From reactants to products via simple hydrogen-bonding networks: Information transmission in chemical reactions

Giuseppe Brancato*, Frédéric Coutrot†, David A. Leigh†‡, Aden Murphy§, Jenny K. Y. Wong†, and Francesco Zerbetto*‡

*Dipartimento di Chimica ''G. Ciamician,'' Università degli Studi di Bologna, via F. Selmi n.2, 40126, Bologna, Italy; †Department of Chemistry, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JJ, United Kingdom; and §Nanoarchitectonics Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 5, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8565, Japan

Edited by Jack Halpern, University of Chicago, Chicago, IL, and approved February 5, 2002 (received for review December 22, 2001)

The transmission of information is ubiquitous in nature and often occurs through supramolecular hydrogen bonding processes. Here we report that there is a remarkable correlation during synthesis between the efficiency of the hydrogen-bond-directed assembly of peptide-based [2]rotaxanes and the symmetry distortion of the macrocycle in the structure of the final product. It transpires that the ability of the flexible macrocycle-precursor to wrap around an unsymmetrical hydrogen bonding template affects both the reaction yield and a quantifiable measure of the symmetry distortion of the macrocycle in the product. When the yields of peptide rotaxane-forming reactions are high, so is the symmetry distortion in the macrocycle; when the yields are low, indicating a poor fit between the components, the macrocycle symmetry is relatively unaffected by the thread. Thus during a synthetic sequence, as in complex biological assembly processes, hydrogen bonding can code and transmit ''information''—in this case a distortion from symmetry—between chemical entities by means of a supramolecularly driven multicomponent assembly process. If this phenomenon is general, it could have far reaching consequences for the use of supramolecular-directed reactions in organic chemistry.

continuous symmetry measure | hydrogen bonds | molecular recognition | rotaxanes | symmetry distortion

Hydrogen bonding routinely codes and transmits information during the course of biological activity, most celebratedly with nucleic acids, but also in the selective signaling and interactions of proteins, carbohydrates, hormones, and numerous other functional biomolecules (1). Here we show that information contained in hydrogen bonding motifs can be stored over the course of a multistep reaction pathway that starts from simple building blocks and ends in a set of unconventional molecular architectures, namely peptide-based rotaxanes (molecular systems in which a macrocyclic ring is locked onto a linear ''thread'' by two bulky ''stoppers'') (2, 3). Subsequently, the information (symmetry) can be quantitatively extracted (in the form of its distortion) from the solid-state structures of the mechanically interlocked products and related to the yields of the chemical reactions that formed them.

The synthesis and properties of rotaxanes and molecular shuttles based on various dipeptide threads and benzylic amide macrocycles have been described (4–8). In fact, the structural tolerance of the assembly process allows the preparation of rotaxanes derived from oligopeptide sequences of at least 2–5 amino acid residues as long as they contain at least one non-N-terminal glycine residue (9). The five-component ''clipping'' reactions (Fig. 1) produce rotaxanes because cooperative multipoint hydrogen bonding between the open chain precursor **1** (which, in the absence of a suitable template, preferentially adopts a linear *syn–anti* conformation) and the thread **2** promotes a conformational change that brings the reactive end groups in close proximity leading to rapid cyclization of **1**. During the crucial phase of the reaction, the forming macrocycle

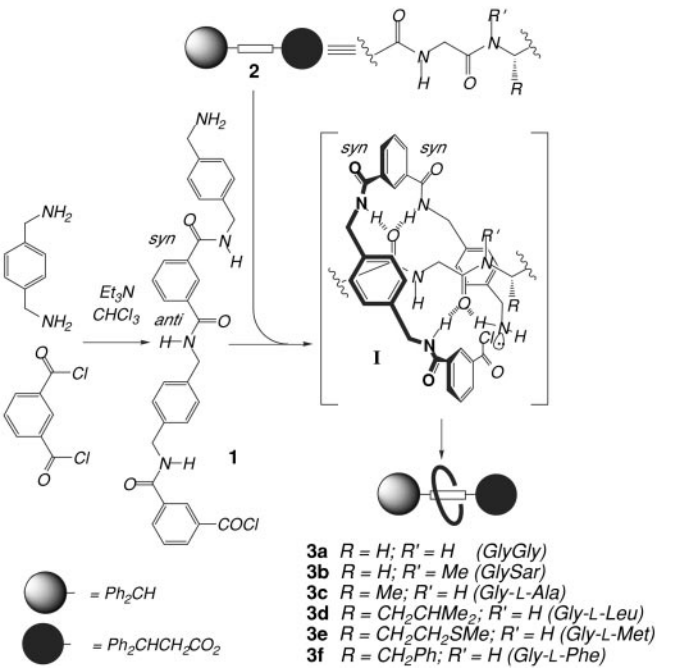

Fig. 1. Rotaxane formation by means of peptide-based hydrogen bond templates. Sar, sarcosine.

