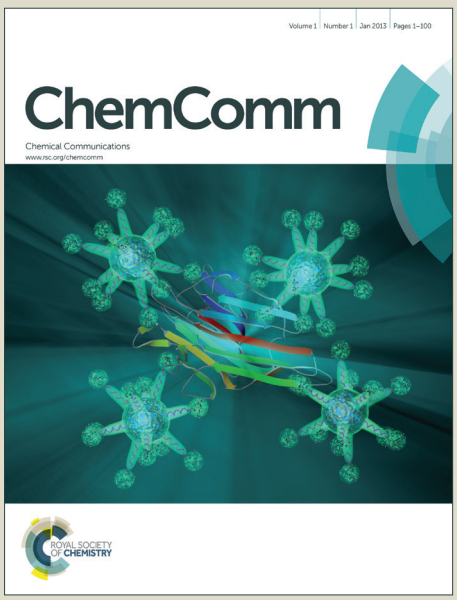
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Constitutional self-selection from dynamic combinatorial libraries in aqueous solution through supramolecular interactions

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

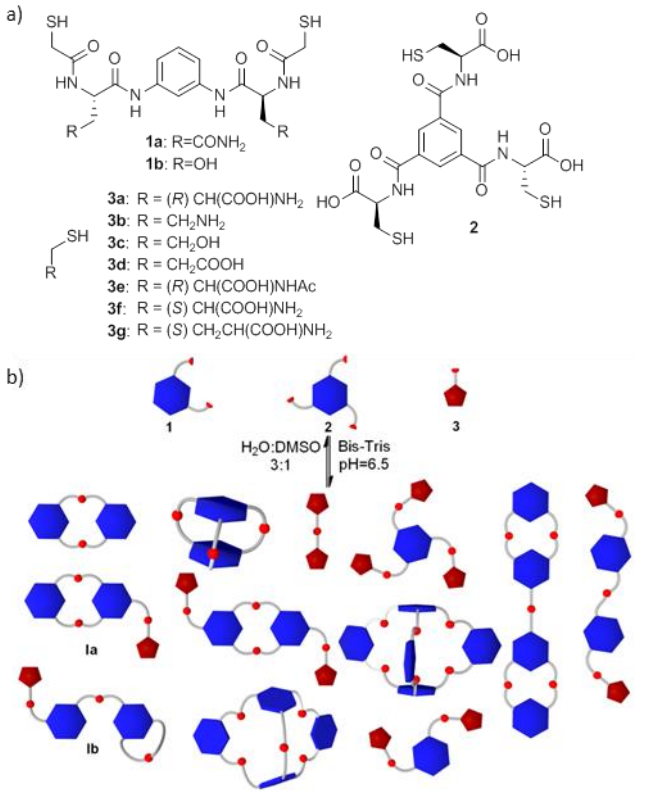
www.rsc.org/

Jordi Solà*, Maria Lafuente, Joan Atcher, and Ignacio Alfonso*

We describe the predominant formation of a specific constitution arising from the combination of building blocks with different topologies through disulphide chemistry in a Dynamic Combinatorial Library (DCL). The supramolecular interactions established by a zwitterionic cysteine moiety are responsible for the self-selection of one product among all virtual members of a large library.

Self-assembling and self-organization are inherent processes in Nature that occur by the spontaneous folding and assembling of relatively simple chemical entities.[1] For that, the delicate inter- and intramolecular interactions are established and regulated by proof-reading and self-correction.[2] Accordingly, the chemical information stored in a given chemical structure is decoded to render beautifully complex and fully functional biomolecular systems. For decades, supramolecular chemists have been fascinated and inspired by this biomolecular machinery. More recently, Constitutional Dynamic Chemistry (CDC) and Dynamic Combinatorial Chemistry (DCC) have proposed the implementation of complex dynamic systems formed by members able to exchange under thermodynamic control.[3] These dynamic systems (Dynamic Combinatorial Libraries, DCLs) are excellent benchmark models for decoding the chemical information stored in the combination of simple building blocks towards the emergence of new assembled structures.[4]

