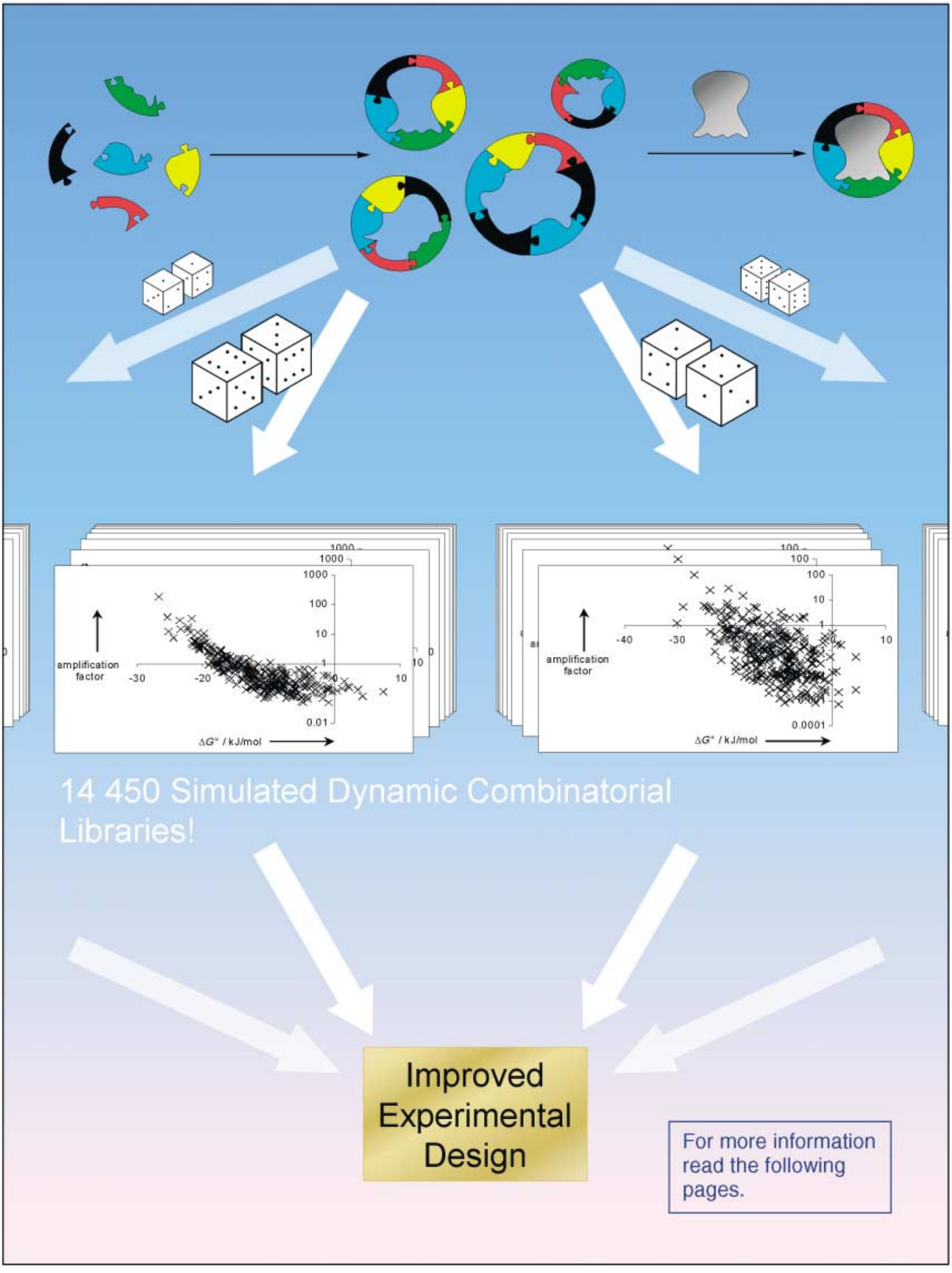

 DOI: 10.1002/chem.200400300

Correlation between Host–Guest Binding and Host Amplification in Simulated Dynamic Combinatorial Libraries

Peter T. Corbett, Sijbren Otto,* and Jeremy K. M. Sanders*[a]

Abstract: We present a versatile computer model of diverse dynamic combinatorial libraries, and examine how molecular recognition between library members and a template can be used to amplify the best binders. The correlation between host–guest binding and amplification was examined for a set of 50 libraries with >300 components each over a wide range of template and building block concentrations. Depending on these concentrations correlations vary from poor (when using a large excess of template) to good (for very dilute libraries and/or substoichiometric template concentrations), highlighting the need to choose the experimental conditions for dynamic combinatorial libraries thoughtfully.

