Programmed Chemical Systems: Multiple Subprograms and Multiple Processing/Expression of Molecular Information

Jean-Marie Lehn*[a]

Dedicated to Pierre Boulez on the occasion of his 75th birthday

Abstract: Programmed chemical systems rest on the structural information stored in a molecular framework and on its reading and processing through non-covalent interactional algorithms to yield specific supramolecular entities. Beyond single-code self-assembly, which generates exclusively a single, specific superstructure, several codes may be implemented in the same overall program, thus opening the possibility to perform multiprogramming. Furthermore, the reading and processing of the same structural information through different interactional algorithms may lead to several different output entities, amounting to multiple expression of molecular information. Such features are revealed in the formation of double helicates, the assembly of metallosupramolecular architectures, and the differential reading of hydrogen bonding patterns in a molecular strand. They open novel perspectives within the framework of programmed chemical systems, concerning multiple processing capacity, and have intriguing implications from the biological point of view.

Keywords: molecular information · molecular recognition · programmed chemical systems · self-assembly · supramolecular chemistry

Introduction

One of the major lines of development of chemical science resides in the ever clearer perception, deeper analysis, and more deliberate application of information features in the elaboration and transformation of matter, thus tracing the path from merely condensed matter to more and more highly organized matter, towards systems of increasing complexity.

[a] Prof. Dr. J.-M. Lehn
Laboratoire de Chimie Supramolèculaire
ISIS, Universitè Louis Pasteur
4, rue Blaise Pascal, 67000 Strasbourg (France)
Fax: (+33) 3-88-41-10-20
E-mail: lehn@chimie.u-strasbg.fr

Programmed Chemical Systems

Supramolecular chemistry has paved the way towards perceiving chemistry as an *information science*. Its most far-reaching contribution to chemical science is the introduction and implementation of the concept of molecular information and its corollaries, with the aim of gaining control over the organization of matter. Through the appropriate manipulation of intermolecular non-covalent interactions, it has developed progressively into a chemistry of molecular information that involves the storage of information at the molecular level, in the structural features, and its retrieval, transfer, and processing at the supramolecular level, through specific interactional algorithms, which operate through molecular recognition events based on interaction patterns (hydrogen bonding arrays, sequences of donor and acceptor interactions, ion coordination sites).

Systems presenting such features may be considered as *programmed supramolecular systems*.[1–3] They involve the explicit application of molecular recognition as a means of controlling the evolution of supramolecular species, assemblies, and devices as they build up from their components through *self-processes* (self-assembly, self-organization, self-recognition…).[1, 2] They are based on the incorporation into molecular components of suitable instructions for the spontaneous but directed generation of well-defined supramolecular entities. Depending on the design of the interaction patterns between the components, more or less strict programming of the output species will be achieved. The program is molecular, the information being contained in the covalent structural framework; its operation through non-covalent, intermolecular recognition algorithms is supramolecular.

In correspondence with their external behavior, as members of an assembly, the internal scene of molecular species can be ruled by means of patterns of non-bonded interactions that operate at the *intramolecular* level and may be put to use for enforcing specific molecular geometries, such as helicity.[4]

Combining both the molecular and the supramolecular levels defines *programmed chemical systems* in their generality. An important feature of such systems is their *robustness*. In a *robust system* the instructions are strong enough for ensuring the stability of the process, that is, the self-assembly

  0947-6539/00/0612-2097 $ 17.50+.50/0

is stable towards modifications of parameters such as concentrations and stoichiometries of the components, presence of foreign species, and so forth. Conversely, when the assembly only occurs in a narrow range of parameter values, the system is unstable and presents a *singularity*; it may also display a bifurcation or a switching point between different assemblies.

The generation of a given superstructure results, in its simplest form, from the operation of a *single-code* self-assembly program. A step beyond consists in exploring the possibility to devise systems of higher complexity that would operate in multimode fashion and present *multiple-programming* features, based on the implementation of several codes within the same overall program. Specifically, the system may then behave either as a *linear combination* of the subroutines, each yielding its predetermined substructure, or as a *cross-combination* with interference between the subprograms.