establishes a complicated pattern of weaker secondary interactions (including π–π, CH–π, dipole–dipole, van der Waals, etc.) with the template that depend overwhelmingly on the ability of the two amide groups of the thread to coordinate the incoming molecule by means of hydrogen bonding, as shown in supramolecular complex **I**. In general, the better the fit—both electrostatically and sterically—that the interacting surfaces of ''host'' and ''guest'' can adopt as the flexible macrocycle-precursor wraps around the thread in **I**, the higher the yield of the reaction will be.

Apart from being implicated in directing the covalent-bond-forming reactions, multipoint hydrogen bonding between the unsymmetrical thread and the macrocycle in the rotaxane necessarily also distorts the ring from its original isolated symmetry

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: CSM, continuous symmetry measurement.

Data deposition: Crystallographic data for **3a–f** (excluding structure factors) are available from the Cambridge Crystallographic Data Centre as supplementary publication numbers CCDC-141367 (**3a**), CCDC-179-101321 (**3b**), CCDC-147201 (**3c**), and CCDC-141368 to CCDC-141370 (**3d–f**). Copies of the data can be obtained free of charge on application to CCDC, Cambridge CB2 1EZ, United Kingdom (e-mail: teched@chemcrys.cam.ac.uk).

‡To whom reprint requests may be addressed. E-mail: david.leigh@ed.ac.uk or gatto@ciam.unibo.it.