Keywords: combinatorial chemisty · computer simulations · host–guest systems · templated synthesis · thermodynamics

Introduction

Dynamic combinatorial chemistry (DCC) combines synthesis and screening in a single-step approach to the discovery of new functional molecules.[1,2] A set of building blocks is allowed to combine through reversible bond formation, creating a library that is under thermodynamic control. Addition of a template causes stabilization and hence amplification of the host compounds[3] that bind effectively to the template. The key concept, that amplification and host–guest binding strength are strongly correlated may be intuitively reasonable but is experimentally untested.

Using computer simulations, Severin et al.[4] have constructed small example DCLs where the correlation between amplification and binding strength is very poor, posing a challenge to the usefulness of the DCL technique. We describe here a versatile computer simulation approach to explore this phenomenon in large DCLs. Whereas the work by Severin et al. highlights a set of selected cases where amplification and binding strength do not correlate, our work systematically varies "experimental conditions" and maps under which conditions the correlation between binding strength and amplification is acceptable.

Severin's simulation considers a library of trimers made from equimolar amounts of three building blocks: *A*, *B* and *C*. If two of the trimers—*AAA* and *ABC*—are both stabilized by a factor of 1000,[5] then only the *ABC* trimer will be amplified, and the concentration of *AAA* will actually go down. In this example, amplification efficiency and stabilisation are not well correlated. This observation challenges the generality and practicality of the dynamic combinatorial library approach, but how serious and how general is this challenge? How well do the amplifications in an average library reflect the host–guest binding constants? Can libraries be set up under conditions where good correlation between amplification efficiency and binding constants can be guaranteed beforehand? These are questions that we will address herein.

Gaining statistically significant answers to these questions requires the comparison of binding constants and amplification efficiencies for a large number of libraries. Since obtaining the necessary information experimentally is a formidable task for even a single library, the only practical approach is to use computer simulations.

Methods of Calculation

We simulated[6] dynamic combinatorial libraries made using an arbitrarily chosen number of seven different building blocks[7]—*A* to *G*—which were allowed to assemble into all possible oligomers up to tetramers,[8] giving 28 dimers, 84 trimers and 210 tetramers.[9] The exchange processes in the library were simulated using a set of equilibrium constants such that the concentrations of library members matched the statistical preference for hetero-oligomers over homo-oligomers. A preference for dimers

[a] P. T. Corbett, Dr. S. Otto, Prof. J. K. M. Sanders
Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge CB2 1EW (UK)
Fax: (+44) 1223-336017
E-mail: jkms@cam.ac.uk

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

DOI: 10.1002/chem.200400300

over trimers and tetramers was also programmed, to reflect the behaviour of a library that does not result in polymer formation.[10]

A binding constant (K) was randomly assigned to each host, drawing from a normal distribution of $\log_{10}K$, with a standard deviation of 1 and a mean of 2. Lognormal distributions of binding constants have been used in a previous theoretical study of dynamic combinatorial chemistry,[11] and appear to be good descriptors of ensembles of biological receptors.[12] The standard deviation of 1 (or values close to it) is commonly used, whereas literature values[11,12] of the mean in these examples range from zero to 3.4. We have chosen a mean of 2 which gives binding constants in the high end of the distribution that are roughly in accord with those that we have observed experimentally for the stronger binders in our DCLs.[2a–g,k]

The concentration of each of the library members was calculated, both in the presence and absence of template. The ratio of these concentrations gave the amplification factor for each host.

