Translation of rod-like template sequences into homochiral assemblies of stacked helical oligomers

Quan Gan[1][†], Xiang Wang[1], Brice Kauffmann[2], Frédéric Rosu[2], Yann Ferrand[1] and Ivan Huc[1]*

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 to another. The process is useful to synthesize structures and thus functions that might be difficult to create otherwise, and it 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 RNA[3]. 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 optimized[7–10]. Here we show a high-fidelity enzyme-free translation of long rod-like alkylcarbamate oligomers into well-defined sequences of stacked helical aromatic oligoamides. The features present in the rods, which include the number and distance between carbamate functions and stereogenic centres, template the self-assembly of complementary stacks of helices that each have 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 (foldamers[11]) that may, for example, serve as scaffolds to organize 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 on winding aromatic oligomers around *para*-phenylenes[12,13], polyalkylammoniums[14] or dicarbamates derived from α,ω-diaminoalkanes[15,16]. In the latter case, arylamide foldamer hosts[17] may be either single[15] or double[16] helical once wound around the rod (Fig. 1a, Supplementary Fig. 8 and Supplementary Table 1). Our set of building blocks comprises the short oligomers **1**, **2** and **4** along with newly synthesized longer sequences **3**, **5** and **6** (Fig. 1b and Supplementary Fig. 1) and rods **7–29** (Fig. 1c and Supplementary Figs 2–7). Single-station rods **7–14** served to assess the stability of host–guest complexes as a function of guest length. Association constants in CDCl$_3$ were systematically determined by ^{1}H NMR titrations (Supplementary Figs 9–27 and Supplementary Table 2). 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', which means **15** is included in (**6**)$_2$). The obtained values (K_a = 590 l mol^{-1} at 293 K, $\Delta H = -41.3$ kJ mol^{-1}, $\Delta S = -89.0$ J K^{-1} (Supplementary Fig. 28)) reveal a strong enthalpically driven process with a large entropic barrier. The complex formation involves hydrogen bonding between carbonyl groups of the guest and amide protons of the 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 the one hand, the distance between hydrogen-bond donors on the helices (that is, the number of helix turns) and, on the other hand, the length of the alkyl chains that connect hydrogen-bond acceptors on the rods[15]. In single helical complexes, shrinking the rod by a single CH$_2$ 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 that differ by 3–4 CH$_2$ units with comparable affinities[16]. This prior knowledge hinted at the possibility of loading numerous single and/or double helices, each of a defined length, on rods that possess the complementary binding stations arranged in a chosen sequence. The sequence would thus template the assembly of an organized multihelical supramolecular polymer arrangement (Fig. 1a). Templation would be facilitated by the facts that: (1) complex formation does not have to involve a threading mechanism, which requires the stepwise introduction of the helices in the order of their arrangement on the rod, as it may also occur via an unfolding of the helix and its refolding around the rod, which allows error correction in the process—thermodynamic products may thus form regardless of the order in which components are assembled; (2) 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 an 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 with a different spatial organization. Thus, we first aimed to control helix handedness using stereogenic centres on the rod. A series of rods with one helix-binding station and bearing one or two terminal chiral groups was prepared (**16–20** (Supplementary Fig. 2)) and their host–guest complexes with single or double helical hosts of matching length were investigated (Supplementary Figs 29–34 and 36–37). As illustrated by the emergence of a major species in the ^{1}H NMR spectra (Fig. 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 (Fig. 2e, Supplementary Fig. 42 and Supplementary Table 3) allowed us to unambiguously assign M helicity as being favoured by a (R,R) chirality on the rod. This resulted in a strong induced negative circular dichroism (CD) band at 342 nm, whereas the (S,S) enantiomer induced a positive band (Fig. 2c). Diastereoselectivity was moderate when the rod possessed a single chiral group (**3** ⊃ **18** (Fig. 2d)) and also when the host was a double helix ((**5**)$_2$ ⊃ **17** (Fig. 2f, Supplementary Fig. 43 and Supplementary Table 4)). Using a chiral terminal naphthylethyl