Disulphide chemistry has been exploited widely for the generation of DCLs in aqueous media.[5] However, most of the systems described are based on molecules with two reacting sites (bipodal). Thus, although some impressive (and somehow unexpected) geometries have been described such as catenanes and knots,[6] the tri-dimensional space covered is relatively small, and the libraries are mostly directed towards the generation and interconversion of cyclic oligomers.[7] The formation of cage-like organic structures has seldom been described by oxidation of tripodal building blocks or by combination of tripodal and bipodal building blocks.[8] The stoichiometry and the presence of a template were critical in the latter. More recently Sanders has described the diverse topologies obtained by combination of tri- and monothiols in water.[9] In order to expand the structural and topological diversity of supramolecular assemblies by DCC we decided to study the effect of combining mono-, di- and trisulphides in the same DCL. In principle, by combining a series of building blocks of different topology, a broad structural variety is expected. We chose to combine pseudopeptidic C_2-symmetric dithiols **1a-b** described earlier by our group[10] with the tripodal building block **2**[8] and a series of monothiols **3a-g**. Thus, one may expect the formation of several compounds with different architectures some of which are schematically represented in Scheme 1b.



Scheme 1. a) Thiol building blocks used in this study. b) Schematic representation of some of the possible architectures generated by the combination of topologically different building blocks.

We started by mixing trithiol **2** with neutral dithiols **1a** or **1b** at pH 6.5 in an aqueous solution containing 25% DMSO. The use of DMSO allows the faster oxidation and exchange within a few hours even at a slightly acidic pH.[11] The mixtures were analysed after 24h and 48h with no significant changes. Several compounds can be detected in accordance with the results published in literature.[8-10] Our DCLs, however, presented less variety of cyclic products arising from oxidation of dithiols **1a-1b** (Fig. 1a, blue trace) as they prefer to form mixed products with **2** (presumably cage-like compounds of formula 1_22_2 and 1_32_2)

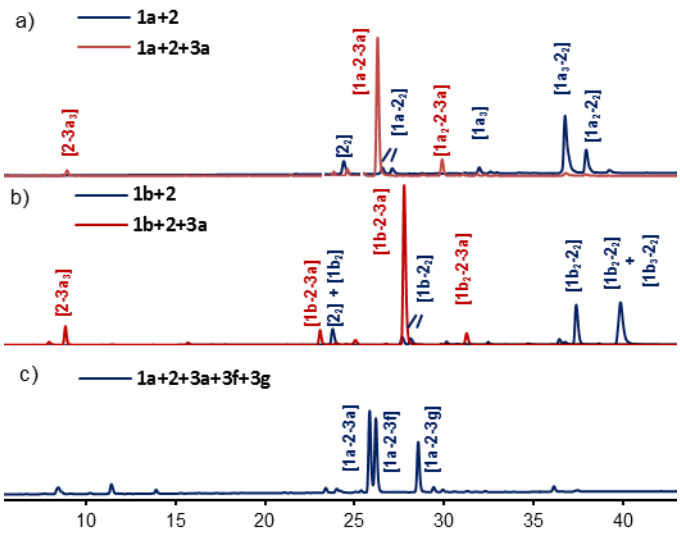


Fig. 1. a, b) DLCs formed by **1a,b** and **2** (0.5 mM each) in aqueous solution (25% DMSO, pH 6.5) in the presence (red) or absence (blue) of cysteine (2.5 mM). c) DLC formed by **1a** (0.5 mM) and **2** (0.5 mM) in the presence of L-Cys, D-Cys and homo-Cys (1.0 mM each).

To our surprise, when the same experiment was repeated for **1a** and **2** in the presence of 2.5 mM of cysteine (enough to saturate all free thiols) a product containing each of the three building blocks (**1a23a**) was predominantly formed (Fig. 1a, red trace). The other minor detectable species are the combination of **2** with 3 cysteines and a structure of the formula **1a$_2$23a**. Since Ser-derivative **1b** presented similar behaviour (figure 1b) we focused our studies on Asn-derivative **1a**. The distribution of products was mostly unchanged when varying the concentration of Cys from 0.5 mM (equimolar with other components) to up to 100 mM, showing the stability of the product formed.[12] In addition, we combined building blocks **1a** and **2** in different proportions (1:2 and 2:1) in the presence of 1 eq of Cys per thiol. In these experiments almost all the limiting reagent was consumed in the same product of formula **1a23a** whereas the reagent in excess formed either adducts with Cys or the corresponding homodimer (ESI†). We attribute the selection of the major product formed in the mixture to some sort of self-recognition within the building blocks during the assembly process. Thus, stabilization through supramolecular interactions would lead to the formation of a major, more stable compound. Reversibility tests confirmed the reaching of the thermodynamic equilibrium: addition of **3a** after 24h to a sample containing **1a** and **2** gave essentially the same mixture as the one formed by **1a**+**2**+**3a** from the beginning (ESI†).

