Systems Chemistry

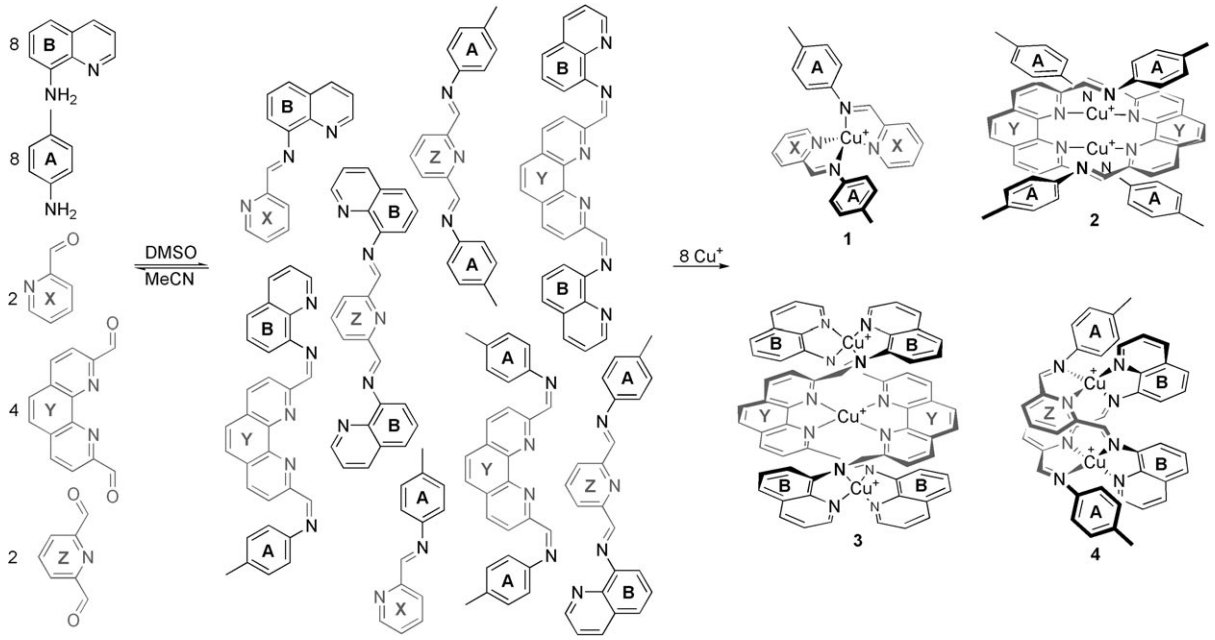
DOI: 10.1002/anie.200703877

Self-Assembly in Systems of Subcomponents: Simple Rules, Subtle Consequences**

*Rupam J. Sarma and Jonathan R. Nitschke**

A defining feature of living organisms is the ability of their biomolecular machinery to create a highly complex set of functional structures from a relatively limited set of basic building blocks.[1] Although the same subcomponents are employed across a wide variety of structures, and although the linkages that connect them are in many cases formed reversibly, under thermodynamic control,[2] biological structures have evolved not to interfere with each others' assembly: A high degree of compartmentalization occurs during assembly processes in living systems. In many cases, this compartmentalization is physical, involving the spatial separation of two systems by a membrane, for example. In other cases, particularly in prokaryotes, no such physical separation exists, and the systems must rely upon self-sorting of chemical species to avoid "crosstalk",[3] the unwanted sharing of subcomponents between systems.[4] In seeking to understand and mimic natural self-sorting, chemists have investigated a variety of model systems,[5] which have contributed to the understanding of the principles that underlie their behavior.

Herein we describe a self-sorting system,[6] shown in Scheme 1, for which we have deciphered the basic rules that govern which products will be observed under what circumstances. Although these rules are simple, the complexity of the overall system and the degree of control obtained represent advances in the state of the art. By controlling the stoichiometry of the five starting subcomponents, any arbitrary subset of the four product structures, in any relative proportion, may be prepared. The self-assembly rules governing this system are thus deterministic and may be used as programming instructions for the selection of "output" structures based upon "input" chemical species.



Scheme 1. The Cu^{I}-templated formation of structures **1**–**4** from amines **A** and **B** together with aldehydes **X**, **Y**, and **Z**.