[1]Université de Bordeaux, CNRS, IPB, CBMN – UMR 5248, Institut Européen de Chimie et Biologie, 2 rue Robert Escarpit, 33600 Pessac, France. [2]Université de Bordeaux, CNRS, INSERM, IECB – UMS3033 – US001, Institut Européen de Chimie et Biologie, 2 rue Robert Escarpit, 33600 Pessac, France. [†]Present address: School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. *e-mail: i.huc@iecb.u-bordeaux.fr
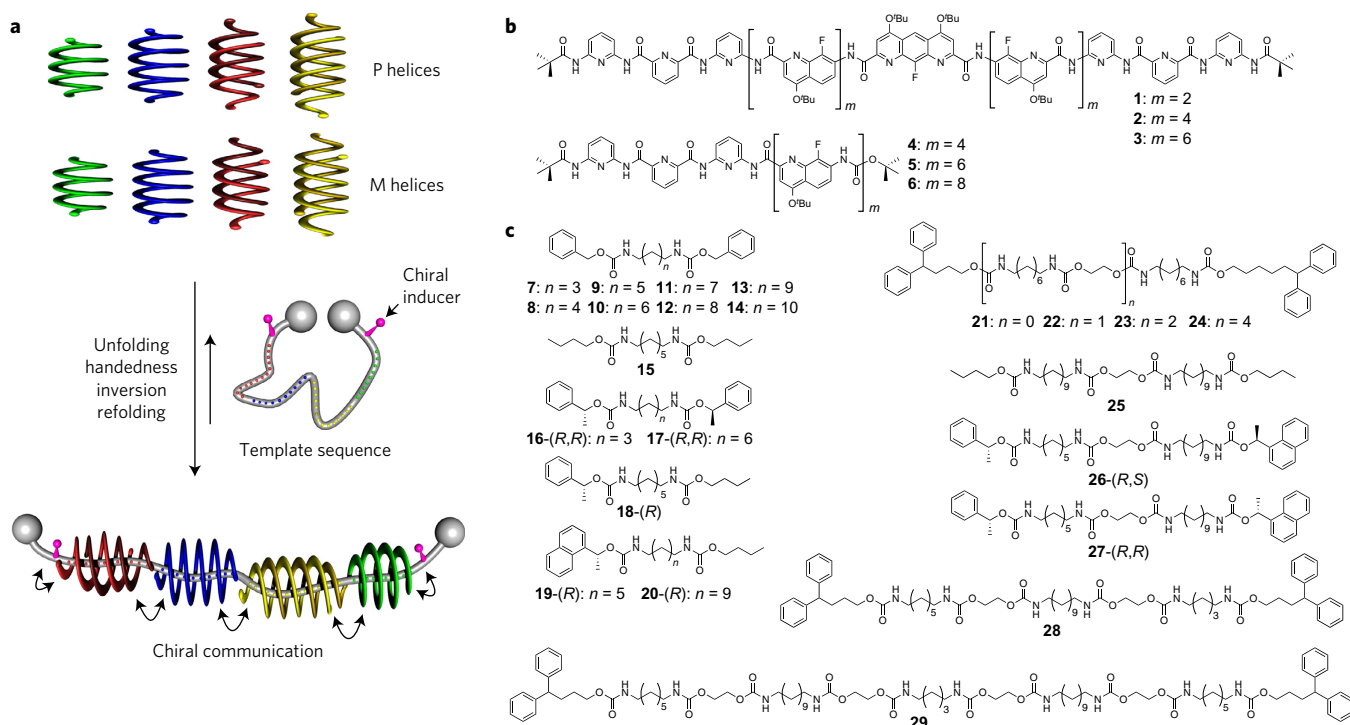
Figure 1 | Principle of rod-to-foldaxane translation and molecules used. a, Schematic representation of the controlled homochiral assembly of single- and double-stranded helices of various lengths onto a long dumbbell-shaped template that possesses complementary binding stations for each helix and terminal stereogenic centres. **b**, Formulae of aromatic oligoamides that form single helical (**1–3**) and double helical (**4–6**) host–guest helix–rod complexes. In the absence of a guest, all the oligomers exist predominantly as double helices at thermodynamic equilibrium in 1 mM CDCl₃ solutions. **c**, Formulae of oligocarbamate guest–template sequences.
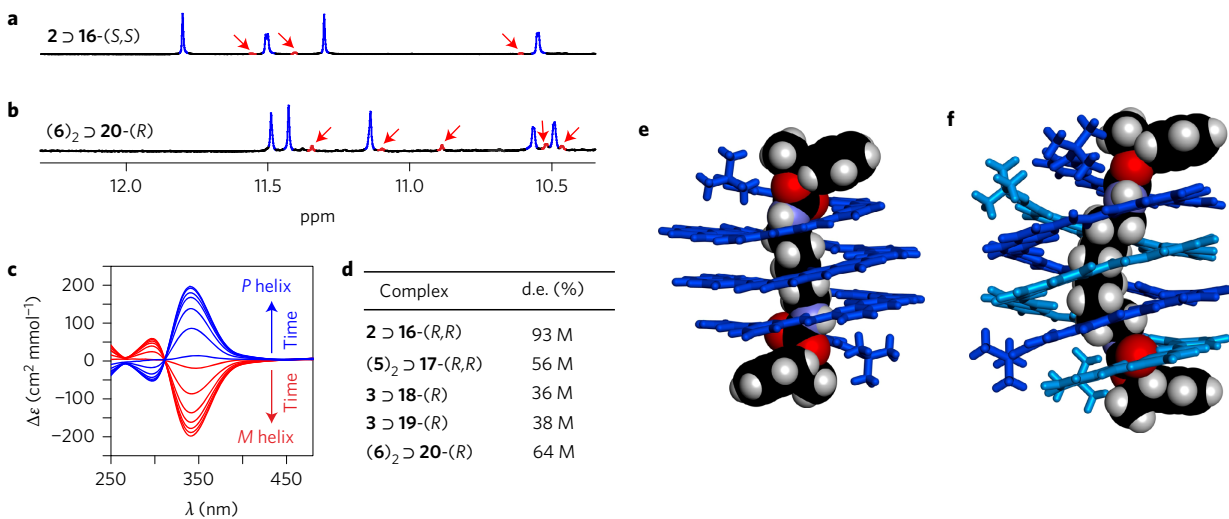
**Figure 2 | Foldaxane assembly, diastereoselectivity and helix handedness induction. a,b**, Part of the 700 MHz ¹H NMR spectra in CDCl₃ at 298 K of **2** ⊃ **16**-(S,S) with [**2**] = 1 mM and [**16**-(S,S)] = 2 mM, and (**6**)₂ ⊃ **20**-(R) with [(**6**)₂] = 1 mM and [**20**-(R)] = 2 mM (**b**). Signals of major diastereomeric complexes are denoted in blue whereas minor complexes are denoted in red and indicated with red arrows. **c**, CD spectra of **2** (20 µM, 313 K) in CDCl₃ at different time intervals (5, 30, 60, 90, 120, 150 and 210 min) after the addition of 3 equiv. **16**-(S,S) (blue) or **16**-(R,R) (red). **d**, d.e. values of helix-rod complexes defined as d.e. = ([M helix] – [P helix])/([P helix] + [M helix]) × 100. The preferred helical sense is indicated next to the 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 are omitted for clarity.

group (**3** ⊃ **19** and (**6**)₂ ⊃ **20**) gave rise to similar handedness induction. Altogether, the five entries shown in Fig. 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 is required to

decipher the interplay between all these parameters, the several cases we report here already show efficient handedness induction.