We shall here analyze briefly some recent advances towards such systems, try to identify the concepts that they suggest, and speculate about the perspectives these may open. No exhaustive description of the surrounding research panorama will be given, but only ad hoc presentations will be cited which serve the purpose. The discussion will be based mainly on inorganic systems, but other types of interactions may be considered as well.

Inorganic self-assembly[2, 3, 5] involves the generation of well-defined metallo-supramolecular architectures from suitably designed organic ligands and specific metal ions. It is directed by the structural information stored in the ligands and the coordination algorithm defined by a given set of metal ions. It has allowed the generation of a variety of entities of high structural complexity, such as linear and circular helicates,[2, 5–8] cage compounds,[9–11] grid-type arrays,[9, 12–14] and so on.

Two Different Helicates from the Same Ligand Strand

Linear sequences of discrete metal-ion binding sites may serve as molecular strands which contain information that may be read out and processed by metal-ion complexation. Thus, linear ligands containing repeating bidentate complexation units such as bipyridine (B), or terpyridine (T) separated into discrete sites by a suitable spacer, combine with metal ions of tetrahedral (e.g., Cu^{I}, Ag^{I}) or octahedral (e.g., Fe^{II}) coordination geometry respectively, to produce homostrand double-helical metal complexes, duplex helicates.[2, 6, 7] Heteroduplex helicates are assembled for instance from tritopic BBB and TTT strands by means of the five-coordinate Cu^{II} ions.[15] Considering the formation of the DNA double helix from complementary polynucleotide strands, the nucleotide components and the base pairing by hydrogen bonding find their correspondence in the ligand binding sites and the metal-ion coordination, respectively.

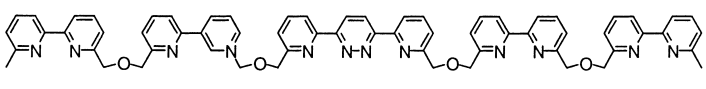
An important further step is represented by linear ligands that contain a given sequence of binding subunits which undergo self-assembly into heterosite homo- or heteroduplex complexes with identical or different metal ions (homo- or heterometallic). This is the case for instance with the mixed-site strands BTB and TBT, and Cu^{I}, Cu^{II}, or Fe^{II} metal ions, which generate the corresponding double helicates (Figure 1). One may note that the $[(BTB)Cu^{II}Cu^{II}Cu^{II}(TBT)]$ combination amounts to the translation of one strand into the other one by virtue of the Bn–T correspondence established by the five-coordinate Cu^{II} ions.[16]

Of direct relevance to the present purpose is the case of a BBT strand, which may generate two different homoduplexes when two different sets of metal-ion coordination geometries are brought into operation: a head-to-head double helicate with $[2(\text{tetrahedral}) + 1(\text{octahedral})]$ and a head-to-tail arrangement with $[1(\text{tetrahedral}) + 2(\text{pentacoordinated})]$ (Figure 1). Thus, the same ligand instructions lead to different products depending on the way in which they are processed. This feature has wide implications as has been pointed out earlier[17] and as will be discussed more extensively below.

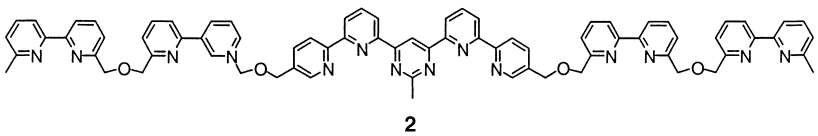
Two Different Metallomacrocycles from the Same Ligand Strand

Beyond systems operating on a single-code program that governs the self-assembly of a specific superstructure, such as a double helicate, a cage complex, a grid-type array, etc., it may be possible to combine several such codes as subroutines of an overall program.