As shown in Scheme 1, the reaction of amines **A** and **B** with aldehydes **X**, **Y**, and **Z** in acetonitrile/DMSO produced a dynamic library[7] of imines in equilibrium with the starting materials and half-formed products (not shown) in which only a single aldehyde group of **Y** or **Z** reacted with **A** or **B**. The

[*] Dr. R. J. Sarma, Dr. J. R. Nitschke
University of Cambridge
Department of Chemistry
Lensfield Road Cambridge, CB2 1EW (UK)
Fax: (+44) 223-793-215
E-mail: jn34@cam.ac.uk

[**] This work was supported by the Walters-Kundert Charitable Trust, the Swiss State Secretariat for Education and Research and the Swiss National Science Foundation.

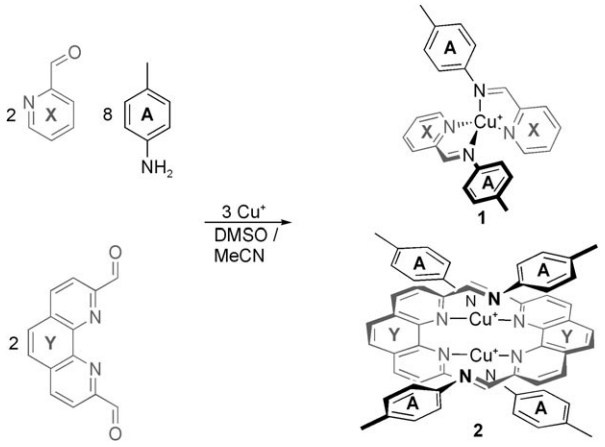
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

addition of copper(I) tetrafluoroborate resulted in the clean formation of a mixture of structures **1**–**4**, which were identified by NMR spectroscopy and ESI-MS. Complexes **1**,[8] **2**,[9] **3**,[10] and **4**[11] could also be prepared individually from their constituent subcomponents and copper(I), as has been previously reported.

Although it might be imagined that each of the intermediate imine structures shown in Scheme 1 could interact in a variety of ways with Cu^{I} ions, only four products were observed after thermodynamic equilibration. Most of the imines were eliminated from the initial dynamic library[12] which indicated that the Cu^{I} ions exerted a strong template effect[13] upon this library.

This selectivity may be encapsulated as a rule of valence satisfaction:[11] *The smallest possible structures will be formed in which all Cu^{I} ions are tetracoordinate and all nitrogen atoms are bound to a Cu^{I} ion*. In structures incorporating dialdehydes **Y** and **Z**, such structures are di- or trinuclear helicates, because imine ligands derived from these subcomponents are poorly configured to chelate a single Cu^{I} ion pseudotetrahedrally.[11] This rule does not preclude the formation of a mixed-ligand dicopper helicate incorporating one ligand from each of **2** and **4**. ESI-MS indicated the presence of such a product in only very small quantities (ca. 1%), however, and we were not able to identify its resonances in NMR spectra.

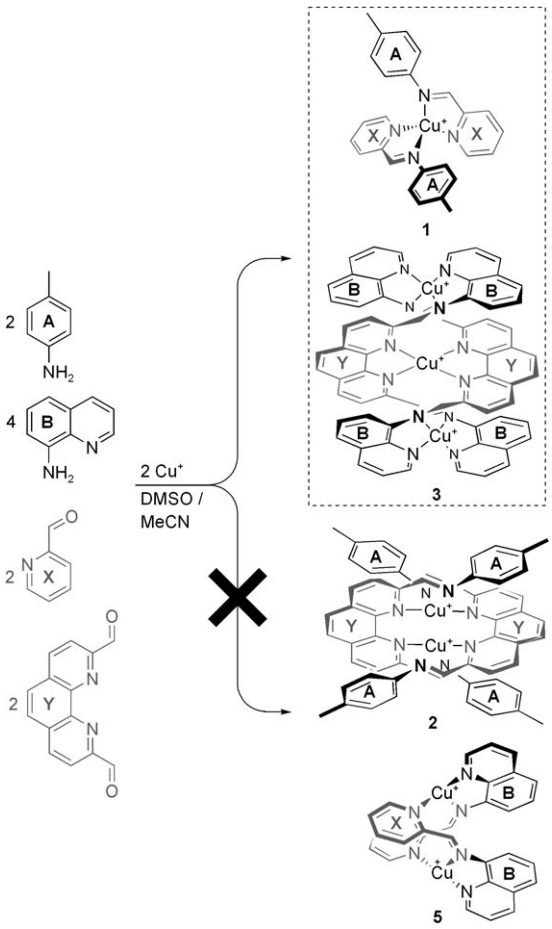
This rule allows us to predict the outcome of self-assembly reactions involving subsets of the full array of subcomponents of Scheme 1. In simple cases, this outcome is relatively straightforward, as shown in Scheme 2. The presence of one amine (**A**) and two aldehydes (**X** and **Y**) only allows two