Next, we assessed the helix–helix end-to-end handedness communication when multiple helices were loaded on a multistation rod. The initial attempts used the double helix (**4**)₂ as a host and showed that its affinity for guests of matching length was too low
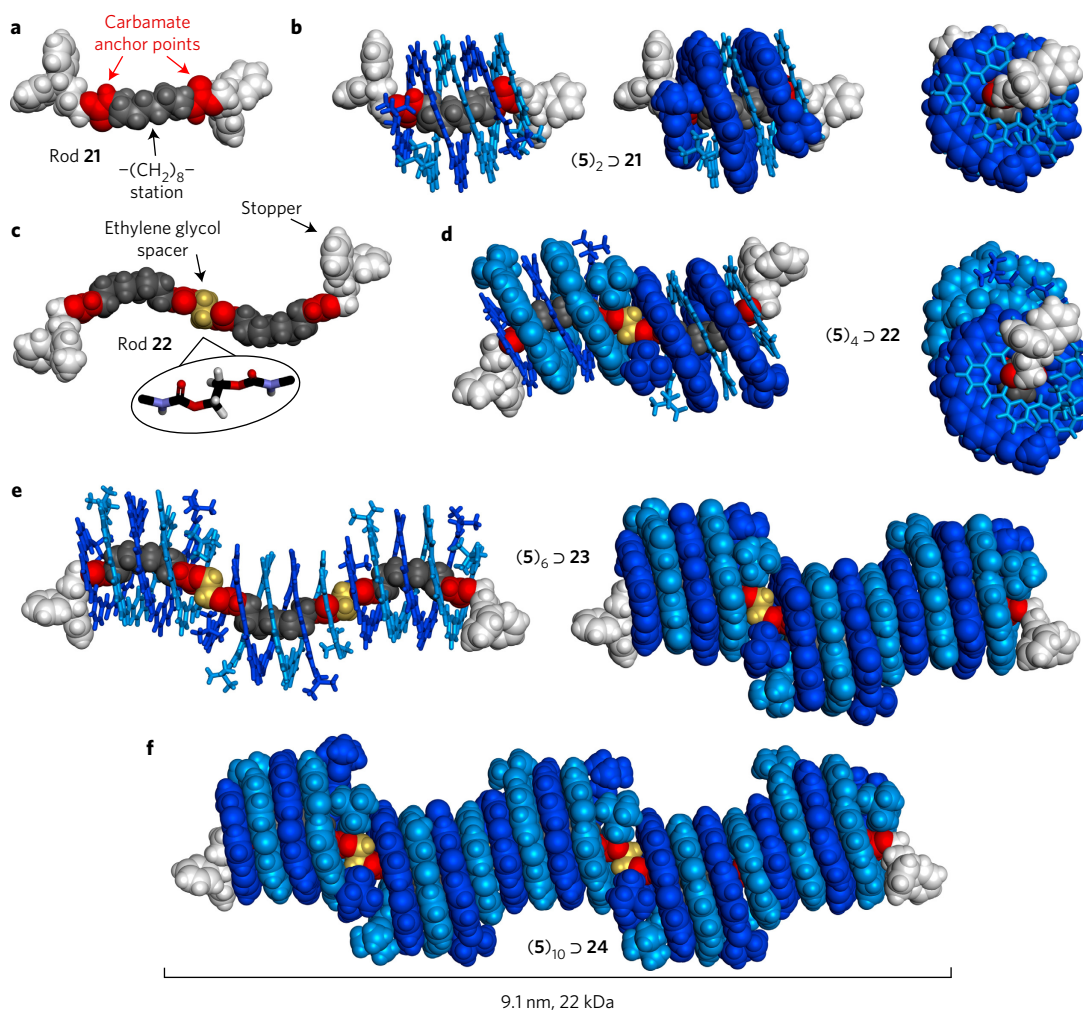
Figure 3 | Foldaxane structure elucidation. a,b, Solid-state structures elucidated by X-ray crystallography of dumbbell rod **21** (**a**), and side views (left and centre) and top view (right) of (**5**)$_2$ ⊃ **21** (**b**). The rod is shown in space-filling representation. Either one or both strands of the double helix are shown in the tube representation. **c,d,** Rod **22** (**c**), and side view (left) and top view (right) of (**5**)$_4$ ⊃ **22** (**d**) in space-filling and tube/space-filling representations for the rod and the double helix, respectively. **e,** Side views of (**5**)$_6$ ⊃ **23**. The rod is shown in a space-filling representation. The helices are shown in tube (left) or space-filling (right) representations. **f,** Energy-minimized molecular model using the MMFFs of the structure of (**5**)$_{10}$ ⊃ **24**. Isobutoxy side chains and included solvent molecules are 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.