This has been realized with ligands **1** and **2**, which contain two types of complexation subunits that code for the assembly of different structures, double-helicate and $[2 \times 2]$-grid structures, respectively, with given metal ions, thus implementing "double subroutine self-assembly" processes. They may produce an inorganic architecture of either type **A** or **B**, depending on whether the processing of the two sets of instructions occurs in independent or in combined fashion, respectively (Figure 2). Binding by ligand **1** of Cu^{I} ions of tetrahedral coordination geometry was found to give only structure **B**.[18] However, the instructions in ligand **1** do not allow a univocal control of the self-assembly process at the outset.



1



2

0947-6539/00/0612-2098 $ 17.50+.50/0 *Chem. Eur. J.* **2000**, *6*, No. 12

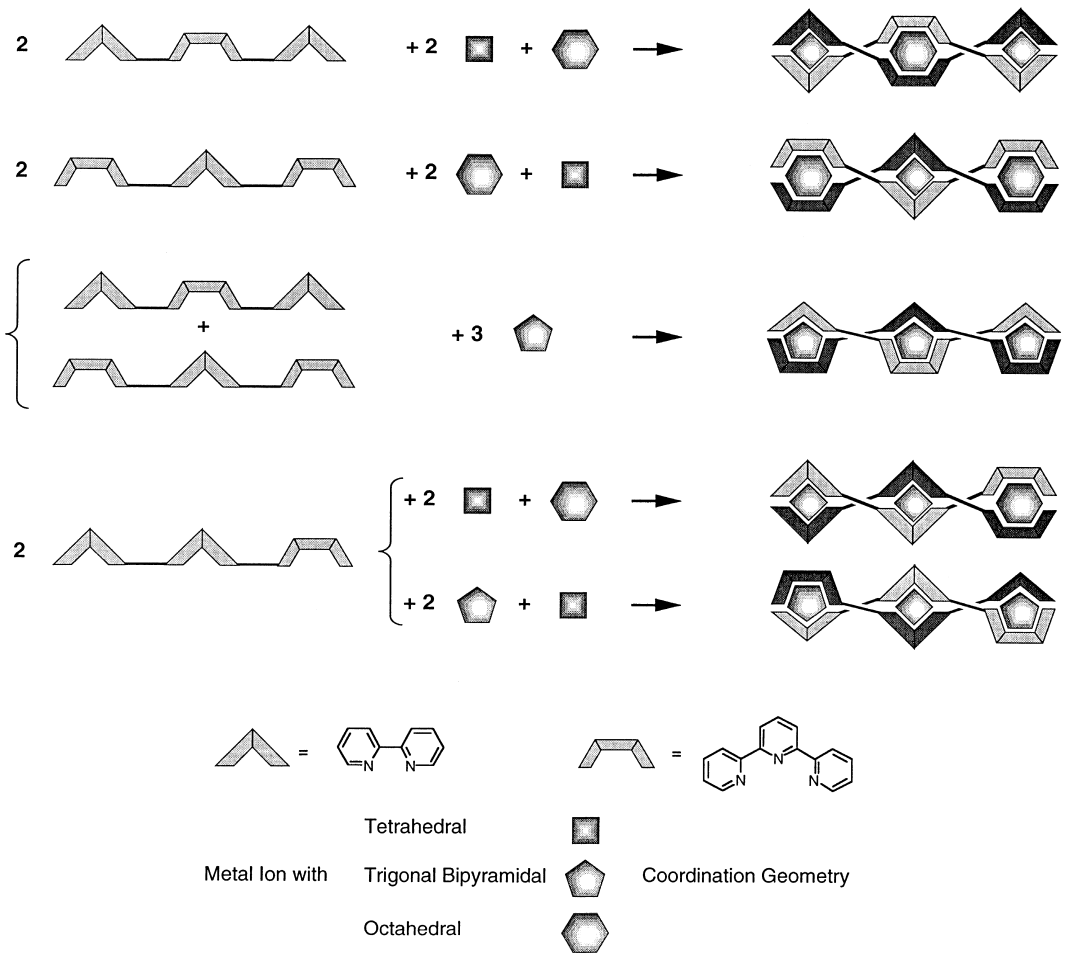


Figure 1. Schematic representation of the self-assembly of linear tritopic ligands (left) containing a defined sequence of binding subunits with specific metal ions (center) to yield predetermined double helicates (right). The ligands are designated as (from top to bottom): BTB, TBT, BTB, TBT, and BBT. The last example (bottom) represents a case in which the reading of the same instructions (binding subunits) by different interaction algorithms (sets of metal ions of different coordination geometry) yield different outputs (double helicates). The bidentate and tridentate binding subunits, represented here as bipyridine B and terpyridine T, respectively, may of course be in principle of other types.[5]