In order to ascertain what structural elements were crucial for the selection, other monothiols (**3b-e**) were tested in the libraries. All of the thiols performed notably worse than Cys in selecting the corresponding major product. These experiments showed that both the ammonium and the carboxylate groups are needed in order to successfully select a particular structure, with the ammonium group being more critical, as a hydrogen bond donor in that position clearly enhances recognition. In fact, the experiments done with thiols **3a-e** showed the trend NH_3^+>OH>NHAc (ESI†) to effectively select a predominant constitution from the mixture.

We turned then to investigate the structure of the multicomponent molecule to see what possible interactions would lead to the formation of the major product. In principle two architectures are possible for a molecule of the formula **1a23a**: the symmetrical structure **Ia** and the less symmetrical **Ib** (Scheme 1). We considered **Ia** more likely as it would present a more compact structure in solution.

To confirm this possible arrangement we reproduced the synthesis of **1a23a** on a preparative scale. The NMR spectra of the purified product in buffered water at pH 6.5 with 15% DMSO-d_6 were acquired using water signal suppression by excitation sculpting (Fig. 2a). All the signals were successfully assigned by combination of ^{1}H-NMR and 2D-TOCSY experiments. All expected protons are visible in the ^{1}H-NMR spectra apart from the three protons corresponding to the alpha carbons of the amino acids from bipodal and tripodal fragments, which fall in the region of water suppression. However, these protons are visible in the TOCSY experiment. Three doublets corresponding to the three amide bonds coupled to the chiral centre of the amino acids are clearly visible in a 2:2:1 ratio (H_A, H_B, H_E). In addition, protons H_C and H_{18} from the phenylenediamine fragment, and H_4 and H_2 from the trimesic acid appear as singlets. Overall, this data is consistent only with the formation of the more symmetrical product **Ia**.

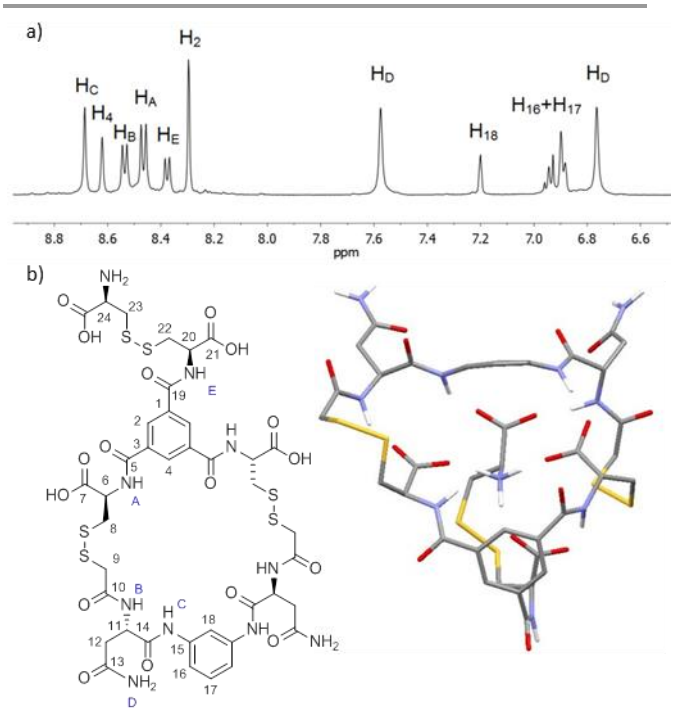


Figure 2. a) Partial ^{1}H-NMR spectrum (500 MHz, buffered H_20 with 15% DMSO-*d6*, 298 K) showing the aromatic and amide region for molecule **1a23a**. b) Proposed structure for major product **1a23a** (**Ia**) with atom labelling (left) and structure obtained by molecular modelling (right).

