J|A|C|S COMMUNICATIONS

Published on Web 04/14/2010

The Impact of the Size of Dynamic Combinatorial Libraries on the Detectability of Molecular Recognition Induced Amplification

R. Frederick Ludlow and Sijbren Otto*

Centre for Systems Chemistry, Stratingh Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received February 25, 2010; E-mail: s.otto@rug.nl

In little over a decade dynamic combinatorial chemistry[1] has developed into a powerful approach for the discovery of new synthetic receptors and ligands for biomolecules. The technique has recently also been used to identify new catalysts,[2] self-replicating molecules,[1k−m,3] and catenanes.[1d,4]

In a dynamic combinatorial library (DCL) building blocks combine through reversible linkages to generate a mixture of products that are in equilibrium through a continuous exchange of the constituent building blocks. Molecular recognition between the molecules in the library or with externally added templates will stabilize selected species, such that the system tends toward optimizing the noncovalent interactions between template and library members. Ideally, this leads to a shift of the equilibrium toward those species that form the most efficient noncovalent interactions; i.e., the best synthetic receptor or the best ligand is amplified selectively. However, there is often a trade-off between achieving large amplification factors[5] and the extent to which amplification is selective for the strongest binders.[6−8]

One of the last remaining unanswered fundamental questions in dynamic combinatorial chemistry relates to the issue of library size. Larger DCLs have a higher probability of containing stronger binders, yet detecting all library members is no longer possible. This raises the question of whether it is better to screen only (many) small libraries, whether larger libraries are more efficient, or whether there is an optimum size? These issues have received very little attention, despite the fact that library size is a key parameter in the experimental design.

Previous theoretical work by Moore[9] and ourselves[10] has indicated that template-induced amplifications remain significant in larger libraries. However, the crude models of the DCLs used in these studies[11] do not allow for probing the behavior of libraries made from several different building blocks. Indeed, no theoretical guidelines currently exist to help the experimentalist to choose how many building blocks to use when preparing a dynamic combinatorial library.

Equally little guidance can be obtained from the experimental work on dynamic combinatorial chemistry as this has mostly been limited to relatively small libraries, often made from only a single building block forming less than 10 library members. However, there are a few reports in which larger libraries have been used successfully. For example, Miller has identified a ligand that binds to an RNA fragment of HIV-1 from a resin-bound DCL of 11 325 members.[12] We recently succeeded in identifying a synthetic receptor for ephedrine from a solution-phase DCL of similar size.[13] These results[14] prove that it is experimentally feasible to screen large libraries. However, they do not reveal whether it is wise to use such large DCLs.

Obtaining a statistically meaningful answer to the question of how large a dynamic combinatorial library should be to have the highest probability of identifying a truly outstanding binder requires an excessively (vide infra) large number of experiments. As there currently exist insufficient data to probe this experimentally we performed an extensive series of computer simulations.

We now report the results of these studies in which we investigated the effect of library size on (i) the probability of detecting *any* amplification; (ii) the probability of detecting the strongest binding library member present; and (iii) the binding affinity of the most amplified *detectable* library member. In parallel we investigated how the above parameters depend on experimental conditions (template and building block concentrations). We show that within the range of libraries sizes investigated (65−4828 compounds), bigger libraries produce better binders. Importantly, the affinities of these compounds are higher than statistically expected on the basis of the fact that more compounds are screened when using larger libraries.

Using our dedicated DCLSim software[7c] we simulated the response of a set of different libraries made from 4, 6, 8, 10, 13, or 16 building blocks to the introduction of a template.[15] Considering oligomeric species up to tetramers this gave DCLs of 65, 203, 486, 990, 2366, and 4828 compounds, respectively. Each library was simulated under a set of 81 different experimental conditions (9 different template and building block concentrations ranging from 0 to 1 M and 100 μM to 1 M, respectively). Following literature precedents[9,10] template binding constants of each individual library member were assigned randomly from a log-normal distribution (mean binding constant 100 M^{-1} and standard deviation 10 M^{-1}).[16] To arrive at a data set from which statistically meaningful conclusions can be derived, 100 libraries with different randomly assigned template binding affinities were simulated for each set of experimental conditions. This amounted to a total of 97 200 in silico DCLs.[17]