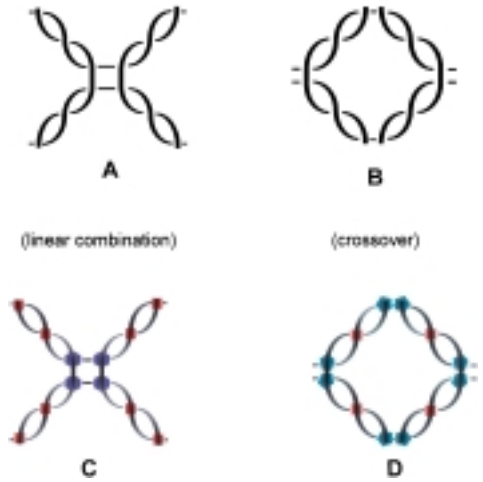


Figure 2. Schematic representation of the coordination architectures potentially accessible from double-subroutine-type ligands such as **1** and **2**, depending on whether the subprograms operate independently (**A**, **C**) or in a combined fashion (**B**, **D**). Architectures **C** and **D** result from the enforced processing of the binding information in ligand **2** by two different sets of metal ions of specific coordination geometries; squares, pentagons, and hexagons represent tetra-, penta-, and hexacoordinated metal ions, respectively.

To achieve this, it is necessary to design a ligand strand whose self-assembly with a given set of metal-ions enforces a *unique* output architecture. Furthermore, in such a case, if different output superstructures are in principle accessible, the processing of the *same* structural and binding information through different coordination algorithms, that is, with different sets of metal ions, may allow the directed generation of *different*, specific, and predictable self-organized architectures. This is realized with ligand **2** whose properties are described in detail in the following article.[19] It generates the architectures **C** or **D** (Figure 2) depending on the set of metal ions put into operation: the formation of **C** requires one equivalent of an ion of octahedral coordination such as Fe^{II}, Co^{II}, Ni^{II} and two equivalents of Cu^{I} per ligand, whereas **D** requires two equivalents of a pentacoordinated ion, such as Cu^{II} and one equivalent of Cu^{I} per ligand molecule.

Two Different Hydrogen-Bonded Assemblies from the Same Molecular Strand

Multiple expression of molecular information is of course not limited to systems based on metal-ion coordination, but can

also occur with other types of processing interactions such as hydrogen bonding. Thus, a given linear sequence of H-bond donor(D)/acceptor(A) patterns may adopt, in principle, different geometries and generate different supramolecular assemblies upon interaction with different complementary templates, exemplifying that the information contained in the patterns may be expressed differently depending on how it is read through a given interactional algorithm.

In the case of the conformationally flexible receptor strand **3**, which contains a linear sequence of four DAD H-bonding subunits (Figure 3), a dynamic library of numerous different conformers is obtained by rotation around the various bonds.[20] The outcome of a selection within this library and the nature of the assembly generated through binding of an effector, depend on the arrangement of the hydrogen-bonding sites in the template employed. With an ADA imide template, one would expect a linear readout of the strand to give a mixture of the different conformers of the supramolecular entity thus formed (Figure 3, bottom). In contrast, upon the binding of a double-faced, Janus type, ADA/ADA cyanurate template, curvature is introduced into the backbone of the strand **3**, so that binding of two effector units may generate three conformers of C, S, and helical shape (Figure 3, top).