to achieve quantitative binding of several stations on a single rod at low millimolar concentrations (Supplementary Table 2). The longer oligomer **5** was thus prepared and shown to bind as a double helix to the single-station rod **21** with a K_a of 1,700 l mol^{-1} in CDCl$_3$ at 298 K (Figs 3b and 4b, Supplementary Figs 40 and 44 and 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, which results in a 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 and give 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 Fig. 38). On mixing (**5**)$_2$ with **22**, **23** or **24**, ^{1}H NMR spectroscopy initially showed complex patterns that simplified over time, and eventually resulted

in the emergence of a major species. Integration of the rod and helix signals established that the final aggregate stoichiometry corresponds to the binding of helices to all the stations on each rod (Fig. 4c–e and Supplementary Fig. 39). For **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 the helices on a given rod have the same handedness (Fig. 3d,e, Supplementary Figs 45 and 46 and Supplementary Tables 6 and 7). Thus, contacts between helices with a 'like' handedness are more favourable than contacts between helices with an 'unlike' handedness, which results 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 (**5**)$_{10}$ ⊃ **24**, of which a molecular model was built that showed a structure 9 nm long (Fig. 3f and Supplementary Movie 1). Indeed, given the behaviour of (**5**)$_4$ ⊃ **22** and (**5**)$_6$ ⊃ **23**, we find no reason to suggest the exclusive formation of a conformational isomer of (**5**)$_{10}$ ⊃ **24** other than all P or all M. Single crystals of this very large complex (22 kDa) could also be obtained, but the diffraction intensity was too weak to resolve the structure. The crystal structures of (**5**)$_4$ ⊃ **22** and (**5**)$_6$ ⊃ **23** revealed direct face-to-face π–π
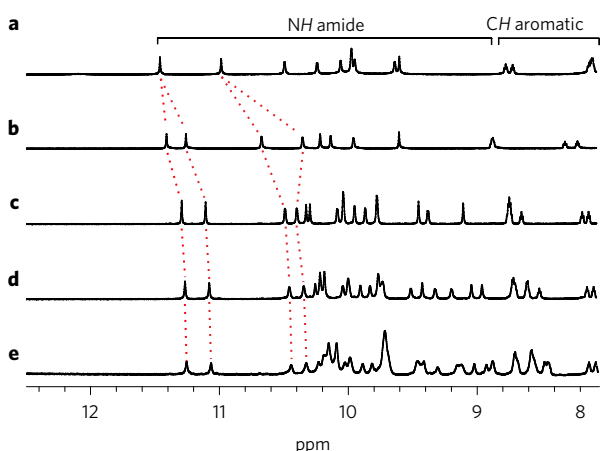
Figure 4 | Solution evidence of the uniqueness of self-assembled foldaxanes. a–e, Parts of the 700 MHz ^{1}H NMR spectra in CDCl$_3$ (0.25 mM) at 298 K that show the amide and some aromatic resonances of $(5)_2$ (**a**), $(5)_2 \supset 21$ (**b**), $(5)_4 \supset 22$ (**c**), $(5)_6 \supset 23$ (**d**) and $(5)_{10} \supset 24$ (**e**). The red dashes illustrate the shielding of terminal aromatic amide protons.

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 Fig. 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 multihelical aggregates.

The formation of homomeric multihelical complexes was further studied in solution using double-station rod **25** and double helix $(6)_2$,