While we know the concentrations of all library members for the in silico DCLs, in an experimental library many of these may fall below the limits of detection of ordinary analytical equipment. To arrive at conclusions that can be used to guide experimental design, a detection limit needs to be imposed on the data. Based on our experience with the LC-MS analysis of large libraries, we estimate that it is possible to detect any compound that is amplified at least 2-fold and for which the difference between its concentration in the untemplated relative to the templated library represents at least 1% of the concentration of the most abundant species in the mixture.[18,19] This detection threshold means that as the library gets larger the probability that we are able to detect amplification events diminishes. For the larger libraries and for most experimental conditions (template and total building block concentrations), only a few percent of the library members are in fact detectable, while the vast majority of the library members will go unnoticed (see Figure S3). Also when only considering the best binders in the library the majority of these will go unnoticed (see Figure S4). Thus, the very best binder may remain undetectable while only

10.1021/ja1013689

suboptimal binders are observed.[20] Is this problematic? This depends on the extent to which the library size influences the affinity of the most amplified compounds that remain within detection limits. To quantify this we looked at the average[21] affinity of the best binder from among the three highest detectable amplifications.[22] The results of this analysis are shown in Figure 1 as a function of the experimental conditions. Clearly, with increasing library size the affinity of the detected library members increases as well. Thus, even though detection efficiency falls as library size increases, the probability of discovering a very strong binder rises more rapidly.

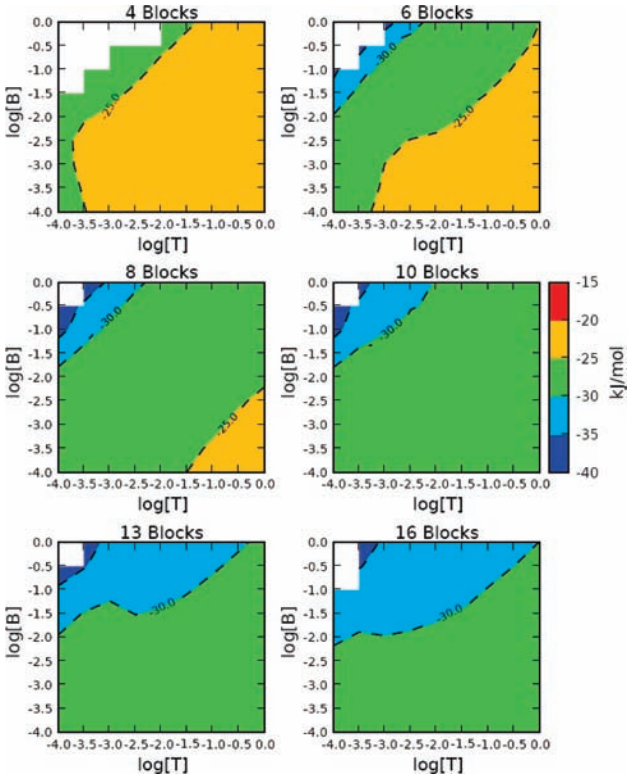

Figure 1. Binding energy of the best binder from among the three most amplified detectable library members for various library sizes (i.e., number of building blocks). Each graph surveys different combinations of template (T) and building block (B) concentrations. The white areas in the upper left corners reflect the scarcity of amplification data in libraries with very low template to building block ratios.

One might expect to obtain stronger binders in the larger libraries simply because the absolute number of compounds that are screened increases (albeit slowly) with library size. We have estimated how much the statistically expected affinity[23] of the best binder from among the detectable library members increases as the library size increases (Table 1). Under the typical experimental conditions shown in the table, this increase amounts to approximately 4−5 kJ/mol upon increasing the library size from 4 to 16 building blocks. The actual affinity that we observe for a given library size is several kJ/mol larger than the statistically expected affinity, and this difference increases as the library size increases. This reflects the fact that stronger binders tend to be more efficiently amplified and therefore have an increased probability of being present at detectable concentrations. Thus, the adaptive nature of dynamic combinatorial libraries facilitates their use under conditions where the majority of library members are not detectable.

We have compared the difference in affinity of the strong binders that we are able to detect with the very best binders in the system that remained below detection limits. The results (Figure S5) show that under typical experimental conditions the best detectable binders are only 2−4 kJ/mol weaker in affinity than the very best binders present. The difference is only weakly dependent on library size, increasing slowly as more building blocks are added. While it may be possible to obtain the very best binder by a parallel screen of a very large number of small sublibraries, it is doubtful whether the substantial additional effort in screening can be justified given the relatively small gain in expected affinity of the best binder.