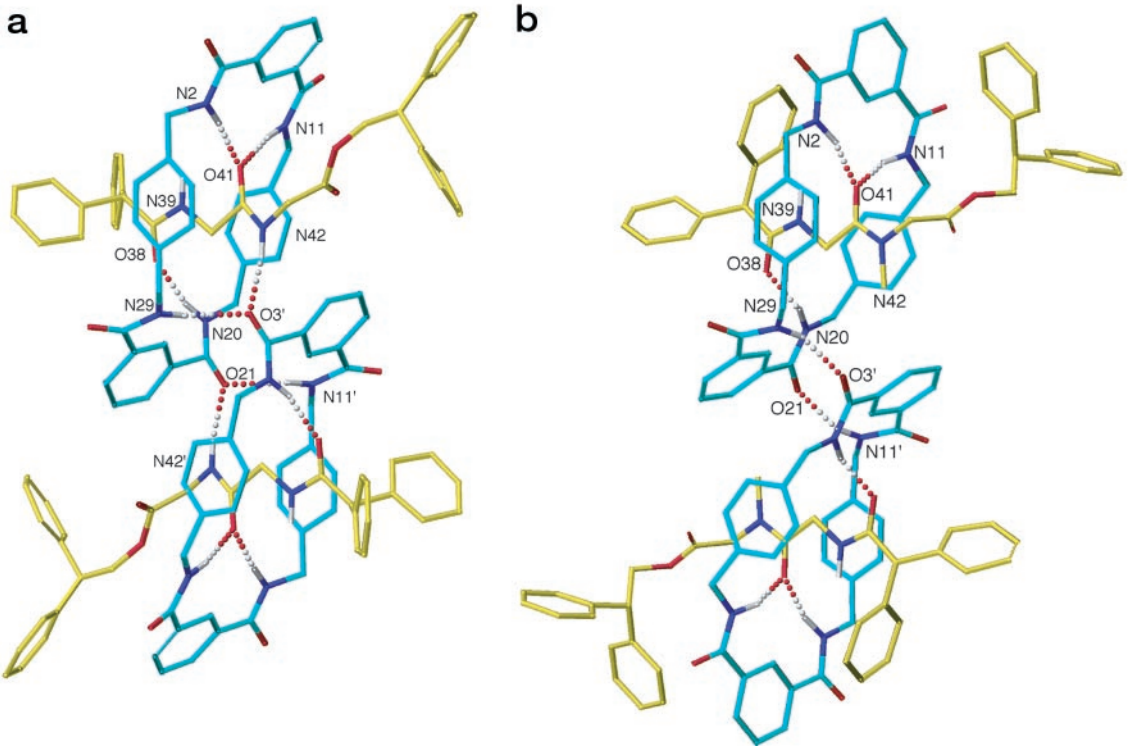

Fig. 2. The solid-state structures of peptido[2]rotaxanes **3a–f** as determined by x-ray crystallography. Carbon atoms of the macrocyclic ring are shown in light blue, and the carbon atoms of the peptide threads are shown in yellow; oxygen atoms are depicted red, nitrogen atoms are depicted in dark blue, and sulfur, green. Non-amide hydrogen atoms have been removed for clarity; those indicated were placed in chemically reasonable positions. (*a*) Gly-Gly rotaxane **3a**, intramolecular hydrogen bond distances: O41–N2 = 3.01 Å; O41–N11 = 3.19 Å; O38–N20 = 2.87 Å; intermolecular hydrogen bond distances: O21–N11′ = O3′–N29 = 3.06 Å; O3′–N42 = O21–N42′ = 2.88 Å; hydrogen bond angles: O41–H–N2 = 174.0°; O41–H–N11 = 170.1°; O38–H–N20 = 165.3°; O21–H–N11′ = O3′–H–N29 = 149.5°; O3′–H–N42 = O21–H–N42′ = 169.1°. (*b*) Gly-Sar rotaxane **3b**, intramolecular hydrogen bond distances: O41–N2 = 3.01 Å; O41–N11 = 3.09 Å; O38–N20 = 2.75 Å; intermolecular hydrogen bond distances: O3′–N29 = O21–N11′ = 2.93 Å; hydrogen bond angles: O41–]H–N2 = 163.5°; O41–H–N11 = 157.5°; O38–H–N20 = 142.7°; O3′–H–N29 = O21–H–N11′ = 147.4°. (*Figure continues on the opposite page.*) (*c*) Gly-L-Ala rotaxane **3c**, intramolecular hydrogen bond distances: O38–N29 = 3.30 Å; O38–N20 = 2.93 Å; O41–N11 = 2.99 Å; O41–N2 = 3.01 Å; hydrogen bond angles: O38–H–N29 = 174.1°; O38–H–N20 = 162.7°; O41–H–N11 = 170.9°; O41–H–N2 = 167.5°. (*d*) Gly-L-Leu rotaxane **3d**, intramolecular hydrogen bond distances: O38–N29 = 2.98 Å; O38–N20 = 3.26 Å; O41–N11 = 3.05 Å; O41–N2 = 3.13 Å; hydrogen bond angles: O38–H–N29 = 156.8°; O38–H–N20 = 153.6°; O41–H–N11 = 172.5°; O41–H–N2 = 162.0°; selected dihedral angles: C7–C8–C10–O10 = 158.0°; C5–C4–C3–O3 = 158.7°; C25–C26–C28–O28 = 155.4°; C23–C22–C21–O21 = 140.2°. (*e*) Gly-L-Met rotaxane **3e**, intramolecular hydrogen bond distances: O38–N29 = 3.18 Å; O38–N20 = 2.90 Å; O41–N11 = 3.10 Å; O41–N2 = 2.97 Å; hydrogen bond angles: O38–H–N29 = 157.5°; O38–H–N20 = 162.3°; O41–H–N11 = 165.4°; O41–H–N2 = 171.7°; selected dihedral angles: C7–C8–C10–O10 = 154.2°; C5–C4–C3–O3 = 150.9°; C25–C26–C28–O28 = 149.6°; C23–C22–C21–O21 = 141.1°. (*f*) Gly-L-Phe rotaxane **3f**, intramolecular hydrogen bond distances: O38–N29 = 3.02 Å; O38–N20 = 3.15 Å; O41–N11 = 2.98 Å; O41–N2 = 3.17 Å; hydrogen bond angles: O38–H–N29 = 160.3°; O38–H–N20 = 149.4°; O41–H–N11 = 169.3°; O41–H–N2 = 164.8°; selected dihedral angles: C7–C8–C10–O10 = 156.0°; C5–C4–C3–O3 = 155.7°; C25–C26–C28–O28 = 143.6°; C23–C22–C21–O21 = 150.7°. Crystallographic data for **3a–f** (excluding structure factors) are available from the Cambridge Crystallographic Data Centre as supplementary publication numbers CCDC-141367 (**3a**), CCDC-179-101321 (**3b**), CCDC-147201 (**3c**), and CCDC-141368 to CCDC-141370 (**3d–f**). Copies of the data can be obtained free of charge on application to CCDC, Cambridge CB2 1EZ, United Kingdom (e-mail: teched@chemcrys.cam.ac.uk).

