Supramolecular Chemistry

Self-Organization of Oligomeric Helical Stacks Controlled by Substrate Binding in a Tobacco Mosaic Virus Like Self-Assembly Process**

*Anne Petitjean, Hélène Nierengarten, Alain van Dorsselaer, and Jean-Marie Lehn**

A major line of investigation in supramolecular chemistry is to understand the processes underlying the self-organization of matter[1,2] and to implement them in artificial systems.[3] Synthetic models that reproduce the recognition properties of nucleic acids,[4,5] proteins,[6] and carbohydrates,[7] as well as the organization of membranes,[8] are useful tools for the study of naturally occurring processes. The self-assembly of viruses represents a biological touchstone for the elaboration of complex architectures on the basis of molecular information. The Tobacco Mosaic Virus (TMV), one of the best-characterized viral systems,[9] is composed of 2130 identical protein subunits, which self-assemble around the viral RNA into a helical supramolecular polymer to define a tower whose height is determined by the polynucleotide substrate. Herein we describe the self-organization of synthetic helical subunits around a molecular cationic strand as a model for the self-assembly of the TMV.

The individual protein building blocks of the TMV self-assemble to yield cylindrical or lock-washer-type helical oligomers of different lengths depending on the pH and ionic strength of the medium. The specific recognition between the protein units and their nucleic acid substrate leads to the exclusive formation of a complex in which the RNA fits inside the nanotube-type helical assembly and dictates its length.

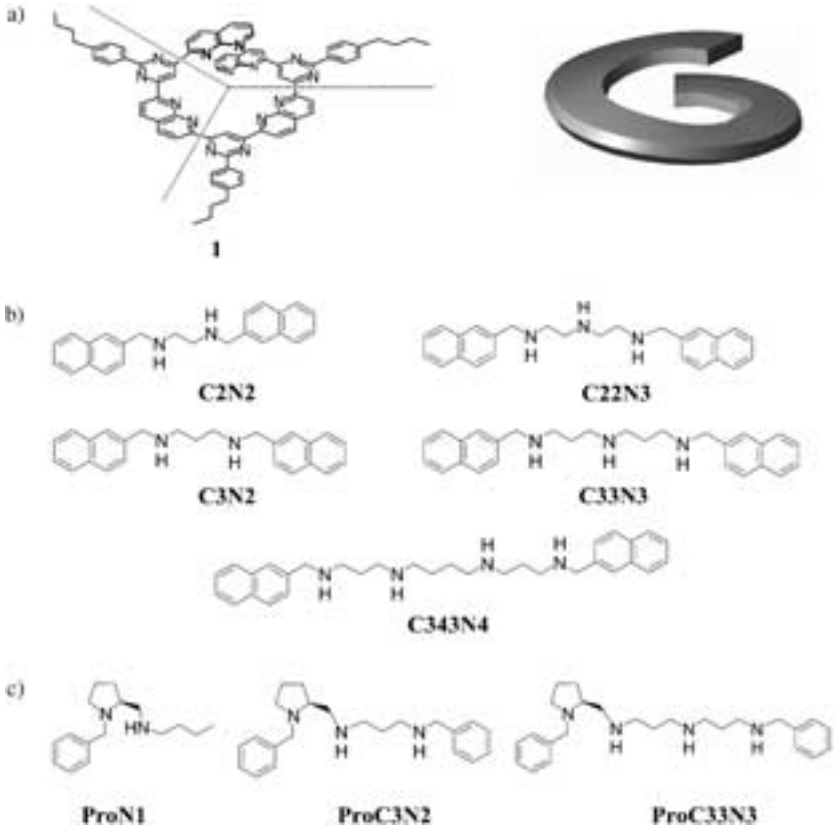

Figure 1. Structures of the compounds: a) helical strand **1**, in which the 3.5 identical motifs encoding for the global helical conformation are highlighted; b) achiral oligoamine substrates; c) chiral oligoamine substrates.