Some other structural considerations can be done in view of the ^{1}H-NMR spectra. H_E displays a relatively low chemical shift

compared to H_A, which could be indicative of a twisting out of the aromatic ring plane. This also explains the large chemical shift difference between H_2 and H_4, both from the tripodal aromatic ring. On the other hand, H_C and H_B appear in the downfield amide region, which is consistent with some kind of hydrogen bonding phenomena, most likely with any of the carboxylate anions present in the molecule. To shine some light in the possible tri-dimensional structure of the compound in solution we performed a conformational search using MMFFaq force field. Molecular modelling rendered a folded structure with several stabilizing interactions as shown in Fig. 2b.

The most stable conformer shows a structure where the pendant Cys folds over the macrocyclic cavity. The ammonium cation resides in the centre of the macrocycle and is stabilized by the three carboxylates that arise from the trimesic acid moiety. The carboxylate group of the cysteine residue would as well be stabilized by hydrogen bonding interactions with the macrocycle amide protons (H_C/H_B). Other hydrogen bonding interactions between the macrocyclic peptidic amides and the carboxylates would further stabilise the structure. Interestingly a *cis* amide bond is observed for the tripodal amide that is not part of the macrocycle. We can also observe different disposition (in/out) for the carbonyls of the amides both in the bipodal and tripodal moieties. This data is in agreement with what we observed by 2D-NOESY experiments. NOE correlations were visible for proton H_A with both H_4 and H_2 which would indicate a dynamic rotation of the amide bonds. Similar behaviour is observed for H_C and H_{18},H_{16} showing that some degree of conformational freedom is present.

Unluckily, we could not assign unambiguous through-space correlations that supported the proposed folding, since proton H_{23} (from the pendant cysteine) overlaps with protons H_9 (within the macrocycle backbone). Besides, as stated before, protons H_6, H_{11} and H_{19} fall in the water suppression region and thus, no NOE correlations could be assigned for these protons neither. Nevertheless, further evidence of the folded structure was obtained when the sample was acidified with TFA (10 eq), because some significant shifts of the NMR signals were observed (ESI†). Thus, H_E and H_A moved downfield and to a similar chemical shift while H_B and H_4 moved upfield. This is consistent with an unfolding of the structure upon protonation, which would leave all the amide or aromatic protons of the tripodal moiety in a more similar chemical environment.

Finally, to prove that small structural changes had a decisive effect on the product distribution we tested the unnatural enantiomer of cysteine (D-Cys, **3f**) and structurally related homocysteine (**3g**). D-Cys formed a major product containing all the building blocks with no major differences with L-Cys. Homocysteine, however, formed a major product **1a23g**, but other compounds were also readily detected. Strikingly, in a competition experiment L-Cys was slightly preferred over D-Cys (51:49, ESI†) and in a competition experiment between L-, D-, and homocysteine the tendency showed that the preference is L>D>homo (39:37:24) demonstrating that subtle structural differences have an impact on the formation of the major species (see figure 1c). This provides further evidence of a folded structure in solution as homocysteine would have a looser fit.

Conclusions

We show the importance of the synergic action of subtle recognition events within the components of a DCL. Combination of building blocks of different topology could result in a plethora of possible structures. However, cooperative supramolecular interactions between the building blocks result in the self-recognition and the selective amplification of a singular constitution even when it is statistically disfavoured. The structural importance of these interactions is proved by the difference observed in the chromatographic profile of the DCL when delicate changes are introduced. We expect that our findings will improve the comprehension of the behaviour of complex dynamic libraries of compounds.

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO, CTQ2012-38543-C03-03) and EU (FP7-PEOPLE-2012-CIG-321659). Personal financial support for J.S. (MINECO, Ramón y Cajal contract), M.L. (MINECO, FPI fellowship) and J.A. (CSIC and European Social Fund, JAE-predoc fellowship) are gratefully acknowledged. We thank Dr. Yolanda Pérez for helpful assistance with the NMR experiments.