Scheme 2. Since **A**, **X**, and **Y** may only be used to construct the ligands of **1** and **2**, these are the only products observed.

imines to be formed; the path from these two imine ligands to complexes **1** and **2** is thus clear-cut.

A more complex system is generated from two amines and two aldehydes. For example, when the four subcomponents shown in Scheme 3 were mixed with Cu^{I} in the proportions shown, only products **1** and **3** were observed by NMR spectroscopy and ESI-MS. Notably, this system allowed for the clean sorting of four subcomponents into two independent "baskets" (product structures) using a single metal-ion



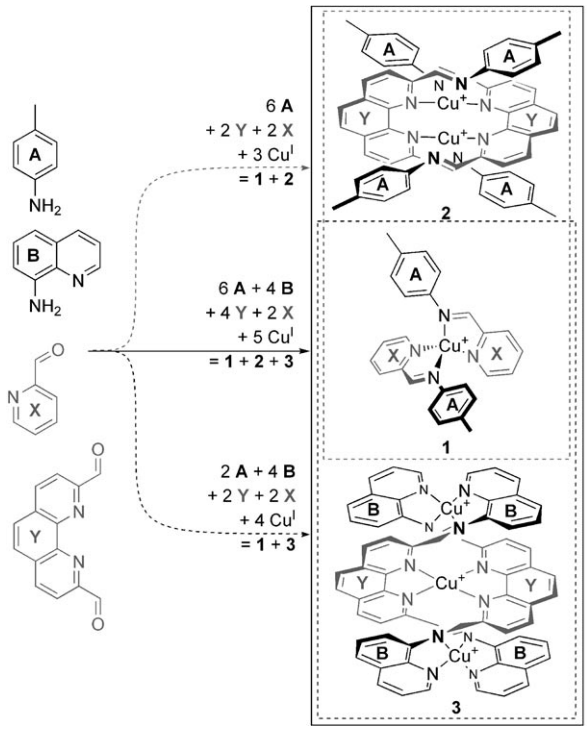
Scheme 3. Although **1**, **2**, and **3** are all individually allowed, only **1** and **3** are observed in this system for the given stoichiometry.

template. The only other system reported to do this required two separate metal ions, Cu^{I} and Fe^{II}.[14]

Both dialdehyde **Y** and aniline **A** were present in the system shown in Scheme 3, ostensibly permitting the formation of "allowed" structure **2**. The removal of **Y** and **A** leaves **X** and **B** to form **5**. This structure, however, is a "forbidden", as the Cu^{I} centers are not tetrahedrally coordinated. Avoidance of forbidden **5** thus leads to the suppression of allowed **2**, and only **1** and **3** are observed as products.

This result demonstrates that the observed set of products cannot be predicted only upon the basis of the stability of individual products. The rule of valence satisfaction must be applied to the system as a whole; all product structures must obey it.

Although **1** and **3** are the sole products observed in the system of Scheme 3, a simple change in stoichiometry altered the composition of this product mixture. As shown in Scheme 4, product mixtures consisting of **1**+**2**, **1**+**3**, and **1**+**2**+**3** are all accessible by altering the relative proportions of **A** and **B** and of **X** and **Y**. When **X** is present, **A** must react preferentially with **X** to form **1**, since **X** cannot form an allowed product with **B**. Since **Y** may generate either **2** or **3**, the ratio of **A** and **B** after the formation of **1** determines the ratio of **2** to **3**. NMR spectra showing the conversion of a



Scheme 4. By changing the stoichiometry of **A**, **B**, **X**, and **Y**, any combination of **1**, **2**, and **3** may be selected as products.

mixture of subcomponents **A**, **B**, **X**, and **Y** into a mixture of **1**, **2**, and **3** after the addition of Cu^{I} are shown in Figure 1.