[*] Dr. A. Petitjean,[+] Prof. Dr. J.-M. Lehn
Laboratoire de Chimie Supramoléculaire
ISIS, 8 allée Gaspard Monge, BP 70028,
67083 Strasbourg Cedex (France)
Fax: (+33) 3-9024-5140
E-mail: lehn@isis.u-strasbg.fr
Dr. H. Nierengarten, Prof. Dr. A. van Dorsselaer
Laboratoire de Spectrométrie de Masse Bioorganique
ECPM, 25 rue Becquerel,
67088 Strasbourg Cedex (France)
[+] Present address:
Department of Chemistry M/C 127-72, Caltech,
1200 E. California Blvd., Pasadena, CA 91125 (USA)

[**] A.P. is grateful to the French Ministère de L'Éducation Nationale, de la Recherche et de la Technologie for financial support, to Dr. J.-A. Bousquet for assistance with the CD measurements, and to Dr. André Matthis (Institut Charles Sadron, Strasbourg) for his kind help with powder diffraction studies.

The naphthyridine-based helical oligomer **1**[10] (Figure 1 a) may be considered as a structural analogue of the one-turn, lock-washer-type protein oligomers occuring in the assembly of TMV. Whereas the latter helical species are based on the *noncovalent* interaction of individual proteins (16.3 per turn), **1** is preorganized into a helix through *covalent* linkage of helicity-enforcing motifs (3.5 per turn) that are based on the *transoid* conformational preference of *ortho*-linked aromatic azaheterocycles.[11] The convergence of several strong electric dipoles of the naphthyridine groups ($\mu = 4.1$ D in benzene, 25 °C)[12] towards the center of the helix creates a polar internal void of ≈ 3.5 Å diameter within **1** which is well-suited to accommodate small cationic guests such as alkali or guanidinium ions. These, in turn, promote the interaction with a free helix through ion–dipole and van der Waals interactions to result in the formation of polymolecular assemblies of

 DOI: 10.1002/ange.200353492

helices **1**.[10] The pH- and ionic strength (salt)-induced polymerization of TMV proteins into helical supramolecular architectures in aqueous solutions is thus mirrored by the cation-promoted polyassociation of helix **1** in organic solvents (Figure 2).

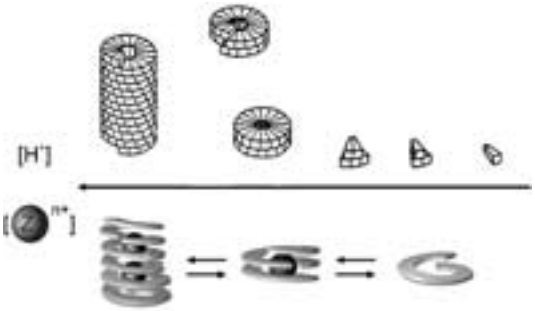

Figure 2. Schematic illustration of the parallel between the pH- and ionic-strength-dependent self-assembly of the TMV protein units and the cation-promoted aggregation of helical components **1** which may involve inorganic (alkali, lead(II)) or small organic (guanidinium) ions.[10]

As single cations are able to induce the aggregation of helices, we anticipated that a "string of cations" could serve as a template for the controlled oligomerization of helices, with the size of the overall architecture possibly dictated by the number of positive charges within the string. Therefore, two families of linear oligoammonium cations were synthesized by standard procedures (Figure 1b and c) and were characterized by ^{1}H and ^{13}C NMR spectroscopy, electrospray mass spectrometry (ESMS),[13] and elemental analysis. The symmetrical oligoammonium ions of the first family (Figure 1b) were obtained from primary diamines and contain internal secondary amines separated by two to four methylene groups. They are designated by **C*m*N*n***, in which ***n*** stands for the number of amine sites and ***m*** denotes the sequence of the number of methylene groups in all the spacers linking the successive N sites. The naphthyl end groups were introduced to allow easy identification of the complexes with **1**.[13b] The second family consists of proline-appended chiral oligoamines in which the protonation sites are separated by propylene spacers (Figure 1c).