which forms the most-stable complexes and was thus the least prone to dissociate on heating (Supplementary Fig. 35). On saturation of the two-station rod, two duplexes are bound to form $(6)_4 \supset 25$. Aside from the major PP/MM species, minor signals assigned to a PM diastereomer allowed us to calculate a d.e. value of above 90%. This value did not change significantly with temperature, which suggests that the 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 the types of information that a rod may contain to be translated into an organized sequence of helices: chiral groups to control absolute helix handedness, several stations to bind to several helices and different stations to bind to different helices, for example, a single and a double helix (Fig. 5e). Rods **26** and **27** possess two distinct and long stations that ensure the high thermodynamic stability of their complexes with the matching helices, separated by an ethylene glycol spacer to induce a 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 $(6)_2$. Rods **26** and **27** differ from the relative stereochemistry of their two stereogenic centres, which may favour the binding of two hosts with opposite handedness in the case of **26** or the same handedness in the case of **27**, and thus act either antagonistically or synergistically with respect to the helix–helix handedness communication. Indeed, on mixing **3** and $(6)_2$ with either **26**-(R,S) or **27**-(R,R), very different outcomes resulted. With **26**-(R,S), a complex NMR spectrum formed that corresponded to a mixture of four possible P(P)$_2$, M(M)$_2$, P(M)$_2$ and M(P)$_2$ **3**·$(6)_2 \supset 26$ complexes, none of which had a strong prevalence (Fig. 5c). In contrast, with **27**-(R,R), a sharp ^{1}H NMR spectrum
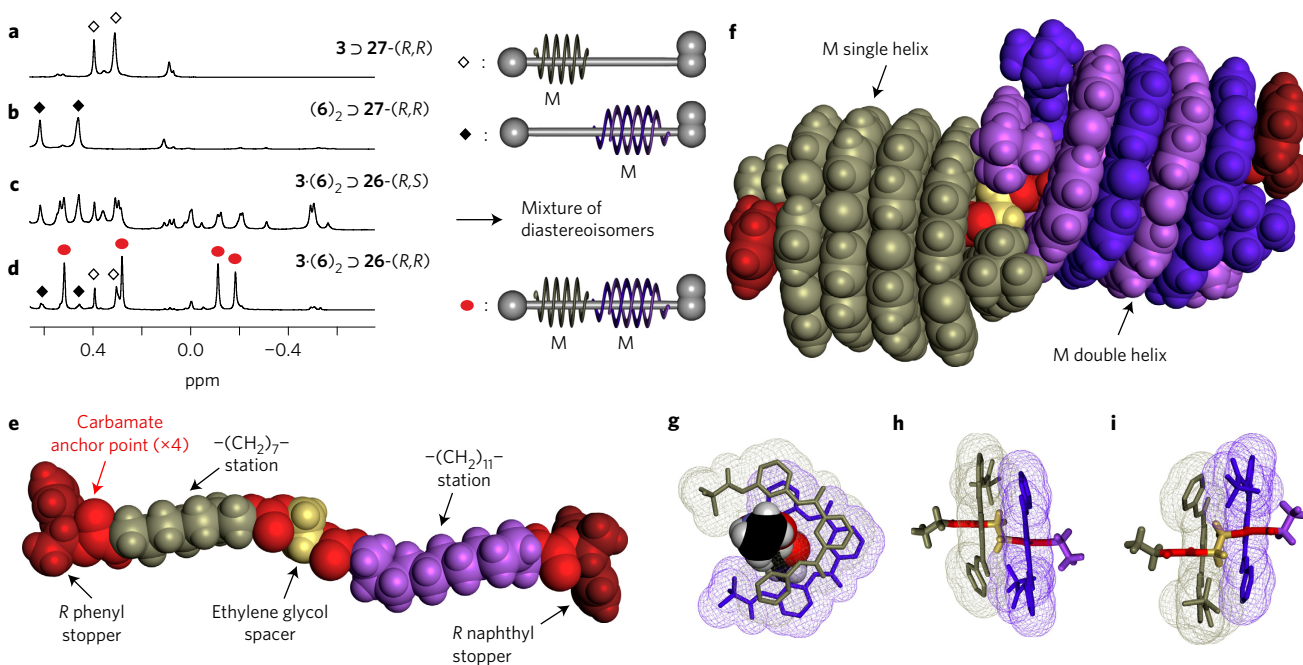
Figure 5 | Formation of heteromeric stacks of oligomers on heteromeric rods. a–d, Parts of the 400 MHz ^{1}H NMR spectra that show resonances of the pivaloyl end groups of the helices in CDCl$_3$ at 298 K of **3** $\supset$ **27**-(R,R) (1 mM) (**a**), $(6)_2 \supset$ **27**-(R,R) (1 mM) (**b**), **3** (1 mM) and $(6)_2$ (1 mM) in the presence of rod **26**-(R,S) (1 equiv.) (**c**) and **3** (1 mM) and $(6)_2$ (1 mM) in the presence of rod **27**-(R,R) (1 equiv.) (**d**). White diamonds, M-**3** $\supset$ **27**-(R,R); black diamonds, M-$(6)_2 \supset$ **27**-(R,R); red circles, homochiral complex M-**3**·M-$(6)_2 \supset$ **27**-(R,R). **e,f**, Side views the structure of rod **27**-(R,R) (**e**) and the crystal structure of M-**3**·M-$(6)_2 \supset$ **27**-(R,R) (**f**). **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) that show 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 a space-filling representation (**g**) or in a tube representation (**h,i**). The volumes of the pyridine trimers are shown as grey taupe and purple isosurfaces, respectively, in a quad-mesh representation. Isobutoxy side chains and included solvent molecules are omitted for clarity.
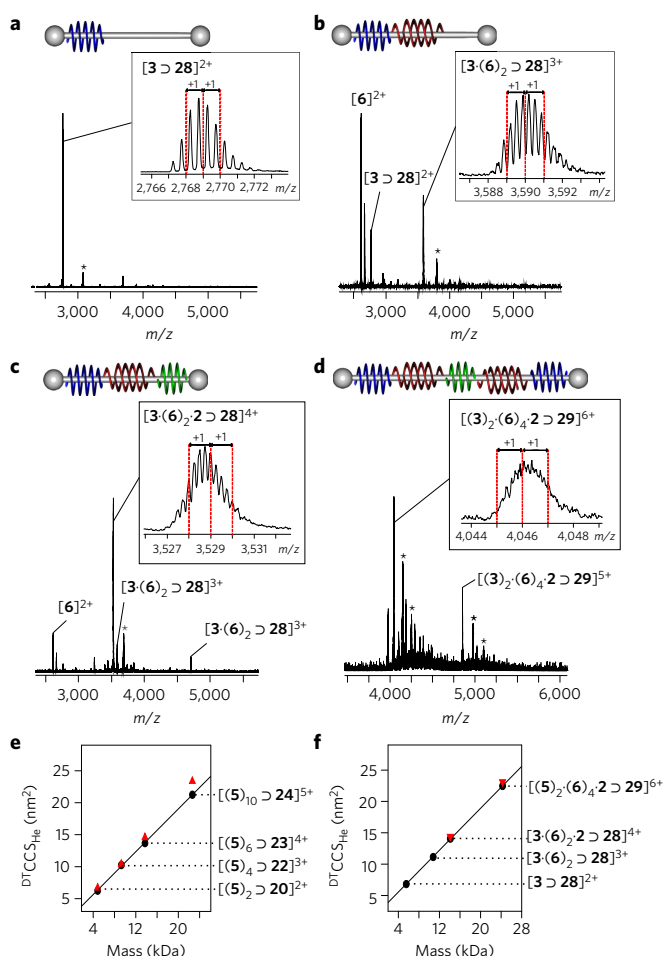
Figure 6 | Detection of heteromeric foldaxane sequences on heteromeric rods using IMMS. a–d, ESI-MS of $3 \supset 28$ (**a**), $3\cdot(6)_2 \supset 28$ (**b**), $3\cdot(6)_2\cdot2 \supset 28$ (**c**) and $(3)_2\cdot(6)_4\cdot2 \supset 29$ (**d**) ([foldaxane] = 150 µM in CHCl$_3$). The peak annotation $[n]^{z+}$ 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 an asterisk correspond to the expected masses +625 from an unidentified contaminant. **e,f**, Comparison of $^{DT}CCS_{He}$ for either experimental (filled black circles) or calculated assuming a rigid cylinder model (filled red triangles) of the homomeric assemblies $(5)_2 \supset 21$, $(5)_4 \supset 22$, $(5)_6 \supset 23$ and $(5)_{10} \supset 24$ (**e**) and the heteromeric assemblies $3 \supset 28$, $3\cdot(6)_2 \supset 28$, $3\cdot(6)_2\cdot2 \supset 28$ and $(3)_2\cdot(6)_4\cdot2 \supset 29$ (**f**).