(for example, hydrogen bonds simultaneously present between the macrocycle and the different amide groups of the thread cannot be of identical energies, and thus the symmetry of the macrocycle is intrinsically broken). Loosely speaking, the stronger the hydrogen bonding between the components, the larger the deformation of the symmetry caused by the thread can be. Synthetic yields and symmetry distortion in the product are therefore influenced by the same phenomenon. A direct relationship between the two, however nontrivial, would imply that the reactions that form the rotaxanes conserve and transmit information from reactants to products. An interesting corollary to this relationship would be that an increase in the hydrogen bonding ability of certain sections of an unsymmetrical template may not necessarily increase a reactions yield; it could be that only the part resulting in symmetry distortion of the final product carries information related to the efficiency of the assembly process. Importantly, such a distortion is amenable to prediction by molecular modeling calculations.

The x-ray crystal structures of six dipeptide rotaxanes (**3a–f**), each containing the same benzylic amide macrocycle, show varying conformations and co-conformations (ref. 9; "co-conformation" refers to the relative positions and orientations of the mechanically interlocked components with respect to each other; ref. 10) of the interlocked components as illustrated in Fig. 2. Despite cyclization of the open-chain precursor **1** occurring about a near-identical template in each case (the variations in the structure of the threads all occur at the periphery of the template region shown in **I**), the yields of the reactions vary from 32% to 62%.

To quantify various aspects of symmetry measures, several approaches have been proposed (11–23). One is the continuous symmetry measure (CSM; refs. 17–23), which has proved suc-