The interaction of the symmetrical oligoamines **C*m*N*n*** with the helix **1** was monitored by means of ESMS analysis[13] of a freshly prepared mixture of **1** (500 μL of a 1.3×10^{-4} M stock solution in $CHCl_3$), trifluoroacetic acid (92 μL of a 7.0×10^{-4} M stock solution in $CHCl_3$; 1.0 equiv relative to **1**), and the amine in various proportions. Figure 3 shows the ES mass spectra of **1** in combination with the different oligoamines. The spectrum of **1** alone (data not shown) displays very weak signals of different oligomers in which one to three residual cations (H^+, K^+, Na^+) form complexes with 1—4 equivalents of **1** (see Figure 5 left, for a graphical illustration).

When **1** was mixed with 0.5 equivalents of the diamines **C2N2** and **C3N2** (**1**/diammonium = 2:1) in the presence of 1 equivalent of protons, only those complexes in which all the interaction sites (ammonium groups) are occupied were detected by means of ESMS. In the case of **C2N2**, a small peak at $m/z = 1487.0$ was detected that corresponds to the monoadduct $(\mathbf{1}\text{–}\mathbf{C2N2}H)^+$ but accounted at most for 7% of the overall signal. The two triamines **C*mm*N3** exhibit different behavior. In the case of **C33N3**, the nature of the major adduct depends on the relative proportions of the triammonium species and **1** in the mixture. Introduction of a stoichiometric amount of **1** with respect to N sites (**1**/**C33N3** = 3:1) leads to incomplete complexation of the receptor (Figure 3d bottom), with the fully saturated adduct (m/z 1283.7) present as the major species ($(\mathbf{1}_3\text{–}\mathbf{C33N3}H_3)^{3+}$/$(\mathbf{1}_2\text{–}\mathbf{C33N3}H_2)^{2+} = 1.5$). In the presence of an excess of **1** (**1**/**C33N3** = 5:1 and 10:1), the unsaturated complex almost completely disappears to the benefit of the fully saturated complex. With **C22N3** however, even a large excess of receptor (10:1) only yields minor amounts of the fully saturated species (Figure 3c). One explanation for this discrepancy may arise from the lower pK_a value of **C22N3** relative to that of **C33N3** since the third positive charge has to be accommodated closer to the two peripheral charges in **C22N3**. Nevertheless, as protonation of **1** is not observed, **C22N3** must be fully protonated as no other accessible base is present in the medium ($CHCl_3$).

Another factor arises from the steric requirements for the formation of the trimeric complex, if the van der Waals thickness of aromatic groups (3.4–3.5 Å) and the distance between two consecutive ammonium sites are taken into account. In the $-NH_2^+CH_2CH_2NH_2^+-$ fragment, the largest distance between two consecutive ammonium groups is about 3.8 Å when they adopt an *anti* conformation (based on molecular-modeling studies). Stacking two strictly planar aromatic surfaces on top of each other requires a minimal distance of 3.4–3.5 Å between the terminal ammonium sites. The helical units **1** are not planar and therefore need more room to fit. In the adduct with **C2N2**, the shorter distance between the two consecutive binding sites can be compensated by a shift of the two units of **1** towards the end groups. However, in the saturated adduct of **C22N3,** the proximity of the N sites could hinder the insertion of the central unit of **1** which would result in a smaller binding constant of this third receptor and hence a lower proportion of the saturated adduct. On the other hand, with two consecutive ammonium binding sites separated by about 5 Å in **C3N2** and **C33N3** there is enough room for the binding of a receptor unit to each ammonium site such that a ratio of 5:1 between **1** and **C33N3** allows quasi-complete saturation of the triammonium substrate (Figure 3d, middle).