Results and Discussion

The results from a typical randomly generated library are shown in Figure 1 a, which compares the free energy of binding[13] of each of the receptors to their amplification factors, using total concentrations of 10 mM for both the building blocks and the template. The correlation between the binding strength and the amplification factor is satisfactory, and the best receptor is amplified to the greatest extent. Figure 1 b shows the results for another randomly generated DCL, that only differs from that in Figure 1 a in the way the binding constants are distributed over the various receptors. Here, the correlation is poor, and the best receptor is hardly amplified at all. The results in Table 1 explain the cause of the remarkable difference in the behaviour of these two libraries. In the library in Figure 1 b, the dimer *CG* has a fair affinity for the template, and a high initial concentration. This combination results in *CG* accounting for 74% of the available *C* and *G* in the presence of the template. The concentration of the best host, *CCCE*, critically depends upon the availability of the building block *C*, and so *CCCE* is only amplified to a small extent despite its high affinity for the template. In the library in Figure 1 a, however, there are no dimers with a stronger free energy of binding than $-18.5\ \mathrm{kJ\,mol^{-1}}$, and so there is no significant "drag" on building blocks that make up the best host, *DDDE*. These examples demonstrate that the disruption of the correlation between amplification and binding is not limited to small libraries as described by Severin,[4] but can also occur in larger libraries.

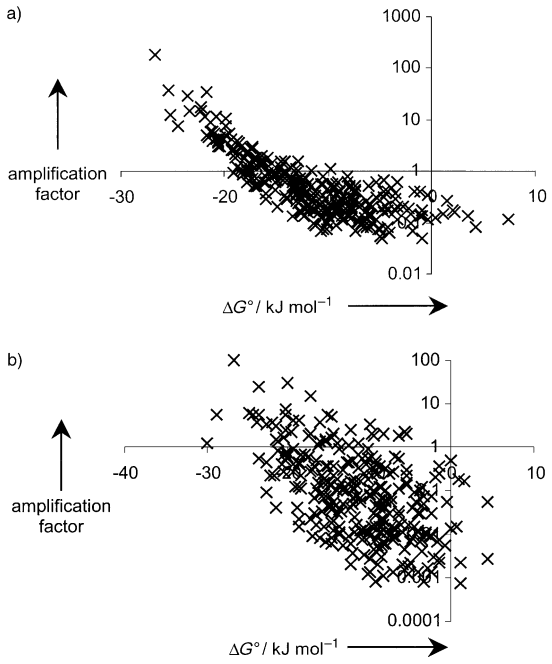

Figure 1. The relationship between amplification and free energy of binding for all of the hosts in two randomly-generated DCLs, that differ only in the way the binding constants are distributed over the various hosts. In both DCLs, the total concentration of the building blocks and the concentration of the template is 10 mM. R^2 values for the correlation between free energy and the logarithm of the amplification factor (taking only significantly amplified hosts into account) are 0.72 and 0.24, respectively.

In an attempt to assess how the correlation between binding affinity and amplification depends on experimental conditions, DCLs were simulated for a range of template and building block concentrations. Seventeen values in the range of 0.1 mM to 1 M were chosen for each concentration, giving a total of 289 different "experimental" conditions. To get statistically significant data, a set of fifty DCLs was randomly generated, and the library composition was simulated for each of the 289 sets of conditions, thus generating a total of 14450 DCL simulations. For each simulation, the correlation between binding energy and the logarithm of the amplification factor was quantified by calculating the correlation coefficient R^2 using all of the hosts that were significantly amplified (at least doubled in concentration).[14] Some typical examples of libraries exhibiting different R^2 values are shown in the Supporting Information.

Table 1. Concentration and amplification data for selected hosts in the DCLs shown in Figure 1.

Hosts from Figure 1 a					Hosts from Figure 1 b				
Ranking[a]	Host	$\Delta G^0_{\mathrm{binding}}$ [kJ mol^{-1}]	c [μM][b]	Amplification factor	Ranking	Host	$\Delta G^0_{\mathrm{binding}}$ [kJ mol^{-1}]	c [μM]	Amplification factor
1	*DDDE*	−26.7	6.22	183.9	1	*CCCE*	−30.0	0.04	1.2
2	*FGG*	−25.4	102.68	38.8	2	*ACE*	−28.7	30.30	5.7
3	*BBBE*	−25.2	0.43	12.8	3	*BDDF*	−26.7	9.74	96.0
4	*AAAF*	−24.5	0.27	7.9	4	*AA*	−24.8	562.57	6.1
–	–	–	–	–	5	*BE*	−24.2	962.37	5.2
101	*AE*	−14.6	223.31	1.2	6	*BBFF*	−23.8	0.20	3.8
114	*AD*	−13.5	218.66	1.2	7	*CG*	−23.6	1050.72	5.7

[a] Ranking indicates order of binding affinity. [b] The concentration in the templated libraries.