Notes and references

Department of Biological Chemistry and Molecular Modelling, IQAC-CSIC, Jordi Girona 18-26, Barcelona, Spain.
E-mail: jordi.sola@iqac.csic.es; ignacio.alfonso@iqac.csic.es
Fax: +34 932 045 904; Tel: +34 934 006 100

† Electronic Supplementary Information (ESI): Experimental details, synthetic procedures, additional HPLC traces and NMR spectra is available See DOI: 10.1039/c000000x/

1. G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418-2421; J.-M. Lehn, *Angew. Chem., Int. Ed.*, 2013, **52**, 2836-2850.
2. J.-M. Lehn, *Chem. Soc. Rev.*, 2007, **36**, 151-160; J.-M. Lehn, *Angew. Chem. Int. Ed. Engl.*, 1990, **29**, 1304-1319.
3. R. F. Ludlow and S. Otto, *Chem. Soc. Rev.*, 2008, **37**, 101-108; J. Li, P. Nowak and S. Otto, *J. Am. Chem. Soc.*, 2013, **135**, 9222-9239; J. R. Nitschke, *Nature*, 2009, **462**, 736-738; G. M. Whitesides and R. F. Ismagilov, *Science*, 1999, **284**, 89-92; F. B. L. Cougnon and J. K. M. Sanders, *Acc. Chem. Res.*, 2011, **45**, 2211-2221; P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.* 2006, **106**, 3652-3711.
4. S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders and J. F. Stoddart, *Angew. Chem., Int. Ed.*, 2002, **41**, 899-952; S. Otto, *Curr. Opin. Drug Discovery Dev.*, 2003, **6**, 509-520; I. Huc and J.-M. Lehn, *Actual. Chim.*, 2000, 51-54; R. Custelcean, *Top. Curr. Chem.*, **322**, 193-216; E. Moulin, G. Cormos and N. Giuseppone, *Chem. Soc. Rev.*, 2012, **41**, 1031-1049; N. Giuseppone, *Acc. Chem. Res.*, 2012, **45**, 2178-2188; S. Otto, *Acc. Chem. Res.*, 2012, **45**, 2200-2210.
5. S. P. Black, J. K. M. Sanders and A. R. Stefankiewicz, *Chem. Soc. Rev.*, 2014, DOI: 10.1039/C3CS60326A.
6. M.-K. Chung, S. J. Lee, M. L. Waters and M. R. Gagne, *J. Am. Chem. Soc.*, 2012, **134**, 11430-11443; F. B. L. Cougnon, N. Ponnuswamy, N. A. Jenkins, G. D. Pantos and J. K. M. Sanders, *J. Am. Chem. Soc.*, 2012, **134**, 19129-19135; N. Ponnuswamy, F. B. L. Cougnon, J. M. Clough, G. D. Pantoş and J. K. M. Sanders, *Science*, 2012, **338**, 783-785.
7. M. Rauschenberg, S. Bomke, U. Karst and B. J. Ravoo, *Angew. Chem., Int. Ed.*, 2010, **49**, 7340-7345; M. Eisenberg, I. Shumacher, R. Cohen-Luria and G. Ashkenasy, *Bioorg. Med. Chem.*, 2013, **21**, 3450-3457.
8. K. R. West, K. D. Bake and S. Otto, *Org. Lett.*, 2005, **7**, 2615-2618; A. R. Stefankiewicz, M. R. Sambrook and J. K. M. Sanders, *Chem. Sci.*, 2012, **3**, 2326-2329.
9. A. R. Stefankiewicz and J. K. M. Sanders, *Chem. Commun.*, 2013, **49**, 5820-5822.
10. J. Atcher, A. Moure and I. Alfonso, *Chem. Commun.*, 2013, **49**, 487-489.
11. J. Atcher and I. Alfonso, *RSC Adv.*, 2013, **3**, 25605-25608.
12. For concentrations of cysteine higher than 10 mM a precipitate appears. However HPLC analysis of the mixture has little change in absorbance showing that the major part of the precipitate is cystine, which is invisible at 254 nm.