To investigate this length-regulation process, spermine was functionalized with the same end groups to give the longer tetrammonium chain **C343N4** (Figure 1b) in which consecutive ammonium binding sites are separated by spacers containing three or four methylene groups. Upon mixing **C343N4** and **1** in a 1:7 ratio under acidic conditions (1 equivalent of acid relative to **1**), the ESMS spectrum (not shown) showed the coexistence of two species in equal quantities: the expected $(\mathbf{1}_4\text{–}\mathbf{C343N4}H_4)^{4+}$ adduct was accompanied by the lesser-charged $(\mathbf{1}_4\text{–}\mathbf{C343N4}H_3)^{3+}$ species, but four ligands were bound in both cases. The presence of the latter species, which accommodates four ligands with one

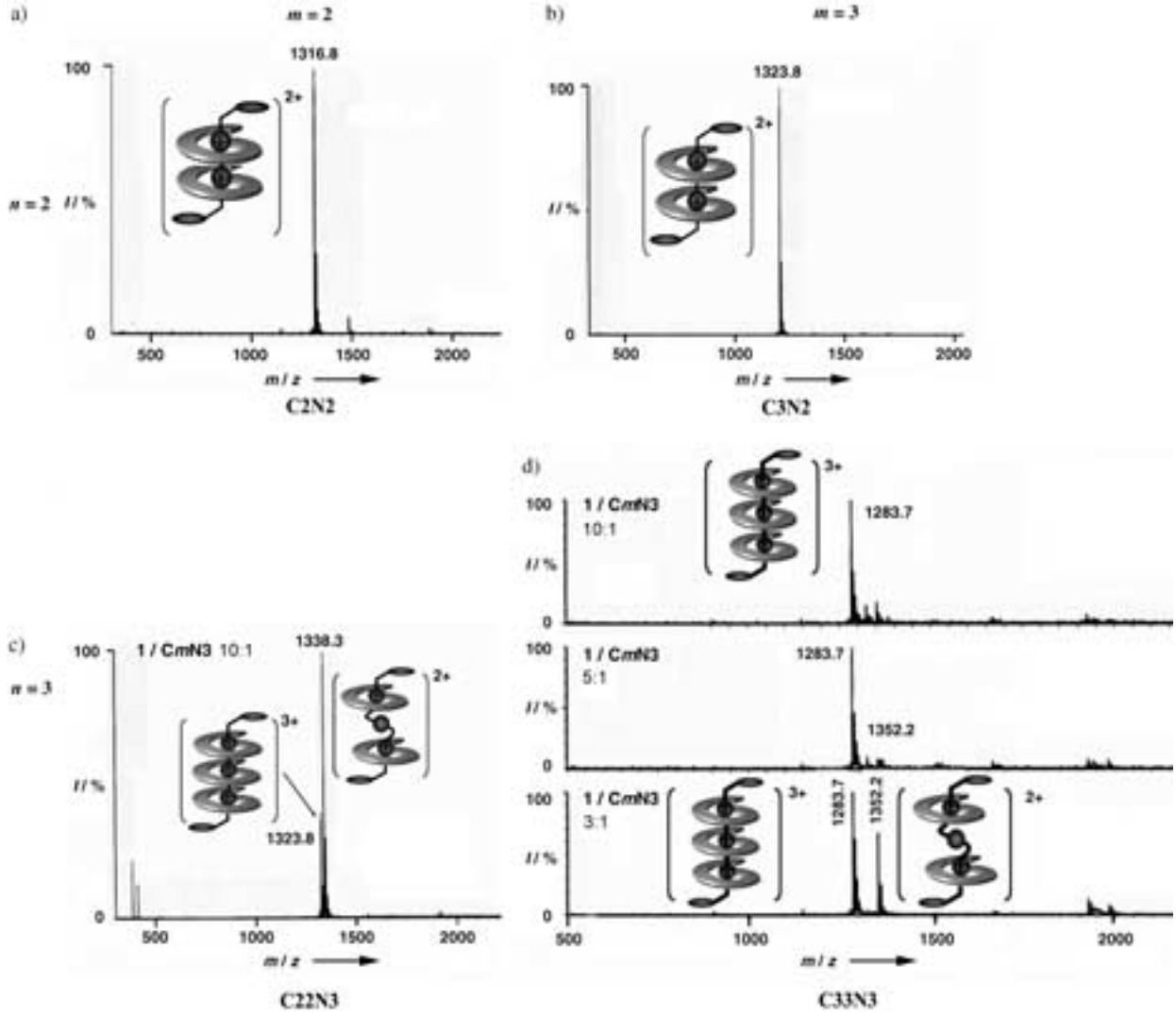

Figure 3. ESMS spectra obtained by mixing **1**, TFA (1 equiv) and the oligoamines **C*m*N*n*** (1/*n* equiv) in $CHCl_3$. In the case of **C*m*N3**, different **1**/**C*m*N3** ratios were investigated and are indicated. For experimental details see reference [13].