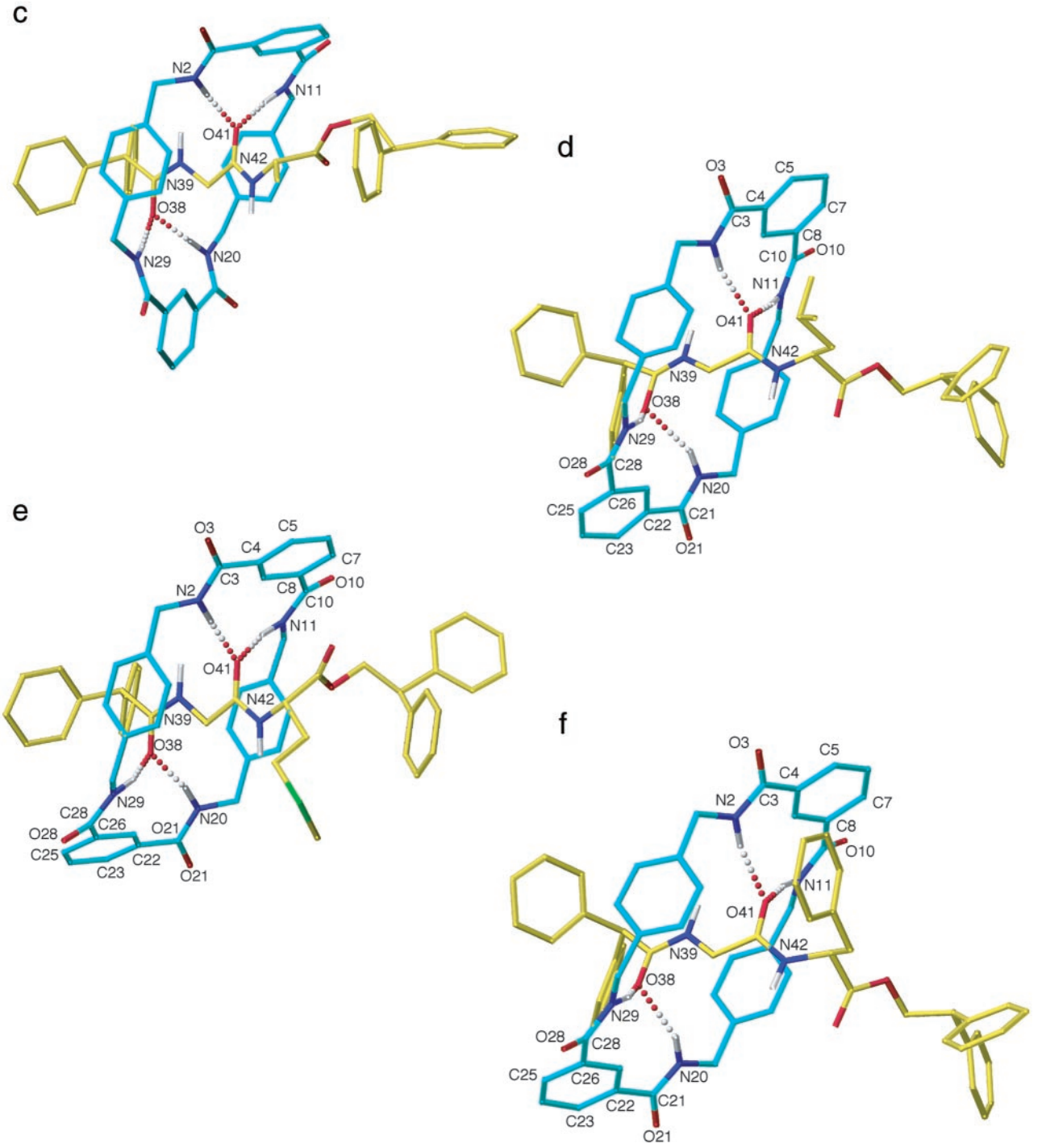

Fig. 2. (*Continued.*)

cessful in a number of cases (17–23). In this approach, Avnir and collaborators calculate CSM or the symmetry deviation, S, as the norm of the vector defined by the difference between the optimal idealized structure—obtained by numerical techniques—and the actual structure. A simple equation connects S to the Cartesian coordinates of the atoms:

$$S = \frac{100}{nA^2} \sum_{i=1}^{n} (p_i - \hat{p}_i)^2, \qquad [1]$$

where n is the number of equivalent points (i.e., atoms) in the symmetry of interest, A is a normalization factor given by the distance between the center of mass of the molecule and its own farthest atom, p_i is the coordinates of the atoms, and $\hat{p}_i$ is the coordinates of the atoms of the optimal structure that are determined by numerical optimization techniques (one for each structure). In this way, in agreement with the approach proposed by Avnir (17–23), S does not measure the distance to a predetermined general structure but to the closest symmetry sought. $S = 0$ coincides with perfect symmetry, i.e., no symmetry lowering, and $S = 100$ collapses the system to a single point.

Table 1. CSM, *S*, and reaction yields, *Y*, for dipeptide [2]rotaxanes 3a–f

	S	*Y*, %
3a Gly-Gly	0.17	62
3b Gly-Sar	0.18	60
3c Gly-L-Ala	0.13	45
3d Gly-L-Leu	0.02	37
3e Gly-L-Met	0.03	36
3f Gly-L-Phe	0.02	32

The benzylic amide macrocycle present in the peptide rotaxanes **3a–f** can exist in boat (C_{2v}), chair (C_{2h}), or half-chair/twist-boat conformations (see Fig. 2). Evaluation of the symmetry measure in different point groups, i.e., with respect to different symmetry operations, was thought likely to give quantities that would not be directly comparable. For example, the chair conformation has an inversion center, which is absent in the boat structure. Forcing its use would have made *S* unrealistically large for the boat conformation. It was therefore decided to limit the calculation of *S* to the twofold axes that (*i*) are roughly parallel to the thread, and (*ii*) pass through the *p*-xylylene fragments. The difference between the two is the preference for the ring to adopt (*i*) the boat or (*ii*) the chair-like conformation. As seen from the x-ray structures, both these conformations are accessible in the rotaxane architecture but have different consequences for the intercomponent hydrogen bonding geometries that can be adopted.

Table 1 shows the *S* values for the macrocycle in the six rotaxane crystal structures. In three cases, **3a–c**, the symmetry distortion is substantial. Since the thread possesses only trivial symmetry, the values are a consequence of the effectiveness with which the macrocycle wraps around the hydrogen-bonding template. The three rotaxane systems with smaller *S* values can therefore be considered to contain threads that are intrinsically less well suited to interacting positively with the macrocyclic ring.