indicated the presence of a dominant species (Fig. 5d) with a structure identified in the solid state as the favoured homohelical arrangement (Fig. 5e,f, Supplementary Fig. 47 and Supplementary Table 8). The structure showed that the helix–helix end-to-end homohelical contact for single helix–double helix communication is similar 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 multihelical 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: $(6)_2 \supset 13$, $3 \supset 9$ and $2 \supset 7$ where **13**, **9** and **7** contain undecylene, heptylene and pentylene chains, respectively. For these three complexes, K_a values in CDCl$_3$ are above $10^4 \, \text{l mol}^{-1}$ and cross-association is less than 0.2%. Complexes between **1** and

shorter guests would not cross-associate, but their use would be hampered by lower binding. To create 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 the information content during multifoldaxane formation. We prepared rods **28** and **29**, 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 (ESI-IMMS) 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, and thus results in a high-fidelity translation of the sequence of binding stations of the rod. Thus, the sequential loading of **3**, $(6)_2$ and **2** yielded each expected intermediate up to the final complex $3\cdot(6)_2\cdot2 \supset 28$ (Fig. 6a–c). Its large size (24 kDa) meant the observation of $3\cdot(6)_2\cdot2\cdot(6)_2\cdot3 \supset 29$ required the instrumentation to be pushed to its limits, but was successful (Fig. 6d). Provided the concentration is high enough to achieve quantitative binding, the formation of these sequences is quantitative when equilibrium is reached (Methods). In addition, IMMS gave access to the collision cross-sections (CCSs) of these objects in the gas phase. Calculated and measured values of foldaxanes from rods **21–24**, **28** and **29** are in excellent agreement (Fig. 6e,f and Supplementary Figs 40 and 41). The linear relation between CSSs and molecular mass is consistent with a rigid cylinder model[20] 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 with 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 other complexes between polymers and helical[14] or macrocyclic[21–25] hosts in which no communication between the hosts takes place. This 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 expanded further and allow the organization in space of various functional groups attached to each helical component. Another extension could optimize the covalent capture[30,31] of these non-covalent assemblies to convert the stack of helices into single molecular polymeric chains.

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Moore, P. B. & Steitz, T. A. The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.* **72,** 813–850 (2003).
2. Ogle, J. M. & Ramakrishnan, V. Structural insights into translational fidelity. *Annu. Rev. Biochem.* **74,** 129–177 (2005).
3. Namba, K. & Stubbs, G. Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science* **231,** 1401–1406 (1986).
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.* **43,** 3695–3699 (2004).
5. Suzuki, K., Sato, S. & Fujita, M. Template synthesis of precisely monodisperse silica nanoparticles within self-assembled organometallic spheres. *Nat. Chem.* **2,** 25–29 (2010).

6. Kondratuk, D. V. *et al.* Supramolecular nesting of cyclic polymers. *Nat. Chem.* **7,** 317–322 (2015).
7. Lewandowski, B. *et al.* Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **339,** 189–193 (2013).
8. He, Y. & Liu, D. R. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nat. Nanotech.* **5,** 778–782 (2010).
9. McKee, M. L. *et al.* Multistep DNA-templated reactions for the synthesis of functional sequence controlled oligomers. *Angew. Chem. Int. Ed.* **49,** 7948–7951 (2010).
10. Meng, W. *et al.* An autonomous molecular assembler for programmable chemical synthesis. *Nat. Chem.* **8,** 542–548 (2016).
11. Guichard, G. & Huc, I. Synthetic foldamers. *Chem. Commun.* **47,** 5933–5941 (2011).
12. Tanatani, A., Mio, M. J. & Moore, J. S. The size-selective synthesis of folded oligomers by dynamic templation. *J. Am. Chem. Soc.* **124,** 5934–5935 (2002).
13. Nishinaga, T., Tanatani, A., Oh, K. & Moore, J. S. Chain length-dependent affinity of helical foldamers for a rodlike guest. *J. Am. Chem. Soc.* **123,** 1792–1793 (2001).
14. Petitjean, A., Cuccia, L. A., Schmutz, M. & Lehn, J.-M. Naphthyridine-based helical foldamers and macrocycles: synthesis, cation binding, and supramolecular assemblies. *J. Org. Chem.* **73,** 2481–2495 (2008).
15. Gan, Q. *et al.* Helix–rod host–guest complexes with shuttling rates much faster than disassembly. *Science* **331,** 1172–1175 (2011).
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.* **50,** 7572–7575 (2011).
17. Zhang, D.-W., Zhao, X., Hou, J.-L. & Li, Z.-T. Aromatic amide foldamers: structures, properties, and functions. *Chem. Rev.* **112,** 5271–5316 (2012).
18. Clayden, J., Lund, A., Vallverdú, L. & Helliwell, M. Ultra-remote stereocontrol by conformational communication of information along a carbon chain. *Nature* **431,** 966–971 (2004).
19. Gan, Q. *et al.* Identification of a foldaxane kinetic byproduct during guest-induced single to double helix conversion. *J. Am. Chem. Soc.* **134,** 15656–15659 (2012).
20. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39,** 1633–1655 (2010).
21. Talotta, C., Gaeta, C., Qi, Z., Schalley, C. A. & Neri, P. Pseudorotaxanes with self-sorted sequence and stereochemical orientation. *Angew. Chem. Int. Ed.* **52,** 7437–7441 (2013).
22. Lee, S., Chen, C.-H. & Flood, A. H. A pentagonal cyanostar macrocycle with cyanostilbene CH donors binds anions and forms dialkylphosphate[3]rotaxanes. *Nat. Chem.* **5,** 704–710 (2013).
23. Belowich, M. E. *et al.* Positive cooperativity in the template-directed synthesis of monodisperse macromolecules. *J. Am. Chem. Soc.* **134,** 5243–5261 (2012).
24. Frampton, M. J. & Anderson, H. L. Insulated molecular wires. *Angew. Chem. Int. Ed.* **46,** 1028–1064 (2007).
25. Harada, A., Li, J. & Kamachi, M. The molecular necklace: a rotaxane containing many threaded α-cyclodextrins. *Nature* **356,** 325–327 (1992).
26. Sánchez-García, D. *et al.* Nanosized hybrid oligoamide foldamers: aromatic templates for the folding of multiple aliphatic units. *J. Am. Chem. Soc.* **131,** 8642–8648 (2009).
27. Zhao, H. *et al.* Chiral crystallization of aromatic helical foldamers via complementarities in shape and end functionalities. *Chem. Sci.* **3,** 2042–2046 (2012).
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.* **51,** 695–699 (2012).
29. Haldar, D., Jiang, H., Léger, J.-M. & Huc, I. Double versus single helical structures of oligopyridine-dicarboxamide strands. Part 2: the role of side-chains. *Tetrahedron* **63,** 6322–6330 (2007).
30. Clark, T. D. & Ghadiri, M. R. Supramolecular design by covalent capture. Design of a peptide cylinder via hydrogen-bond-promoted intermolecular olefin metathesis. *J. Am. Chem. Soc.* **117,** 12364–12365 (1995).
31. Li, J., Carnall, J. M. A., Stuart, M. C. A. & Otto, S. Hydrogel formation upon photoinduced covalent capture of macrocycle stacks from dynamic combinatorial libraries. *Angew. Chem. Int. Ed.* **50,** 8384–8386 (2011).