The results from these simulations are summarized in Figure 2, which shows how the mean (Figure 2a) and the standard deviation (Figure 2b) in R^2 in the fifty libraries vary across the range of conditions. An alternative metric of the correlation between binding affinity and amplification is the fraction of libraries from which the best receptor is the one that is the most strongly amplified.

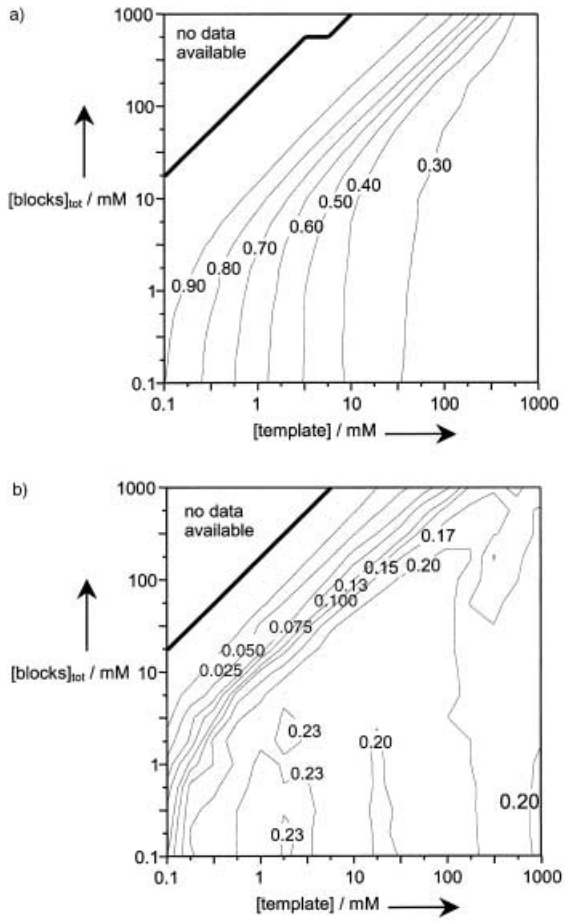

Figure 2. The a) mean and b) standard deviation of the correlation between binding affinity and amplification in DCLs, as a function of template and total building block concentration.

Figure 3a shows the results. The chance of finding the best receptor among the *three* most highly amplified compounds is shown in Figure 3b. At high building block concentrations (>10 mM), the correlation is dependent upon the ratio of template to building blocks. Here, keeping the concentration of template at a tenth of the total concentration of building blocks gives an acceptable R^2 of 0.8 ± 0.1, or a probability of the best host being the one that is most strongly amplified of about 90%. Using a large excess of template gives a much poorer correlation, as this allows for the strong amplification of the dimers with moderate affinity. At lower concentrations (<10 mM), the concentration of the template alone becomes important, and results of a similar quality can be obtained for these libraries by keeping the template concentration at around 0.3 mM—about the affinity of a -20 kJ mol^{-1} ($K=3\times10^4$ M^{-1}) receptor. However, in this region, the correlation coefficient varies more from one library to the next (Figure 2b). It is encouraging that it is rare for a very poor receptor to be significantly amplified, even under the worst conditions.