tricationic substrate, indicates that the ion–dipole and van der Waals interactions within the complex are strong enough to hold one of the ligands around an unprotonated nitrogen site.

Self-organization by selection[3] involves the selective assembly of a mixture of those components that are required for generating the final entity. This is the case in the formation of the "correct" double and triple inorganic double helices (helicates) from a mixture of ligand strands and metal ions,[14] and in the assembly of inorganic grid-type architectures.[15] To test for such selection in the present dynamic system of equilibrating species, competition experiments were carried out with mixtures of the **C*m*N*n*** substrates (***m***, ***n*** = 2 or 3). The selectivity of the self-assembly of units of **1** around the oligoammonium substrates bears relation to the specificity of the protein-coating process in the TMV to the exact viral RNA.

When **1** was mixed with TFA (1 equiv), **C2N2** (0.5 equiv), **C3N2** (0.5 equiv), **C22N3** (0.33 equiv), and **C33N3** (0.33 equiv), the only species observed were (in decreasing abundance): $[\mathbf{1}_2\text{–}\mathbf{C3N2}H_2]^{2+}$ (56%), $(\mathbf{1}_2\text{–}\mathbf{C33N3}H_2)^{2+}$ (26%), $(\mathbf{1}_3\text{–}\mathbf{C33N3}H_3)^{3+}$ (11%), and $(\mathbf{1}_2\text{–}\mathbf{C22N3}H_2)^{2+}$ (7%). It appears therefore that the **C3N2** and **C33N3** strands are strongly selected with the latter being present under both fully and partially complexed forms. On the contrary, **C2N2** and **C22N3** are markedly under-represented in this selection in agreement with the results above. Interestingly, the fully assembled complex with **C33N3** is present in significant proportions despite the required 3:1 stoichiometry, the competition with the other species, and the entropic cost of its full assembly—a noteworthy sign of its stability.

The poorer binding of oligoammonium chains containing short internal spacers between successive charges is a first indication that the helical conformation of the neutral ligand **1** is conserved in the oligoammonium complexes, in agreement with powder diffraction measurements on the K^+, Cs^+, and guanidinium complexes.[10] Related to this, we investigated chirality induction in the self-assembled oligomers through complexation of **1** with chiral oligoammoniums derived from L-proline (Figure 1c). A short C2 spacer between the first two ammonium binding sites was intention-

ally chosen instead of a C3 spacer so that the first interaction site would be closer to the stereogenic center. The receptor that interacts with the first secondary amine may then also experience a positive attraction towards the protonated appended proline, and thus be brought closer to the chiral inducer. Benzyl end groups, instead of naphthyl groups, were selected because of their lower spectroscopic absorption in the region of interest.

The circular dichroism (CD) signals of a mixture of **1** (0.5 mM), TFA (1 equiv), and **ProC3N2** ($n=2$; $1/n$ equiv) or **ProC33N3** ($n=3$) in $CHCl_3$ were monitored in a region where only receptor **1** absorbs light energy (350–400 nm) and the proline substrates show no absorption. Hence any CD signal detected in this region should result from a transfer of chirality from the proline substrate to ligand **1**. Figure 4 shows