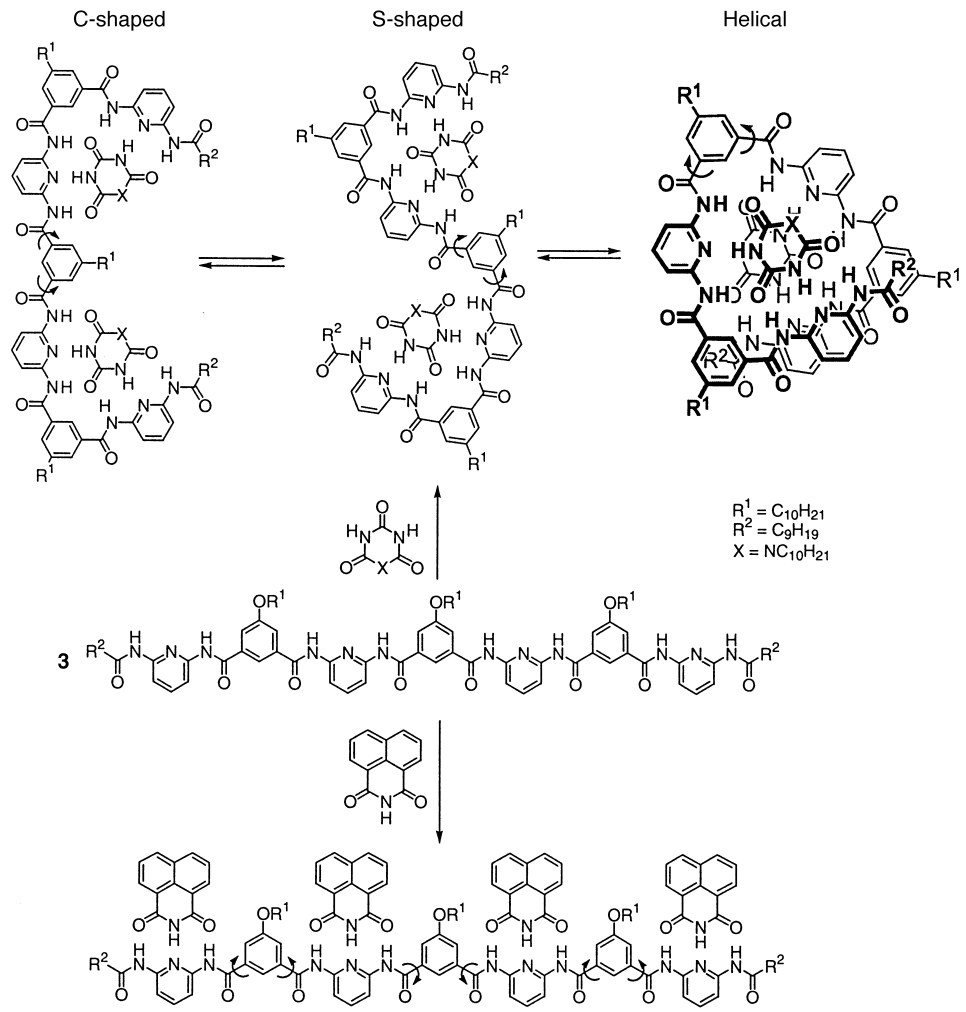


Figure 3. Hydrogen-bonded template-dependent expression of the information stored in the receptor strand **3** as a function of the processing mode through the recognition algorithm of the effector: angular reading/curved conformers of different shapes (top); linear reading (bottom).

The primary recognition event leads to the effector-induced generation of a coiled object, which thereafter undergoes self-assembly into helicoïdal columns. This sequence of events represents a hierarchical self-organization process.[17, 21] On a general level, one may note that such sequential self-assembly displays a conditional behavior and amounts to an IF logic gate, a given assembly being able to form only *if* the previous one has been generated.

Consequently, selection from the dynamic library of possible conformational isomers of **3** occurs differentially depending on how the receptor is read. The initial H-bond mediated substrate – receptor interaction may also be viewed as the deconvolution of a virtual dynamic library[22] of conformers to give a discrete supramolecular object that promotes a subsequent or "second level" self-assembly event.