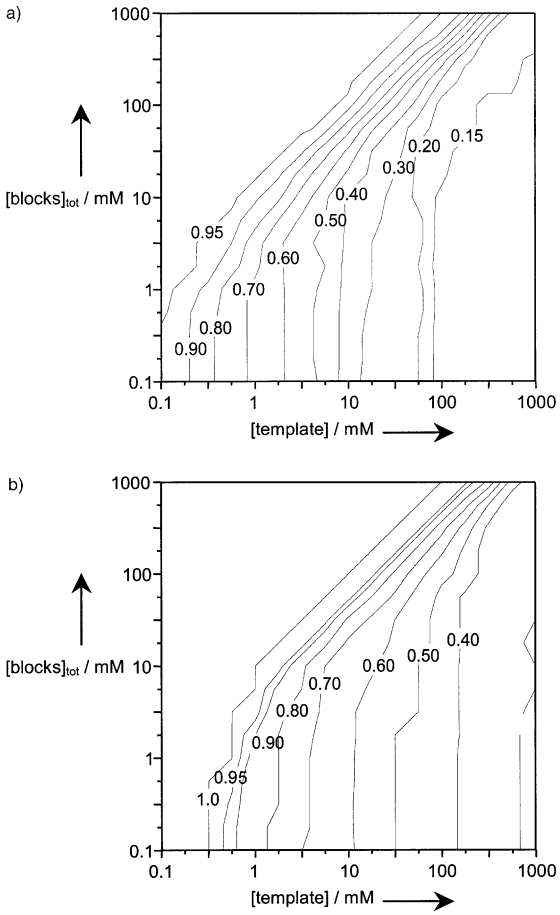

Figure 3. The probability of the most strongly binding host in a DCL being a) the most amplified compound or b) among the three most highly amplified compounds, as a function of template and total building block concentration.

Conclusion

In summary, our simulations have demonstrated that the correlation between host–guest binding and host amplification can vary, depending on how the binding constants happen to be distributed over the various library members. In the libraries we have analyzed, the main disruption of the correlations results from moderately good hosts that are already present in relatively high concentrations in the absence of template and increase in concentration even further upon addition of template to become a "drag" on the building block reservoir. This hampers the amplification of better hosts that require some of the same building blocks.

However, our simulations have identified experimental conditions under which this effect is reduced to an acceptable level. As a rule of thumb, restricting the amount of template to a tenth of the total building block concentration means that the different receptors have to compete for the

template, and the best binder has a high probability of being the most strongly amplified compound. However, restricting the concentration of template tends to reduce amplification factors across the whole library, increasing the chance that good hosts remain undetected because their concentrations are below detection limits. It may therefore be advisable to screen libraries in two rounds, starting with a relatively high (equimolar with respect to the total building block concentration) template concentration. For those libraries where amplification of some hosts is observed, a second round of screening using a reduced amount of template will then be likely to reveal the best receptors. In short, the effects noted by Severin et al. can also be present in large libraries under some experimental conditions, but we have demonstrated here that they need not restrict the usefulness of dynamic combinatorial chemistry in general.

Acknowledgement

We thank Dr. J. E. Redman for alerting us to reference [6]. We acknowledge support from the Royal Society (University Research Fellowship) to S.O. and from EPSRC to J.K.M.S. & P.T.C.

[1] For reviews, see a) S. Otto, *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 509–520; b) S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Curr. Opin. Chem. Biol.* **2002**, *6*, 295–321; c) R. L. E. Furlan, S. Otto, J. K. M. Sanders, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 4801–4804; d) S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem.* **2002**, *114*, 938–993; *Angew. Chem. Int. Ed.* **2002**, *41*, 898–952; e) O. Ramström, T. Bunyapiboonsri, S. Lohmann, J.-M. Lehn, *Biochim. Biophys. Acta* **2002**, *1572*, 178–186; f) S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Drug Discovery Today* **2002**, *7*, 117–125.