The system may thus be said to follow a program during the self-assembly process. First, aldehydes **X** and **Z** react to produce **1** and **4**, removing the corresponding amounts of **A** and **B** from the system. Second, aldehyde **Y** reacts to produce **2** and **3**, the relative proportions of which depend upon the amounts of **A** and **B** left in solution. This program may be quantified; from a mixture containing **a** equivalents of **A**, **b** of **B**, **x** of **X**, **y** of **Y**, and **z** of **Z**,[15] the following products would be observed: **x**/2 equivalents of **1**, (**a**−**x**−**z**)/4 of **2**, (**b**−**z**)/4 of **3**, and **z**/2 of **4**. For these rules to hold, it is necessary to impose the boundary conditions of Equations (1)–(3):[15]

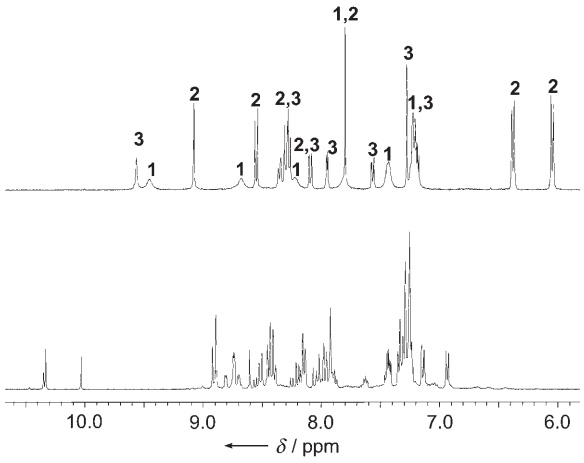


Figure 1. ^{1}H NMR spectra corresponding to the dynamic library formed by subcomponents **A**, **B**, **X**, and **Y** (bottom) and the mixture of products **1**+**2**+**3** (top) generated by addition of Cu^{I} to this library.

$$\mathbf{a}-\mathbf{x}-\mathbf{z} \geq 0 \quad (1)$$

$$\mathbf{b}-\mathbf{z} \geq 0 \quad (2)$$

$$\mathbf{y} = 2\,\mathbf{a} + 2\,\mathbf{b} - 2\,\mathbf{x} - \mathbf{z} \quad (3)$$

The condition imposed by Equation (1) ensures that a sufficient quantity of **A** is present to react with all **X** and **Z** present in the system to produce **1** and **4**. Equation (2) likewise ensures that enough **B** is available to react with all dialdehyde **Z**. Equation (3) requires that the amount of dialdehyde **Y** present in the system be sufficient to react with any residual **A** and **B**, once the requirements of **X** and **Z** have been met. Outside of these boundary conditions, deterministic control is lost. For example, when **A**, **B**, and **Y** are present in quantities such that $2\,\mathbf{a}+2\,\mathbf{b}<\mathbf{y}$, mixtures of different proportions of **2** and **3** are observed. Higher concentrations of Cu^{I} tend to favor the formation of **3**, as we have observed previously.[10]

Table 1 lists the correct stoichiometry of subcomponents required to arrive at an arbitrary equimolar subset of the four products. All of these combinations have been tested under

Table 1: The quantities of subcomponents and Cu^{I} required to generate any arbitrary subset of the structures shown in Scheme 1.

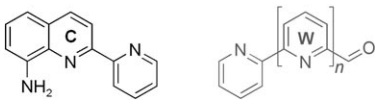
Amine(s)	Aldehyde(s)	Equiv Cu^{I}[15]	Observed product(s)
2**A**	2**X**	1	**1**
4**A**	2**Y**	2	**2**
4**B**	2**Y**	3	**3**
2**A**+2**B**	2**Z**	2	**4**
6**A**	2**X**+2**Y**	3	**1**+**2**
2**A**+4**B**	2**X**+2**Y**	4	**1**+**3**
4**A**+2**B**	2**X**+2**Z**	3	**1**+**4**
4**A**+4**B**	4**Y**	5	**2**+**3**
6**A**+2**B**	2**Y**+2**Z**	4	**2**+**4**
2**A**+6**B**	2**Y**+2**Z**	5	**3**+**4**
6**A**+4**B**	2**X**+4**Y**	6	**1**+**2**+**3**
8**A**+2**B**	2**X**+2**Y**+2**Z**	5	**1**+**2**+**4**
4**A**+6**B**	2**X**+2**Y**+2**Z**	6	**1**+**3**+**4**
6**A**+6**B**	4**Y**+2**Z**	7	**2**+**3**+**4**
8**A**+6**B**	2**X**+4**Y**+2**Z**	8	**1**+**2**+**3**+**4**

