaga T, Tanatani A, Oh K, Moore JS. Chain length-dependent affinity of helical foldamers for a rodlike guest. *J Am Chem Soc.* 2001; 123:1792–1793. [PubMed: 11456794]
14. Petitjean A, Cuccia LA, Schmutz M, Lehn J-M. Naphthyridine-based helical foldamers and macrocycles: synthesis, cation binding, and supramolecular assemblies. *J Org Chem.* 2008; 73:2481–2495. [PubMed: 18315005]
15. Gan Q, et al. Helix-rod host-guest complexes with shuttling rates much faster than disassembly. *Science.* 2011; 331:1172–1175. [PubMed: 21385710]
16. Ferrand Y, Gan Q, Kauffmann B, Jiang H, Huc I. Template-induced screw motions within an aromatic amide foldamer double helix. *Angew Chem Int Ed.* 2011; 50:7572–7575.
17. Zhang D-W, Zhao X, Hou J-L, Li Z-T. Aromatic amide foldamers: structures, properties, and functions. *Chem Rev.* 2012; 112:5271–5316. [PubMed: 22871167]
18. Clayden J, Lund A, Vallverdú L, Helliwell M. Ultra-remote stereocontrol by conformational communication of information along a carbon chain. *Nature.* 2004; 431:966–971. [PubMed: 15496918]
19. Gan Q, et al. Identification of a foldaxane kinetic byproduct during guest-induced single to double helix conversion. *J Am Chem Soc.* 2012; 134:15656–15659. [PubMed: 22950783]
20. Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJR. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev.* 2010; 39:1633–1655. [PubMed: 20419213]
21. Talotta C, Gaeta C, Qi Z, Schalley CA, Neri P. Pseudorotaxanes with Self-Sorted Sequence and Stereochemical Orientation. *Angew Chem Int Ed.* 2013; 52:7437–7441.
22. Lee S, Chen C-H, Flood AH. A pentagonal cyanostar macrocycle with cyanostilbene CH donors binds anions and forms dialkylphosphate[3]rotaxanes. *Nat Chem.* 2013; 5:704–710. [PubMed: 23881503]
23. Belowich ME, et al. Positive cooperativity in the template-directed synthesis of monodisperse macromolecules. *J Am Chem Soc.* 2012; 134:5243–5261. [PubMed: 22303894]
24. Frampton MJ, Anderson HL. Insulated molecular wires. *Angew Chem Int Ed.* 2007; 46:1028–1064.
25. Harada A, Li J, Kamachi M. The molecular necklace: a rotaxane containing many threaded α -cyclodextrins. *Nature.* 1992; 356:325–327.
26. Sánchez-García D, et al. Nanosized hybrid oligoamide foldamers: aromatic templates for the folding of multiple aliphatic units. *J Am Chem Soc.* 2009; 131:8642–8648. [PubMed: 19530733]
27. Zhao H, et al. Chiral crystallization of aromatic helical foldamers via complementarities in shape and end functionalities. *Chem Sci.* 2012; 3:2042–2046.
28. Nakano K, Oyama H, Nishimura Y, Nakasako S, Nozaki K. λ^5 -phospha[7]helicenes: synthesis, properties, and columnar aggregation with one-way chirality. *Angew Chem Int Ed.* 2012; 51:695–699.
29. Haldar D, Jiang H, Léger JM, Huc I. Double versus single helical structures of oligopyridine-dicarboxamide strands. Part 2: the role of side-chains. *Tetrahedron.* 2007; 63:6322–6330.

30. Clark TD, Ghadiri MR. Supramolecular design by covalent capture. Design of a peptide cylinder via hydrogen-bond-promoted intermolecular olefin metathesis. *J Am Chem Soc.* 1995; 117:12364–12365.
31. Li J, Carnall JMA, Stuart MCA, Otto S. Hydrogel formation upon photoinduced covalent capture of macrocycle stacks from dynamic combinatorial libraries. *Angew Chem Int Ed.* 2011; 50:8384–8386.
32. Shvartsburg AA, Mashkevich SV, Baker ES, Smith RD. Optimization of algorithms for ion mobility calculations. *J Phys Chem A.* 2007; 111:2002–2010. [PubMed: 17300182]
33. Siu C-K, Guo Y, Saminathan IS, Hopkinson AC, Siu KW. Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses. *J Phys Chem B.* 2010; 114:1204–1212. [PubMed: 20039660]
34. Kabsch W. XDS. *Acta Cryst.* 2010; D66:125–132.
35. Palatinus L, Chapuis G. SUPERFLIP – a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions. *J Appl Cryst.* 2007; 40:786–790.
36. Sheldrick GM. SHELXT – Integrated space-group and crystal-structure determination. *Acta Crystallogr.* 2015; A71:3–8.
37. Sheldrick GM. Crystal structure refinement with SHELXL. *Acta Crystallogr.* 2015; C71:3–8.
38. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr.* 2010; D66:486–501.
39. Hübschle CB, Sheldrick GM, Dittrich B. ShelXle: a Qt graphical user interface for SHELXL. *J Appl Cryst.* 2011; 44:1281–1284. [PubMed: 22477785]
40. Spek AL. Single-crystal structure validation with the program PLATON. *J Appl Cryst.* 2003; 36:7–13.

Figure 1. Principle of rod-to-foldaxane translation and molecules used.

(a) Schematic representation of the controlled homochiral assembly of single and double stranded helices of various length onto a long dumbbell shaped template possessing complementary binding stations for each helix and terminal stereogenic centres. (b) Formulae of aromatic oligoamides forming single helical (1-3) and double helical (4-6) host-guest helix-rod complexes. In the absence of guest, all oligomers exist predominantly as double helices at thermodynamic equilibrium in 1 mM CDCl_3 solutions. (c) Formulae of oligocarbamate guest template sequences.