Examination of the six crystal structures helps explain this pattern and shows the presence of different types of hydrogen bonds (24), "standard" two-center systems and "bifurcated" three-center bonds. In two cases, **3a** and **3b**, there are also intermolecular hydrogen bonds seen in the solid state that cannot be involved in the mechanism of formation of "isolated" molecules in solution (see below). Importantly, the rotaxane architecture forces the intramolecular bifurcated bonds to be orthogonal to the lone pairs of the acceptor, resulting in their being somewhat weaker than conventional hydrogen bonds where the acidic hydrogen atoms on nitrogen point directly at the oxygen lone pairs (i.e., the region of highest electron density) (this motif is also seen in the x-ray crystal structures of various benzylic amide macrocycle-based catenanes; see ref. 25).

In more detail: (*i*) The Gly-Gly site (rotaxane **3a**) is the least sterically hindered for the macrocycle and also the most flexible, allowing it to adopt the most favorable hydrogen bonding geometries in solution or the solid state. The macrocycle adopts a chair conformation, and the crystal structure shows one standard and one pair of bifurcated hydrogen bonds (plus the intermolecular one), all in the range of 2.87–3.19 Å and 165.3–174.0°, typical geometries for strong amide–amide hydrogen bonds (26); (*ii*) the Gly-Sar (**3b**) system displays a similar hydrogen bonding pattern to Gly-Gly, but the extra methyl group distorts the macrocycle into a less favorable half chair/twist boat-like structure; (*iii*) the Gly-L-Ala rotaxane (**3c**) has two pairs of intercomponent bifurcated (i.e., weaker) hydrogen bonds, and the macrocycle adopts a boat conformation to form medium length (2.93–3.30 Å) but reasonably linear NH···O hydrogen bonds in the range 162.7–174.1°; (*iv*) the last three

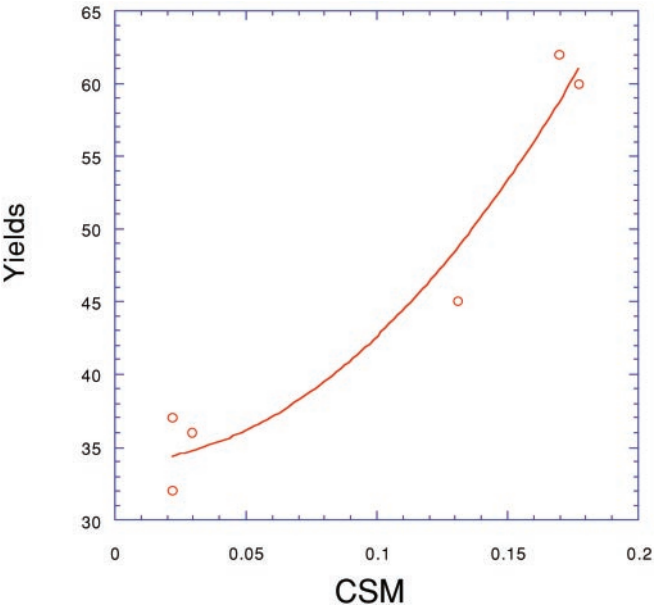

Fig. 3. The relationship between the CSM from the solid-state structure of the peptido[2]rotaxanes and the five-component reaction yield.

systems (**3d–f**) also possess two pairs of bifurcated hydrogen bonds, but the larger steric bulk at their chiral centers prevents the macrocycle from adopting a boat conformation from which it could form the same type of hydrogen bond geometries as **3c**. Thus, although the macrocycle adopts a relatively unstrained chair conformation, this geometry means that several of the NH···O hydrogen bonds have to be longer than normal and distorted from linearity (e.g., N20···O38 = 3.26 Å, N20H···O38 = 153.6° **3d**; N29···O38 = 3.18 Å, N29H···O38 = 157.5° **3e**; N20···O38 = 3.15 Å, N20H···O38 = 149.4° **3f**) and therefore relatively weak. Moreover, to form these hydrogen bonds at all, some of the amide groups of the macrocycle have to twist out of planarity with the benzene rings of the isophthaloyl unit, decreasing the π-electron conjugation (e.g., C23–C22–C21–O21 = 140.2° **3d**, C23–C22–C21–O21 = 141.1° **3e**, C25–C26–C28–O28 = 143.6° **3f**). Comparison of the hydrogen bonding patterns with the symmetry-lowering values, *S*, thus confirms that stronger intercomponent interactions in the solid state do correspond to larger symmetry distortions in the macrocycle. This finding is consistent with the idea that the better the ability of the macrocycle to wrap around the thread and form a close fitting, reciprocal pattern of interactions, the more effective it is at following the electronic and steric "contours" of the thread, adopting a complementary shape to its partner and a distortion in its symmetry.