conditions where each product is present at a concentration of approximately 10 mM; NMR spectra are presented in the Supporting Information. Products within mixtures were identified by ESI-MS of the mixtures and by direct comparison of ^{1}H NMR spectra of pure products with the spectra of mixtures. Attempts to use fingerprinting to identify the components of these mixtures using 2D NMR spectroscopy (COSY, ROESY) were unsuccessful; we attribute this to the wide variation in ^{1}H relaxation times between the various products.

More complex mixtures, having arbitrary product ratios, are predicted to be readily accessible within the boundaries set by Equations (1)–(3), by simply mixing together subcomponents in the ratio in which they are found in the desired collection of product structures.[15] The system of Scheme 1 is thus fully deterministic. The four degrees of freedom[16] associated with the quantities of the inputs **A**, **B**, **X**, **Y**, and

Z are fully reflected in the quantities of outputs **1**, **2**, **3**, and **4** observed after self-assembly.

Examination of ways in which the aldehydes and amines of Scheme 1 might come together reveals how this system might or might not be extended, in keeping with the rule of valence satisfaction. The addition of an amine that contains an additional pyridyl nitrogen atom, such as 2-(2-pyridyl)-8-aminoquinoline **C** (Scheme 5), would lead to an undefined



Scheme 5. Amine **C**, whose addition to the system of Scheme 1 would result in a loss of deterministic control over the product mixture in the presence of one other amine and more than one aldehyde, and aldehydes of type **W**, the addition of any number of which to this system would result in the conservation of control.

outcome when present with amine **A** and two or more aldehydes. Any collection of monoaldehydes of type **W** (Scheme 5) could be added, however, and would lead to a well-defined mixture of products. The key difference is that amine **C** could add three different products to the dynamic library[7] of products of Scheme 1, whereas any collection of aldehydes of type **W** would add only one product per aldehyde. A more complete discussion is presented in the Supporting Information.

The present system thus demonstrates one simple mechanism by which the complex self-organizing systems of biomolecules within prokaryotic cells[4] might avoid interfering with one another. The question of how molecules might organize into mutually noninterfering complex systems is also related to the question of how prebiotic chemical systems became alive.[17]

Still more complex self-organizing systems might be created by adding further layers of dynamic linkages[18] that do not interfere chemically with the imines and the metal coordination of the present system. These additional linkages would impose their own selection rules upon the overall self-assembly process, eventually allowing access by thermodynamic self-assembly to structures of sufficient complexity to generate complex function.[19]

Experimental Section

Full experimental and characterization details, including copies of NMR spectra, are presented in the Supporting Information. In a typical experiment, calculated amounts of amines **A** and **B** and aldehydes **X**, **Y**, and **Z** (as noted for each of the entries in Table 1) and [D_3]MeCN (0.4 mL) were added to a teflon-capped NMR tube, and the corresponding quantity[15] of [$Cu(NCMe)_4$]BF_4 was added. Quantities were chosen such that the concentration of the product copper complexes would be approximately 10 mM. The dark brown mixture thus obtained was degassed and purged three times with argon. The reaction mixture was kept overnight at room temperature and then at 50 °C for 24 h. During this period dark brown solids precipitated. The volume of solvent was reduced to half the original volume under vacuum, and subsequently [D_6]DMSO (0.2 mL) was added, which resulted in dissolution of the precipitated solids. The reaction mixture was again kept at 50 °C for 24 h, after which the volume of solvent was reduced by half and further [D_6]DMSO (0.2 mL) was added. The progress of the reaction was monitored using ^{1}H NMR spectroscopy over 7 d, after which time the spectrum corresponded in all cases to the mixture of products indicated in Table 1.