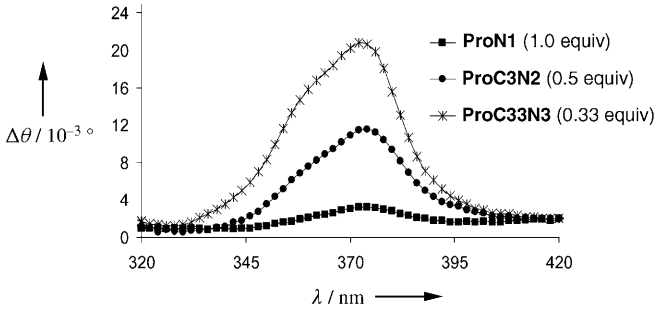

Figure 4. Circular dichroism spectra of a mixture of **1** (0.5 mM), TFA (1 equiv), and the proline-derived oligoamines ($1/n$ equiv) in $CHCl_3$ as a function of wavelength, in a region where only the naphthyridine groups of **1** absorb (the ellipticity $\Delta\theta$ is measured in millidegrees).

the CD curves of the three complexes in the naphthyridine absorption region. The shape of the signals is the same in all three cases with primary and secondary maxima at 372 and 362 nm, respectively.

The appearance of a CD signal indicates that the interaction between **1** and the substrates gives rise to a supramolecular chiral assembly in which the chirality originates from the noncovalently bound substrate. The interaction of **1** with **ProN1** (■, Figure 4) gives only a weak signal, which suggests that the complex does not have a strong chiral organization (ESMS studies suggest that monoammonium cations such as dibenzylammonium are complexed under these conditions). Although introduced in smaller concentrations, **ProC3N2** and **ProC33N3** give much stronger signals, which are a sign that the ligands absorbing in this region are chirally organized. Moreover, the number of interacting sites is increased which may also correspond to a better-defined geometry (the data are reported in ellipticity units, that is, relative to the same number of absorbing molecules **1**).

The CD data therefore indicate that the ligands **1** in the oligomeric species organize in accordance with a chiral pattern that is imposed by the proline-derivatized oligoammonium assembling agents; the similarity of the shape of the signals reflects a similar environment for the different complexes. Combined with the powder diffraction measurements on single cations[10] and the poor complexing ability of sterically hindered substrates, these data support a helical organization of the ligand **1** around the oligoammonium substrate. They indicate that the interaction of **1** with oligoammonium strands leads to the self-assembly of oligomeric chiral stacks of helical units **1** by the optically active polycationic scaffold, and that the size of the supramolecular architectures thus formed is dictated by the number of interaction sites present in the substrate (Figure 5). As stacking is expected to be better for helices **1** of the same chirality, the formation of homochiral stacks involves chiral selection of molecules **1** of the same helicity. On the other hand, diastereomeric stacks may, in principle, form between the optically active polycationic thread and enantiomeric stacks of **1** such that the CD effects represent an overall result for the population of all assemblies.

Taken together, the present ESMS and CD results are in agreement with the model of $(\mathbf{1}_x\text{–}\mathbf{C}m\mathbf{N}n\mathrm{H}_p)^{p+}$ stacks as outlined although they do not provide precise geometrical information. In particular, the ESMS data provide compelling evidence in favor of the structural model proposed.[16] Such stacks also represent rotaxanes[17] based on helical components. Their mechanism of formation may involve partial unwrapping of the helical unit **1** followed by insertion of the substrate chain into the opened cavity, in line with the process of dynamic helicity interconversion of related strands.[11] An alternative mechanism may be slippage, one of the processes that leads to the formation of rotaxanes.[17] Notably, the control of the polyrotaxane assembly[18] by the length of the thread is a process related to that described herein.

The length selection in these synthetic models reproduces the regulation process in the self-assembly of the TMV.