The synthetic yields of the rotaxane-forming reactions are shown in Table 1.[¶] Given the many factors that can contribute to both reaction yields (e.g., solubility of products and intermediates, and different degrees of intramolecular hydrogen bonding in the threads) and crystal structure geometries (solvent, temperature, and concentration at which the crystals were grown), the correlation between the CSM of the solid-state structures and the yields is remarkable and, in particular, implies a close similarity between the structures of the final products and the intermediate supramolecular complex **I**, which leads to the transition state for ring closure.

[¶]The rotaxane-forming reactions were carried out under identical conditions (6, 9) and repeated at least three times. The yields of individual experiments were remarkably reproducible, all falling within a 2% range.

The yields are probably dependent on the relative strengths of the three amide–amide hydrogen bonds between the open-chain precursor **1** and the thread, shown in **I** (the fourth hydrogen bond, between an amine and an amide, will be much weaker because the hydrogen bond acidity, α_2^H, of amines is in the range 0.00–0.26, compared with the typical α_2^H of a secondary amide of 0.38) (27). The two rotaxanes formed in highest yield (**3a** and **3b**, 62% and 60%), thus each have one standard intercomponent hydrogen bond (strong), two bifurcated intercomponent hydrogen bonds (less strong), and a symmetry-lowering *S* of 0.17 and 0.18, respectively. The Gly-L-Ala rotaxane, **3c**, formed in a lower yield (45%), has only bifurcated hydrogen bonds (weak, but linear because it can adopt a boat conformation) and a CSM value of 0.13. The three rotaxanes formed in the most modest yields (32–37%), **3d–f**, each have two pairs of relatively weak bifurcated hydrogen bonds and a CSM value less than 0.04. Ironically, the three threads that have the bulkiest amino acid side chains (i.e., the most pronounced asymmetric centers) give the smallest symmetry distortions in the macrocycle; the reason behind this, the poor fit between the macrocycle and thread, also accounts for their modest yields.

Fig. 3 further illustrates the relationship between CSM and the reaction yields. Although the two-parameter parabola (without the linear term) has a high r^2 value (>0.95), its presence is mainly to assist the eye. [Nonlinear trends of *S* versus physical quantities have been observed (15, 16). It is, however, intriguing to notice that molecular distortions along either a vibrational motion or a bond tend to increase the energy quadratically. For discussion of this point see ref. 28.] CSM, in the special form of continuous chirality measure (21), has been correlated to other dynamic and energy-related quantities such as the binding activity of trypsin/arylammonium inhibitors, D_2-dopamine receptor/dopamine derivative agonists, trypsin/organophosphates inhibitors, acetylcholinesterase/organophosphate and butyrylcholinesterase/organophosphates (21), and has been used to establish relationships between chirality content and stereoinduction (22). In the latter case, three-term parabolas were used to fit computed chirality contents versus enantiomeric excesses, and the r^2 values found there, 0.63–0.92, are comparable to that obtained above.

The fundamental picture that emerges from this study is of a system in which the ability of a flexible macrocycle precursor to wrap around an unsymmetrical hydrogen bonding template determines, and thus relates, both reaction yield and the symmetry distortion in the product. When the yields of peptide rotaxane-forming reactions are high, so is the symmetry distortion in the macrocycle; when the yields are low, indicating a poor fit between the components, the macrocycle symmetry is relatively unaffected by the thread. Thus in synthetic chemistry, as in complex biological systems, hydrogen bonding can code and transmit information between chemical entities by a supramolecularly driven multicomponent assembly process. The case reported here is, however, unique both in terms of its simplicity and the way the information, i.e., symmetry, is stored and subsequently extracted. If this result proves to be a general phenomenon, it could have far-reaching consequences for the use of supramolecular-directed covalent bond-forming reactions in organic synthesis.