Author contributions

Q.G. and X.W synthesised all the 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 the authors discussed the results and commented on the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.H.

Competing financial interests

The authors declare no competing financial interests.

Methods

Preparation and purification of homo- and heteromeric stacks of aromatic amide oligomers on oligourethane rods. Conditions differ depending on whether or not the rod possesses two bulky stoppers. Rods devoid of at least one stopper may thread themselves directly 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 allows their purification using gel-permeation chromatography (GPC). CDCl$_3$ was the preferred medium for easy monitoring by ^{1}H NMR spectroscopy. To optimize the foldaxane formation prior to use, CDCl$_3$ was freed of water and traces of HCl 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 that contained the helical host(s) ([host] > 4 mM) in the same solvent. The sample tube was shaken carefully after each addition and ^{1}H NMR spectra were recorded after 10 min of incubation. Fast threading of the helix on the rod allowed the immediate observation of the corresponding foldaxane. On the addition of a stoichiometric amount of guest, the observed yield is quantitative (yield > 99%, as far as NMR spectroscopy can detect) for affinity constants larger than 10^4 l mol^{-1}. For multiple station guests and/or chiral guests, a longer time (that is, several hours) is required to reach stereochemical equilibria (P versus 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 μl was used to dissolve the powder and subsequently reduced to 50 μl 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, which results in complicated NMR spectra because of the presence of multiple diastereomers. To reach stereochemical equilibrium, incubation at >50 mM for at least five days at 35 °C is required for decaurethanes. The process may be interrupted, the sample cooled to room temperature and diluted to monitor the advancement by ^{1}H NMR spectroscopy, and incubation 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 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.

NMR spectroscopy. NMR spectra were recorded on three different NMR spectrometers: (1) an Avance II NMR spectrometer (Bruker Biospin) with a vertical 7.05 T narrow-bore/ultrashield magnet operating at 300 MHz for ^{1}H NMR observation and 75 MHz for ^{13}C NMR 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 9.4 T narrow-bore/ultrashield magnet operating at 400 MHz for ^{1}H NMR observation by means of a 5 mm direct QNP ^{1}H/^{13}C/^{31}P/^{19}F probe with gradient capabilities; (3) an Avance III NMR spectrometer (Bruker Biospin) with a vertical 16.45 T narrow-bore/ultrashield magnet operating at 700 MHz for ^{1}H NMR observation by means of a 5 mm TXI ^{1}H/^{13}C/^{15}N probe with Z gradient capabilities. Chemical shifts (δ) are reported in ppm relative to the ^{1}H NMR residual signal of the deuterated solvent used. ^{1}H 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 distillation over calcium hydride.

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

Mass spectrometry. ESI-IMMS experiments were performed on an 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 μl h^{-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 funnel radiofrequency amplitude 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 gives 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 ($^{DT}CCS_{He}$) of the ions were obtained using stepped field experiments (five electric fields).

Molecular modelling and theoretical CCS calculations. The starting structures of $(5)_2 \supset 21$, $(5)_4 \supset 22$ and $(5)_6 \supset 23$ were built from their X-ray structures. Structures of $(5)_{10} \supset 24$, $3\cdot(6)_2\cdot2 \supset 28$ and $(3)_2\cdot(6)_4\cdot2 \supset 29$ were energy minimized using the Merck Molecular Force Field static (MMFF) force field implemented in MacroModel version 8.6 via Maestro version 6.5 (Schrödinger). All the simulations were carried out using Hyperche 8.0.10 software (hypercube) and the AMBER force field (parm99) was used. The structures were relaxed and optimized using the Polak–Ribiere algorithm until a root mean square (r.m.s.) energy gradient of 0.05 kcal A^{-1} mol^{-1} is obtained. The r.m.s. deviations on all the atoms compared with the crystal structures obtained for $3 \supset 28$ and $3\cdot(6)_2 \supset 28$ were lower than 0.9 Å, which indicates that this force field can perform satisfactorily. The structures were then submitted to unrestrained molecular dynamics (MD) *in vacuo*. Gas-phase MD simulations were run in a canonical ensemble (at T = 296 K) for 1,000 ps using an integration time step of 1 fs. Structures were collected every 25 ps to calculate their CCSs. Theoretical CCS calculations were obtained using the exact hard-sphere scattering model (EHSSrot)[32] with Siu's atom parameters[33]. For each ion structure, 840,000 trajectories were computed and the average value was used for comparison with the experimental CCS.