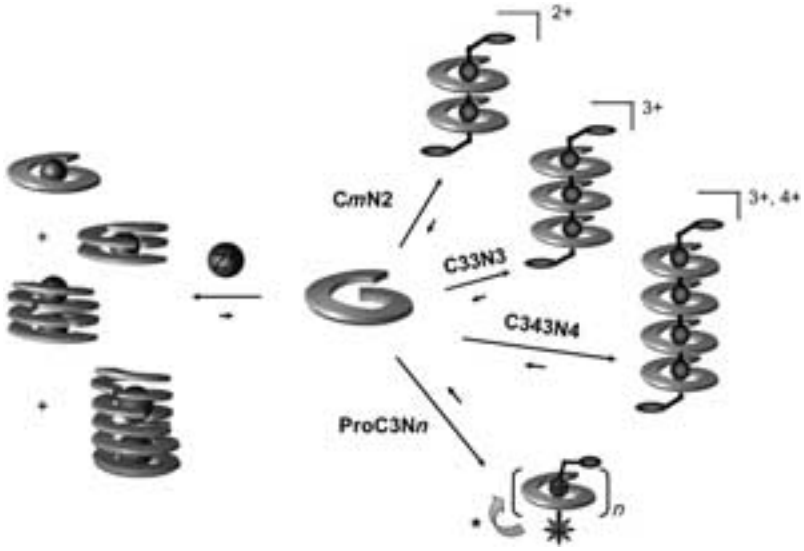

Figure 5. Graphical summary of the supramolecular oligomerization control directed by oligoammoniums substrates versus free oligomerization promoted by monocations (the curved arrow indicates chirality transfer).

However, unlike the viral coat which results from lateral interactions between more than 16 individual proteins assembling into the basic lock-washer entity through non-covalent interactions, helices **1** are covalently preorganized (Figure 2). In both cases, assembly occurs around a central scaffold that imposes supramolecular control of the final architecture. Organization mediated by lateral interactions has been described for systems that exhibit liquid-crystal properties.[19] Overall, the process described herein involves a first level of molecular self-organization to generate a helical building block followed by a further two levels of supramolecular self-organization, namely, the stacking of the helical units and the control of the stack by endo-recognition. It represents a higher level of self-organization[3] along a hierarchical three-stage sequence.

Received: December 10, 2003
Revised: April 30, 2004 [Z53492]

Keywords: amines · helical structures · N ligands · self-assembly · supramolecular chemistry