1. Jeffrey, G. A. & Saenger, W. (1991) *Hydrogen Bonding in Biological Structures* (Springer, New York).
2. Amabilino, D. B. & Stoddart, J. F. (1995) *Chem. Rev.* **95,** 2725–2828.
3. Sauvage, J.-P. & Dietrich-Buchecker, C., eds. (1999) *Molecular Catenanes, Rotaxanes and Knots* (Wiley/VCH, Weinheim, Germany).
4. Leigh, D. A., Murphy, A., Smart, J. P. & Slawin, A. M. Z. (1997) *Angew. Chem. Int. Ed. Engl.* **36,** 728–732.
5. Lane, A. S., Leigh, D. A. & Murphy A. (1997) *J. Am. Chem. Soc.* **119,** 11092–11093.
6. Clegg, W., Gimenez-Saiz, C., Leigh, D. A., Murphy, A., Slawin, A. M. Z. & Teat, S. J. (1999) *J. Am. Chem. Soc.* **121,** 4124–4129.
7. Leigh, D. A., Troisi, A. & Zerbetto, F. (2000) *Angew. Chem. Int. Ed.* **39,** 350–353.
8. Wurpel, G. W. H., Brouwer, A. M., van Stokkum, I. H. M., Farran, A. & Leigh D. A. (2001) *J. Am. Chem. Soc.* **123,** 11327–11328.
9. Asakawa, M., Brancato, G., Fanti, M., Leigh, D. A., Shimizu, T., Slawin, A. M. Z., Wong, J. K. Y., Zerbetto, F. & Zhang S. (2002) *J. Am. Chem. Soc.*, in press. (Published on the Web ASAP contents March 1, 2002.)
10. Fyfe, M. C. T., Glink, P. T., Menzer, S., Stoddart, J. F., White, A. J. P. & Williams, D. J. (1997) *Angew. Chem. Int. Ed. Engl.* **36,** 2068–2069.
11. Murray-Rust, P., Rüsti, H. B. & Dunitz, J. (1978) *Acta Crystallogr. B* **34,** 1787–1793.
12. Gilat, G. (1985) *Chem. Phys. Lett.* **121,** 9–12.
13. Rassat, A. (1984) *C. R. Acad. Sci. Paris II* **299,** 53–55.
14. Maruani, J. & Mezey, P. G. (1990) *Mol. Phys.* **69,** 97–113.
15. Ruch, E. (1972) *Acc. Chem. Res.* **5,** 49–56.
16. Buda, A. B. & Mislow, K. (1992) *J. Am. Chem. Soc.* **114,** 6006–6012.
17. Zabrodsky Hel-Or, H., Peleg, S. & Avnir, D. (1992) *J. Am. Chem. Soc.* **114,** 7843–7851.
18. Zabrodsky Hel-Or, H., Peleg, S. & Avnir, D. (1993) *J. Am. Chem. Soc.* **115,** 8278–8289.
19. Zabrodsky Hel-Or, H., Peleg, S. & Avnir, D. (1993) *J. Am. Chem. Soc.* **115,** 11656–11656.
20. Zabrodsky Hel-Or, H., Peleg, S. & Avnir, D. (1995) *J. Am. Chem. Soc.* **117,** 462–473.
21. Krinan, S. & Avnir, D. (1998) *J. Am. Chem. Soc.* **120,** 6152–6159.
22. Gao, D., Schefzick, S. & Lipkowitz, K. B. (1999) *J. Am. Chem. Soc.* **121,** 9481–9482.
23. Lipkowitz, K. B., Schefzick, S. & Avnir, D. (2001) *J. Am. Chem. Soc.* **123,** 6710–6711.
24. Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding* (Oxford Univ. Press, Oxford).
25. Johnston, A. G., Leigh, D. A., Pritchard, R. J. & Deegan, M. D. (1995) *Angew. Chem. Int. Ed. Engl.* **34,** 1209–1212.
26. Taylor, R. & Kennard, O. (1984) *Acc. Chem. Res.* **17,** 320–326.
27. Abraham, M. H. (1993) *Chem. Soc. Rev.* **22,** 73–83.
28. Brancato, G. & Zerbetto, F. (2001) *J. Phys. Chem. A* **104,** 11439–11442.