[2] For amplification in libraries of macrocycles, see: a) B. Brisig, J. K. M. Sanders, S. Otto, *Angew. Chem.* **2003**, *115*, 1308–1311; *Angew. Chem. Int. Ed.* **2003**, *42*, 1270–1273; b) S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Science* **2002**, *297*, 590–593; c) S. L. Roberts, R. L. E. Furlan, S. Otto, J. K. M. Sanders, *Org. Biomol. Chem.* **2003**, *1*, 1625–1633; d) S. L. Roberts, R. L. E. Furlan, G. R. L. Cousins, J. K. M. Sanders, *Chem. Commun.* **2002**, 938–939; e) R. L. E. Furlan, Y.-F. Ng, G. R. L. Cousins, J. E. Redman, J. K. M. Sanders, *Tetrahedron* **2002**, *58*, 771–778; f) R. L. E. Furlan, G. R. L. Cousins, J. K. M. Sanders, *Chem. Commun.* **2000**, 1761–1762; g) A. L. Kieran, A. D. Bond, A. M. Belenguer, J. K. M. Sanders, *Chem. Commun.* **2003**, 2674–2675; h) E. Stulz, S. M. Scott, A. D. Bond, S. J. Teat, J. K. M. Sanders, *Chem. Eur. J.* **2003**, *9*, 6039–6048; i) B. Fuchs, A. Nelson, A. Star, J. F. Stoddart, S. Vidal, *Angew. Chem.* **2003**, *115*, 4352–4356; *Angew. Chem. Int. Ed.* **2003**, *42*, 4220–4224; j) O. Storm, U. Lüning, *Chem. Eur. J.* **2002**, *8*, 793–798. For other examples of amplification in dynamic combinatorial libraries see: k) S. Otto, S. Kubik, *J. Am. Chem. Soc.* **2003**, *125*, 7804–7805; l) M. Hochgürtel, R. Biesinger, H. Kroth, D. Piecha, M. W. Hofmann, S. Krause, O. Schaaf, C. Nicolau, A. V. Eliseev, *J. Med. Chem.* **2003**, *46*, 356–358; m) M. Hochgürtel, H. Kroth, D. Piecha, M. W. Hofmann, C. Nicolau, S. Krause, O. Schaaf, G. Sonnenmoser, A. V. Eliseev, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3382–3387; n) C. Karan, B. L. Miller, *J. Am. Chem. Soc.* **2001**, *123*, 7455–7456; o) V. Berl, I. Huc, J.-M. Lehn, A. DeCian, J. Fisher, *Eur. J. Org. Chem.* **1999**, 3089–3094; p) H. Hioki, W. C. Still, *J. Org. Chem.* **1998**, *63*, 904–905; q) I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106–2110. For examples of noncovalent DCLs see ref. [1], and references therein.

[3] DCLs may also be constructed where a host molecule is used as a template, and the library members act as guests.

[4] a) Z. Grote, R. Scopelliti, K. Severin, *Angew. Chem.* **2003**, *115*, 3951–3955; *Angew. Chem. Int. Ed.* **2003**, *42*, 3821–3825; b) K. Severin, *Chem. Eur. J.* **2004**, *10*, 2565–2580.

[5] In this case, the differences in intrinsic stability are considered, although the results are equally applicable to stabilisation by a template.

[6] The calculations were performed by the program DCLSim, version 1.0, developed by PTC. Full details of the method are provided in the supporting information. Equilibrium calculations were based on the COGS algorithm: D. D. Perrin, I. G. Sayce, *Talanta* **1967**, *14*, 833–842.

[7] Simulations with 5 or 10 building blocks give similar trends in how the correlation varies with concentration, but with the libraries with fewer building blocks giving somewhat better correlations.

[8] When pentamers are included in the simulation, their overall concentration is so low that they do not have a significant effect on the rest of the DCL.

[9] Isomerism is not considered here—all of the compounds of a particular composition are lumped together.

[10] At 10 mM total building block concentration, the template-free library was partitioned into 90.1% dimers, 9.1% trimers and 0.9% tetramers by amount of building blocks used. This gives a slightly steeper distribution than that experimentally observed in one of our single building block disulfide DCLs[2b].

[11] a) J. S. Moore, N. W. Zimmerman, *Org. Lett.* **2000**, *2*, 915–918; b) P. T. Borbett, S. Otto, J. K. M. Sanders, *Org Lett.* **2004**, *6*, Web release date: 27. April 2004, DOI: 10.1021/ol049398k

[12] Other (arguably more meaningful) distributions are also used, but a lognormal distribution gives as good a fit to observed data as these distributions. S. Rosenwald, R. Kafri, D. Lancet, *J. Theor. Biol.* **2002**, *216*, 327–336.

[13] The binding constants are converted to free energies of binding at 298 K.

[14] Where less than three compounds were amplified by a factor of more than 2, the run was ignored.

Received: March 26, 2004
Published online: May 6, 2004