[1] M. Eigen, *Naturwissenschaften* **1971**, *58*, 465 – 523.
[2] *Self Organizing Systems, The Emergence of Order* (Eds.: F. E. Yates, A. Garfinkel, D. O. Walter, G. B.Yates), Plenum, New York, **1987**.
[3] J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 4763 – 4768.
[4] M. J. Han, J. Y. Chang, *Adv. Polym. Sci.* **2000**, *153*, 1 – 36.
[5] C. Leumann, *Bioorg. Med. Chem.* **2002**, *10*, 841 – 854.
[6] A. G. Cochran, *Curr. Opin. Chem. Biol.* **2001**, *5*, 654 – 659.
[7] C. A. Bertozzi, L. L. Kiessling, *Science* **2001**, *291*, 2353 – 2364.
[8] T. Kunitake in *Physical Chemistry of Biological Interfaces* (Eds.: A. Baskin, W. Norde), Marcel Dekker, New York, **2000**, pp. 283 – 305.
[9] A. Klug, *Angew. Chem.* **1983**, *95*, 579 – 596; *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 565 – 582.
[10] A. Petitjean, L. A. Cuccia, J.-M. Lehn, H. Nierengarten, M. Schmutz, *Angew. Chem.* **2002**, *114*, 1243 – 1246; *Angew. Chem. Int. Ed.* **2002**, *41*, 1195 – 1198.
[11] M. Ohkita, J.-M. Lehn, G. Baum, D. Fenske, *Chem. Eur. J.* **1999**, *5*, 3471 – 3481, and references therein.
[12] J. H. Markgraf, J. F. Skinner, G. T. Marshall, *J. Chem. Eng. Data* **1988**, *33*, 9 – 10.
[13] a) ESMS measurements were performed on a triple-quadrupole mass spectrometer Quattro II (Micromass, Altrincham, UK). The ES source was heated to 60 °C. The sampling cone voltage (Vc) was set at 70 V to avoid any fragmentation processes. The appropriate volume of oligoamine ($6–8 \times 10^{-4}$ M in $CHCl_3$) and TFA (7×10^{-4} M solution in $CHCl_3$, 1 equiv) were successively added to solutions of ligand **1** (1.3×10^{-4} M in $CHCl_3$). After 10 min, sample solutions were introduced into the mass spectrometer source with a syringe pump (Harvard type 55 1111: Harvard Apparatus Inc., South Natick, MA, USA) at a flow rate of 5 μL min^{-1}. Calibration was performed with protonated horse myoglobin. Scanning was performed in the MCA (multichannel analyzer) mode, and several scans were combined to obtain the final spectrum. b) MS signals of the different adducts with the parent oligoammonium chains that bear free primary amines at both ends, are very close and almost equal to the value for a 1:1 complex between **1** and a potassium ion. c) ESMS is not an exact quantitative method. One may assume that the intensities of the peaks of species of the same charge can be compared more accurately than when the charges are different. Species of higher charge usually 'fly' less well so they may be underestimated in this case, further favoring our interpretation.
[14] R. Krämer, J.-M. Lehn, A. Marquis-Rigault, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5394 – 5398.
[15] J. Nitschke, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11 970 – 11 974.
[16] The NMR spectra of the assemblies presented very broad signals displaced to lower chemical shifts; these features support the formation of stacked assemblies. Crystals suitable for X-ray crystallographic studies could not be obtained. UV/Vis and fluorescence spectrophotometric measurements require solutions of much lower concentrations at which aggregates do not form. Other methods of structural characterization are sought. A very approximate binding constant of 1000 M^{-1} was estimated for the interaction of **1** with the cesium cation. The broadness of the NMR signals may indicate nonspecific aggregation, as was observed when **1** was treated with protons in the presence of water.[10] However, the combination of helix **1** with the four oligoammonium chains studied clearly shows the formation of only those complexes that correspond to the number of ammonium sites (ESMS and CD). Still these defined entities may coexist with nonspecific aggregates. Nevertheless, if the entities observed by ESMS were simply undefined aggregates, one would expect the formation of the same mixture irrespective of the oligoammonium compound used. The structural assignment is therefore validated but other types of assemblies may coexist depending on the concentration (the concentration used in the NMR spectroscopic studies is 14-fold higher than that used in the ESMS experiments). A few features of the ESMS measurements deserve to be highlighted in this regard: In contrast to alkaline cations, the introduction of oligoammonium chains leads to very 'clean' ESMS spectra in which no small aggregates, excluding the ammonium substrates, are observed (even the protonated ligand is not detected). Additionally, nonspecific aggregates of specific assemblies (e.g., $(\mathbf{1}_n–\mathbf{C}_m\mathbf{N}_n)_x$) are not seen either, although one may expect a range of nonspecific products to form; it would therefore be surprising not to find any small assemblies contaminating the ESMS spectra.
[17] D. B. Amabilino, J. F. Stoddart, *Chem. Rev.* **1995**, *95*, 2725 – 2828; see, in particular, oligoamine-threaded cucurbituril pseudorotaxanes: K. Kim, *Chem. Soc. Rev.* **2002**, *31*, 96 – 107.
[18] A. Harada, *Acc. Chem. Res.* **2001**, *34*, 456 – 464; T. Dünnwald, R. Jäger, F. Vögtle, *Chem. Eur. J.* **1997**, *3*, 2043 – 2051; S. Anderson, H. L. Anderson, *Angew. Chem.* **1996**, *108*, 2075 – 2078; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1956 – 1959; H. W. Gibson, S. Liu, P. Lecavalier, C. Wu, Y. X. Shen, *J. Am. Chem. Soc.* **1995**, *117*, 852 – 874; G. Wenz, *Angew. Chem.* **1994**, *106*, 851 – 870; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 803 – 822.
[19] V. Percec, *J. Macromol. Sci. Pure Appl. Chem.* **1996**, *33*, 1479 – 1496; M. Suarez, J.-M. Lehn, S. C. Zimmerman, A. Skoulias, B. Heinrich, *J. Am. Chem. Soc.* **1998**, *120*, 9526 – 9532.