Conjectures—Implications—Perspectives

From a general point of view, the results discussed here lead to conjectures which may have far-reaching implications and open novel perspectives within the general framework of programmed chemical systems.[1, 2]

1) The processes involving ligands **1** and **2** correspond to double subroutine self-assembly, the ligand strand containing two subunits which code for different structures. The output architectures generated depend on whether the two sets of instructions are processed independently or in a combined fashion. In more general terms, *multisubroutine self-assembly* may be considered to display three types of behavior: a) it may be *robust*, each sub-program running independently to generate its own encoded output; self-recognition[2, 23] is a related process; b) it may present *crossover*, when the subprograms operate in a combined fashion; c) it may also be of *dominant/recessive* type, one of the subprograms imposing its own output over the other one(s).

2) The three self-assembly processes analyzed above demonstrate that the processing of the *same* ligand information by *different* coordination algorithms (through the use of different sets of metal ions or of different H-bonding effectors)

0947-6539/00/0612-2100 $ 17.50+.50/0 *Chem. Eur. J.* **2000**, *6*, No. 12

allows the controlled generation of *different* output architectures. This can be generalized to and should hold for any type of molecular information and interaction (metal-ion coordination, hydrogen bonding, donor–acceptor, van der Waals). Thus, the information stored in a molecular framework does not necessarily code for a single species only, but may generate different entities depending on how it is read out and processed. The output is not fully determined by the information stored. In such a case, the information may be termed *degenerate*, the degeneracy being lifted by the processing. The operation of different reading algorithms leads to *multiple expression of molecular information*, yielding different output entities in a controlled fashion, through post-informational or post-instructional (post-genomic!) processing events. As a corollary, multiple processing of a single set of instructions allows the generation of *diversity*, since multiple outputs may either coexist, or be potentially accessible (virtual diversity).[22] It thus meets dynamic combinatorial chemistry[22] through the interaction-controlled reversible generation of a set of output entities.

3) *Multiple processing* capacity represents a further step in the design of programmed chemical systems of increasing complexity. One may expect that it will be possible to devise molecular programs of more and more complex architecture, capable of producing a variety of outputs under the strict control of the interactional recognition algorithms. As is already implicit in the cases considered above, such developments lead to the introduction of *parallel processing* into programmed chemical systems,[23a] extending eventually to massively parallel systems in which numerous self-assembly processes would operate in parallel towards the generation a single entity or of several different ones. Thus, beyond the combination of the same, single receptor/ligand with different effectors/ions, one may consider the simultaneous implementation of several different receptors with different effectors (as is the case for self-selection in helicate formation).[23b] Conversely, such chemical systems also open perspectives for *information science* itself, inasmuch as they raise the question of going beyond the usual one to one correspondence, established by a given program, between the input information and a single type of output, towards multiple outputs generated by different modes of processing the same information.

4) The combination of different recognition/instruction features in a molecular program opens a door to the design of self-assembling systems capable of performing *molecular computation*. Recent studies described the use of biomolecules and of DNA-based protocols to solve computational problems.[25] There is no reason why an approach making use of specifically designed non-natural components should not be feasible, with possibly higher diversity, better resistance to fatigue, and more compact/smaller size. It has been argued, that computing through self-assembly may provide a powerful alternative to conventional models.[26] Such potential may be perceived in the coordination-controlled assembly of the double helicates and of metallosupramolecular architectures, as well as in the differential-folding processes induced by effector H-bonding. discussed above. Numerous types of interactions and of recognition units, be they of inorganic or organic nature, are available for exploring these avenues.