Figure 2. Foldaxane assembly, diastereoselectivity and helix handedness induction.

Part of the 700 MHz ^1H NMR spectra in CDCl_3 at 298K of: (a) $2 \supset 16-(S,S)$ with $[2] = 1 \text{ mM}$ and $[16-(S,S)] = 2 \text{ mM}$; (b) $(6)_2 \supset 20-(R)$ with $[(6)_2] = 1 \text{ mM}$ and $[20-(R)] = 2 \text{ mM}$. Signals of major diastereomeric complexes are denoted in blue whereas minor complexes are denoted in red and pointed to with red arrows. (c) Circular dichroism (CD) spectra of **2** (20 μM , 313K) in CDCl_3 at different time intervals (5 min., 30 min., 60 min., 90 min., 120 min., 150 min. and 210 min.) after the addition of 3 equiv. of **16**-(*S,S*) (blue) or **16**-(*R,R*) (red). (d) *d.e.* values of helix-rod complexes defined as $d.e. = ([M\text{helix}] - [P\text{helix}]) / ([P\text{helix}] + [M\text{helix}])$.

helix]) x 100. The preferred helical sense is indicated next to *d.e.* values. (e) Tube (single helix) and space-filling (rod) representation of the crystal structure of *P*-**2**⊂**16**-(*S,S*). (f) Tube (double helix) and space-filling (rod) representation of the crystal structure of *P*-(**5**)₂⊂**17**-(*S,S*). Isobutoxy side chains and included solvent molecules have been omitted for clarity.

Figure 3. Foldaxane structure elucidation.

Solid state structures in the solid state elucidated by x-ray crystallography of (a) dumbbell rod **21**; (b) side views (left and center) and top view (right) of $(5)_2 \supset 21$. The rod is shown in space-filling representation. Either one or both strands of the double helix are shown in tube representation.; (c) rod **22**; (d) side view (left) and top view (right) of $(5)_4 \supset 22$ in space-filling and tube/space-filling representations for the rod and the double helix, respectively; (e) side views of $(5)_6 \supset 23$. The rod is shown in space-filling representation. The helices are shown in tube (left) or space-filling (right) representations. (f) Energy minimised molecular model using the Merck Molecular Force Field static (MMFFs) of the structure of $(5)_{10} \supset 24$. Isobutoxy side chains and included solvent molecules were omitted for clarity. Only the all *P*

helical isomers are shown. The structures belong to centrosymmetrical space groups and thus also contain the all *M* isomers.

Figure 4. Solution evidence of the uniqueness of self-assembled foldaxanes.

Part of the 700 MHz ^1H NMR in CDCl_3 (0.25 mM) showing the amide and some aromatic resonances of: (a) (5)₂; (b) (5)₂⊂21; (c) (5)₄⊂22; (d) (5)₆⊂23 and (e) (5)₁₀⊂24. Red dashes illustrate the shielding of terminal aromatic amide protons.

Figure 5. Formation of heteromeric stacks of oligomers on heteromeric rods.

Part of the 400 MHz ^1H NMR spectra showing resonances of the pivaloyl end groups of the helices in CDCl_3 at 298 K of: (a) $3 \supset 27-(R,R)$ (1 mM); (b) $(6)_2 \supset 27-(R,R)$ (1 mM); (c) 3 (1 mM) and $(6)_2$ (1 mM) in the presence of rod $26-(R,S)$ (1 equiv.); (d) 3 (1 mM) and $(6)_2$ (1 mM) in the presence of rod $27-(R,R)$ (1 equiv.). White diamonds denote $M-3 \supset 27-(R,R)$ whereas $M-(6)_2 \supset 27-(R,R)$ are marked with black diamonds. Red circles denote the homochiral complex $(M-3.M-(6)_2) \supset 27-(R,R)$. Side view of: (e) the structure of rod $27-(R,R)$; (f) the crystal structure of $(M-3.M-(6)_2) \supset 27-(R,R)$. 3 is shown in grey and $(6)_2$ in purple and light purple. (g-i) Details of the crystal structure of $(M-3.M-(6)_2) \supset 27-(R,R)$ showing different views of the stacked pyridine trimers of 3 (grey tube) and $(6)_2$ (purple tube) which both hydrogen bond to rod $27-(R,R)$ shown in space-filling representation (g) or in tube representation (i, j). The volumes of the pyridine trimers are shown as grey taupe and purple isosurfaces, respectively, in quad mesh representation. Side chains ($\text{O}i\text{Bu}$ groups) and included solvent molecules have been omitted for clarity.

Figure 6. Detection of heteromeric foldaxane sequences on heteromeric rods using ion mobility mass spectrometry.

Electrospray mass spectra (ESI-MS) of: (a) $3\supset 28$; (b) $3.6_2\supset 28$; (c) $3.6_2.2\supset 28$; and (d) $3_2.6_4.2\supset 29$, ([foldaxane] = 150 μM in CHCl_3). The peak annotation $[n]^{z+}$ indicates the number and nature of the foldamer n and the charge z of the detected complexes. Only the expected stoichiometries of the oligomers loaded on each rod are detected. Peaks marked with a star correspond to the expected masses $+625$ from an unidentified contaminant. (e,f) Comparison of collisional cross sections ($^{DT}CCS_{He}$, nm^2) either experimental ($\bullet$) or

calculated assuming a rigid cylinder model (▲): (e) homomeric assemblies $5_2\supset 21$, $5_4\supset 22$, $5_6\supset 23$, $5_{10}\supset 24$; (f) heteromeric assemblies $3\supset 28$, $3.6_2\supset 28$, $3.6_2.2\supset 28$, $3_2.6_4.2\supset 29$.