Crystallography. Crystallographic data for host–guest complexes $2 \supset 16$, $(5)_2 \supset 21$ and $(5)_4 \supset 22$ were collected at the Institut Européen de Chimie et Biologie (IECB) X-ray facility (UMS CNRS 3033 – INSERM US001) on a Rigaku MM007 HF rotating anode (0.8 kW). Data were collected at the Cu Kα wavelength with a partial chi goniometer (AFC11). All data-collection strategies were based on Omega scans. The X-ray source was equipped with high flux Osmic Varimax mirrors and a RAPID SPIDER semi-cylindrical image plate detector. The Rigaku CrystalClear suite versions 1.36 and 2.1 were used to index, integrate and scale the data with a multiscan absorption correction.

Crystallographic data for host–guest complexes $(5)_2 \supset 17$, $(5)_6 \supset 23$ and $3\cdot(6)_2 \supset 27$ were collected at the synchrotron ID29 (European Synchrotron Radiation Facility (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 allowed us to tailor the beam sizes down to 75 (full beam), 50, 30, 20 and 10 μm in diameter. Diffraction data were recorded with a fast readout Pilatus 6 M pixel detector (Dectris) that allows data collection with a maximum frame rate of 12 images per second. Data were collected with 0.2° increments (180° or 360° in total) and processed with the XDS package[34].

For all six crystal structures the resolution of the experimental data allowed phasing by direct methods or charge flipping. SUPERFLIP[35], SHELXT[36] and SHELXD[37] were run in parallel on all datasets 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 $(5)_2 \supset 21$, SHELXD for $2 \supset 16$ and $(5)_4 \supset 22$, and SUPERFLIP for $(5)_2 \supset 17$, $(5)_6 \supset 23$ and $3\cdot(6)_2 \supset 27$.

The starting model of the host–guest molecule $(5)_6 \supset 23$ 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 greatly in obtaining a full model ready to refine (the superimposition was made 'by hand' using COOT software[38] and SHELXLE software[39]. All the structures were refined by a full-matrix least-squares method on F^2 with SHELXL-2014 (ref. 37). As a rule of thumb, every time a disorder could be modelled with partial occupation, it was so. However, more often than not, no simple models of the side chains and solvents could be obtained because multiple positions exist and because complexity is introduced by the swapping of side chains and solvent molecules at some positions, which leads to an unstable refinement. Introducing too many restraints on the side chains destabilizes the main chain and thus also must be ruled out. Under such circumstances, an approach that many crystallographers would follow is 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 were modelled using the EADP (equal atomic displacement parameter) constraints 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 opinion, represents real information contained in the data. Such modelling is efficient but, because of radiation-induced damage, the position of some atoms of the side chains could not be established (only for structure $23 \supset (5)_6$). For all the non-hydrogen atoms, attempts to introduce anisotropic displacement parameters were made. However, whenever the ellipsoids adopted an unrealistic shape, an isotropic model was employed. FVAR instruction was used to force some of the isotropic temperature parameters to be equal. To model the anisotropic displacement parameters, ISOR, DELU and RIGU instructions were used. If it was necessary, the geometry of molecules was improved using DFIX, FRAG, FLAT or AFIX66 and AFIX116 SHELX instructions. Hydrogen atoms were positioned in idealized positions and refined with a riding model, with Uiso constrained to 1.2 Ueq value of the parent atom (1.5 Ueq when CH$_3$). 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 (large size of the molecules, large, heavily disordered parts, large solvent content and so on) and decisions made during data refinement (that is, isotropic displacement parameters for non-hydrogen atoms). The SQUEEZE[40] procedure was 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 (guest) molecules.

Data availability. Crystallographic data for $2 \supset 16$, $(5)_2 \supset 17$, $(5)_2 \supset 21$, $(5)_4 \supset 22$, $(5)_6 \supset 23$ and $3 \cdot (6)_2 \supset 27$ are 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 on reasonable request.

References

32. Shvartsburg, A. A., Mashkevich, S. V., Baker, E. S. & Smith, R. D. Optimization of algorithms for ion mobility calculations. *J. Phys. Chem. A* **111,** 2002–2010 (2007).
33. Siu, C.-K., Guo, Y., Saminathan, I. S., Hopkinson, A. C. & Siu, K. W. Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses. *J. Phys. Chem. B* **114,** 1204–1212 (2010).
34. Kabsch, W. XDS. *Acta Cryst.* **D66,** 125–132 (2010).
35. Palatinus, L. & Chapuis, G. SUPERFLIP—a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions. *J. Appl. Cryst.* **40,** 786–790 (2007).
36. Sheldrick, G. M. SHELXT—integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71,** 3–8 (2015).
37. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **71,** 3–8 (2015).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of COOT. *Acta Crystallogr. D* **66,** 486–501 (2010).
39. Hübschle, C. B., Sheldrick, G. M. & Dittrich, B. Shelxle: a Qt graphical user interface for SHELXL. *J. Appl. Cryst.* **44,** 1281–1284 (2011).
40. Spek, A. L. Single-crystal structure validation with the program PLATON. *J. Appl. Cryst.* **36,** 7–13 (2003).