5) From the *biological point of view*, the above processes would amount, for instance, to the generation of different products or the induction of different functions by alternate sensing of the same information through different interaction patterns. This could, in particular, involve different pairing schemes between natural and artificial nucleobases,[27] so that alternate modes of reading DNA or RNA sequences would lead to different DNA to RNA transcription or RNA to protein translation events, respectively, to yield, for instance, different proteins depending on the processing algorithm, defined by specific sets of interaction arrays between the partners. For instance, pairing patterns other than the Watson-Crick A=T, G≡C ones,[28] such as "wobble" pairs G=U and I=U, or involving non-natural code letters,[27] could yield different messenger RNAs from the same DNA, or different proteins from the same mRNA by reading the codons with different tRNAs that contain alternate anticodons for a same codon. Multiple reading would also result from the use of doublet and quadruplet codons in addition to the actual triplets. This corresponds to epigenetic variability/diversity within the general genetic framework. In some ways, point deletions leading to a frameshift amount to reading the remainder of the information in a different way and yield different products.

Multiple reading and expression of molecular information might in particular be relevant for processes involving protein–protein interactions, such as the synthesis of structurally very diverse peptides by multifunctional protein complexes according to the nonribosomal code.[29] Detailed studies of the proteome[30] might reveal such features. The binding of different effectors to the same allosteric receptor site of functional proteins, or of the same effector to different receptors may result in different activities. One may thus be lead to consider, in addition to the one code/one output mode a *one code/several outputs* scheme.

The combination of multiple expression and diversity generation, brought about through differential processing, with reversible build up of binding sites, opens perspectives towards a rather intriguing facet of an emerging adaptive chemistry (which becomes evolutive if the features gained remain acquired):[17] the notion of dynamic information generation and processing, defining *adaptive/evolutive programmed systems*.[31]

[1] J.-M. Lehn, *Angew. Chem.* **1990**, *102*, 1347; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 1304.

[2] J.-M. Lehn, *Supramolecular Chemistry—Concepts and Perspectives*, VCH, Weinheim, **1995**, chapter 9.

[3] *Comprehensive Supramolecular Chemistry*, *Vol. 9* (Eds.: J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vögtle, J.-M. Lehn), Pergamon, Oxford, **1996**.

[4] M. Ohkita, J.-M. Lehn, G. Baum, D. Fenske, *Chem. Eur. J.* **1999**, *5*, 3471, and references cited therein.

[5] a) D. L. Cander, K. N. Raymond, *J. Chem. Soc. Dalton Trans.* **1999**, 1185; b) M. Albrecht, *J. Incl. Phenom. Macro. Chem.* **2000**, *36*, 127; c) S. Leininger, B. Olenyuk, P. J. Stang, *Chem. Rev.* **2000**, *100*, 853.

[6] E. C. Constable, in ref. [3], chapter 6, p. 213.

[7] C. Piguet, G. Bernardinelli, G. Hopfgartner, *Chem. Rev.* **1997**, *97*, 2005.

[8] B. Hasenknopf, J.-M. Lehn, N. Boumediene, A. Dupont-Gervais, A. Van Dorsselaer, B. Kneisel, D. Fenske, *J. Am. Chem. Soc.* **1997**, *119*, 10956; for other circular entities, see for instance: A. Müller, C. Serain, *Acc. Chem. Res.* **2000**, *33*, 2.