Received: August 23, 2007
Published online: November 16, 2007

Keywords: combinatorial chemistry · dynamic covalent chemistry · self-assembly · systems chemistry · template synthesis

[1] L. M. Greig, D. Philp, *Chem. Soc. Rev.* **2001**, *30*, 287 – 302.
[2] J. M. Lehn, *Science* **2002**, *295*, 2400 – 2403.
[3] A. X. Wu, L. Isaacs, *J. Am. Chem. Soc.* **2003**, *125*, 4831 – 4835.
[4] F. M. Harold, *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 544 – 564.
[5] a) R. Krämer, J. M. Lehn, A. Marquis-Rigault, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5394 – 5398; b) K. A. Jolliffe, P. Timmerman, D. N. Reinhoudt, *Angew. Chem.* **1999**, *111*, 983 – 986; *Angew. Chem. Int. Ed.* **1999**, *38*, 933 – 937; c) S. J. Rowan, D. G. Hamilton, P. A. Brady, J. K. M. Sanders, *J. Am. Chem. Soc.* **1997**, *119*, 2578 – 2579; d) T. Kamada, N. Aratani, T. Ikeda, N. Shibata, Y. Higuchi, A. Wakamiya, S. Yamaguchi, K. S. Kim, Z. S. Yoon, D. Kim, A. Osuka, *J. Am. Chem. Soc.* **2006**, *128*, 7670 – 7678; e) H. B. Yang, K. Ghosh, B. H. Northrop, P. J. Stang, *Org. Lett.* **2007**, *9*, 1561 – 1564; f) I. Saur, R. Scopelliti, K. Severin, *Chem. Eur. J.* **2006**, *12*, 1058 – 1066; g) G. Ashkenasy, R. Jagasia, M. Yadav, M. R. Ghadiri, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10872 – 10877; h) P. Mukhopadhyay, P. Y. Zavalij, L. Isaacs, *J. Am. Chem. Soc.* **2006**, *128*, 14093 – 14102.
[6] R. F. Ludlow, S. Otto, *Chem. Soc. Rev.* **2007**, DOI: 10.1039/B611921M.
[7] a) P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652 – 3711; b) J. M. Lehn, A. V. Eliseev, *Science* **2001**, *291*, 2331 – 2332.
[8] D. Schultz, J. R. Nitschke, *J. Am. Chem. Soc.* **2006**, *128*, 9887 – 9892.
[9] J. R. Nitschke, D. Schultz, G. Bernardinelli, D. Gérard, *J. Am. Chem. Soc.* **2004**, *126*, 16538 – 16543.
[10] M. Hutin, R. Franz, J. R. Nitschke, *Chem. Eur. J.* **2006**, *12*, 4077 – 4082.
[11] M. Hutin, G. Bernardenelli, J. R. Nitschke, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17655 – 17660.
[12] S. M. Voshell, S. J. Lee, M. R. Gagne, *J. Am. Chem. Soc.* **2006**, *128*, 12422 – 12423.
[13] a) D. H. Busch, *Science* **1971**, *171*, 241 – 248; b) S. Brooker, Y. Lan, J. R. Price, *Dalton Trans.* **2007**, 1807 – 1820.
[14] D. Schultz, J. R. Nitschke, *Angew. Chem.* **2006**, *118*, 2513 – 2516; *Angew. Chem. Int. Ed.* **2006**, *45*, 2453 – 2456.
[15] The number of equivalents of copper(I) ions required to create a set of coordinatively saturated [L_4Cu^I] structures will be equivalent to the total number of nitrogen atoms present in all subcomponents divided by four: $\mathbf{a}/4 + \mathbf{b}/2 + \mathbf{x}/4 + \mathbf{y}/2 + \mathbf{z}/4$.
[16] In order to satisfy Equation (3), the quantity of one of the subcomponents employed must be determined by the quantity of the other four, which eliminates a degree of freedom. An infinite number of solutions exist for Equations (1) and (2), which thus do not eliminate any degrees of freedom from the system.
[17] R. M. Lemmon, *Chem. Rev.* **1970**, *70*, 95 – 109.
[18] J. M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151 – 160.
[19] See molecular machines special issue: *Acc. Chem. Res.* **2001**, *34*, 410 – 522.