[9] P. N. W. Baxter, in ref. [3], chapter 5, p. 165.
[10] M. Fujita, in ref. [3], chapter 7, p. 253.
[11] a) P. N. W. Baxter, J.-M. Lehn, G. Baum, D. Fenske, *Chem. Eur. J.* **1999**, *5*, 102; b) P. N. W. Baxter, J.-M. Lehn, B. O. Kneisel, G. Baum, D. Fenske, *Chem. Eur. J.* **1999**, *5*, 113.
[12] P. N. W. Baxter, J.-M. Lehn, J. Fischer, M.-T. Youinou, *Angew. Chem.* **1994**, *106*, 2432; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 2284.
[13] G. S. Hanan, D. Volkmer, U. S. Schubert, J.-M. Lehn, G. Baum, D. Fenske, *Angew. Chem.* **1997**, *109*, 1929; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1842.
[14] A. M. Garcia, F. J. Romero-Salguero, D. M. Bassani, J.-M. Lehn, G. Baum, D. Fenske, *Chem. Eur. J.* **1999**, *5*, 1803.
[15] B. Hasenknopf, J.-M. Lehn, G. Baum, D. Fenske, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1397.
[16] V. C. Smith, J.-M. Lehn, *Chem. Commun.* **1996**, 2733.
[17] J.-M. Lehn, in *Supramolecular Science: Where It Is and Where It Is Going* (Eds.: R. Ungaro, E. Dalcanale), Kluwer, Amsterdam, **1999**, p. 287.
[18] D. P. Funeriu, J.-M. Lehn, G. Baum, D. Fenske, *Chem. Eur J.* **1997**, *3*, 99.
[19] D. P. Funeriu, J.-M. Lehn, K. M. Fromm, D. Fenske, *Chem. Eur. J.* **2000**, *6*, 2103.
[20] V. Berl, M. J. Krische, I. Huc, J.-M. Lehn, M. Schmutz, *Chem. Eur. J.* **2000**, *6*, 1938.
[21] See also for instance: a) M. Suárez, J.-M. Lehn, S. C. Zimmerman, A. Skoulios, B. Heinrich, *J. Am. Chem. Soc.* **1998**, *120*, 9526; b) L. A. Cuccia, J.-M. Lehn, J.-C. Homo, M. Schmutz, *Angew. Chem.* **2000**, *112*, 239; *Angew. Chem. Int. Ed.* **2000**, *39*, 233.
[22] J.-M. Lehn, *Chem. Eur. J.* **1999**, *5*, 2455, and references therein.
[23] a) For a case of parallel processing in a multicomponent system by means of luminescent lanthanide complexes, see: S. Faulkner, D. Parker, J. A. G. Williams, in *Supramolecular Science: Where It Is and Where It Is Going* (Eds.: R. Ungaro, E. Dalcanale), Kluwer, Amsterdam, **1999**, p. 53; b) R. Krämer, J.-M. Lehn, A. Marquis-Rigault, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5394.
[24] *Evolutionary Computation Vol. 1 and 2* (Eds.: T. Bäck, D. B. Fogel, Z. Michaleweiz), Institute of Physics, Bristol, **2000**.
[25] L. M. Adleman, *Science* **1994**, *266*, 1021; R. J. Lipton, *Science* **1995**, *268*, 542; F. Guarneri, M. Fliss, B. Bancroft, *Science* **1996**, *272*, 220; M. C. Pirrung, R. V. Connors, A. L. Odenbaugh, M. P. Montague-Smith, N. G. Walcott, J. J. Tollett, *J. Am. Chem. Soc.* **2000**, *122*, 1873; J. Chen, D. H. Wood, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1328; D. Faulhammer, A. R. Cukras, R. J. Lipton, L. F. Landweber, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1385.
[26] M. Conrad, *Nanobiology* **1993**, *2*, 5; P. W. K. Rothemund, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 984.
[27] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1990**, *343*, 33; J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1992**, *356*, 537; D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585; A. K. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 3274.
[28] N. B. Leontis, E. Westhof, *Q. Rev. Biophys.* **1998**, *31*, 399; T. Hermann, E. Westhof, *Chem. Biol.* **1999**, *6*, R335.
[29] H. von Döhren, R. Dieckmann, M. Pavela-Vrancic, *Chem. Biol.* **1999**, *6*, R273.
[30] F. Lottspeich, *Angew. Chem.* **1999**, *111*, 2630; *Angew. Chem. Int. Ed.* **1999**, *38*, 2477.
[31] Some may consider (as indeed have some referees here and there!) that such conceptually loaded writing is at places overly difficult to read and hard to digest. One may contend that there is no reason not to accept for scientific texts, what one is ready to accept (celebrate!) in arts, literature and human sciences, referring, with all due respect, for instance to Emmanuel Kant, Ludwig Wittgenstein or James Joyce…

  0947-6539/00/0612-2102 $ 17.50+.50/0 *Chem. Eur. J.* **2000**, *6*, No. 12