Table 1. Comparison between the Mean Affinity of the Best Binder Statistically Expected for a Given Number of Detectable Library Members and the Actually Observed Affinity for Libraries with an Overall Building Block Concentration of 10 mM

building blocks	[template] (mM)	detectable library members[a]	expected affinity (kJ/mol)[b]	observed affinity (kJ/mol)[c]
4	1	25 (38%)	−22.6	−24.6
4	10	15 (23%)	−21.3	−23.2
10	1	75 (7.6%)	−25.1	−28.0
10	10	63 (6.4%)	−24.7	−27.3
16	1	146 (3.0%)	−26.5	−30.0
16	10	137 (2.8%)	−26.3	−29.4

[a] Mean number of compounds present at concentrations above the detection limit in the library in the presence of the template. The number in brackets refers to the percentage of the total number of compounds in the particular library. [b] Mean affinity of the best binder after n random draws from a normal distribution (mean $K = 100$ M^{-1}; standard deviation $= 10$ M^{-1}), where $n =$ the number of detectable library members. [c] Mean affinity of the best binder among the three most amplified compounds.

What are the optimal experimental conditions for large dynamic combinatorial libraries?[24] Figure 1 shows that for every library size analyzed the highest affinity library members are obtained in the top left corner of the graphs, i.e., using a high concentration of building blocks and a low concentration of template. However, this combination very often fails to produce any template effects at all. This is not captured in Figure 1, which only shows the results for the rare occasions in which a detectable amplification effect occurs. However, Figure 2 shows how the probability of detecting amplification effects rises as the ratio of template to building block increases (i.e., from the top left to the bottom right of the graphs). It is interesting to note that larger libraries are characterized by a higher probability of detecting amplifications, presumably because of the increased probability that they contain strong binders. Comparing Figures 1 and 2 indicates that, when selecting experimental conditions for the DCLs, a trade-off has to be made between the probability of detecting amplifications and the average affinity of the library members that are amplified. It appears that a total building block to template ratio of 10:1 gives a good initial compromise and would be a good starting point for setting up a DCL experiment. Should the experimental system under these conditions still fail to produce any amplification, then the concentration of template may be increased further.

In conclusion, larger DCLs are likely to produce better binders, at least within the range of library sizes investigated herein (up to 4828 members). With increasing library size the probability of making a strong binder rises more rapidly than detection efficiency drops. The affinity of the best binders detected increases more rapidly than expected statistically on the basis of the increase in the number of compounds screened. Furthermore, larger DCLs have a smaller probability of failing to show any amplification than smaller libraries. When choosing template and building block concentrations a trade-off needs to be made between affinity and probability of detection. Using a 10:1 building block to template ratio seems a good compromise. The implications of this work are clear: it is likely to be advantageous to work with DCLs

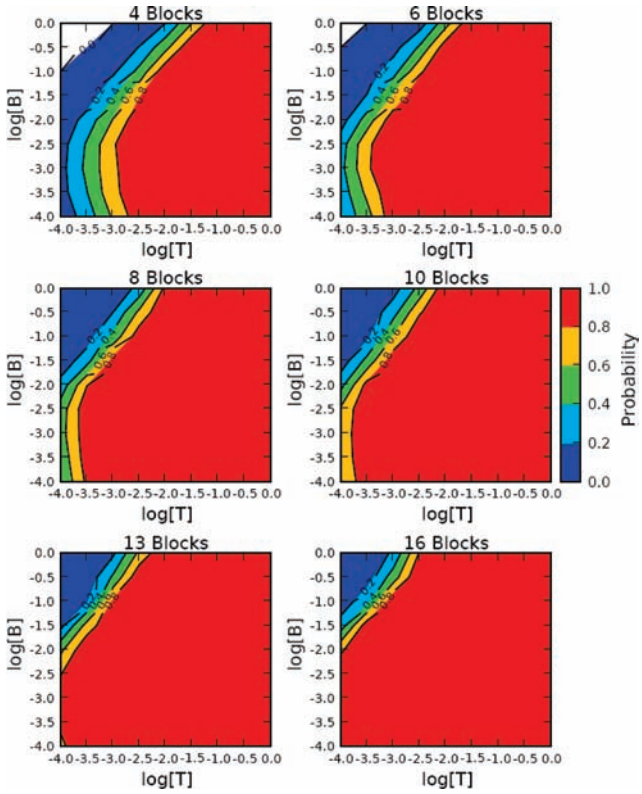

Figure 2. Probability of detecting any amplification for various library sizes (i.e., number of building blocks). Each graph surveys different combinations of template (T) and building block (B) concentrations.

that are much larger than the vast majority reported thus far, provided all building blocks used can potentially contribute to binding affinity.[25]

Acknowledgment. We are grateful for financial support from the EPSRC and the Royal Society.

Supporting Information Available: Details of the computational methods; absolute values of detection thresholds, a comparison between results obtained when applying a detection threshold of 10% instead of 1%, fraction of the detectable library members as a function of library size and experimental conditions; analysis of the difference in affinity of the best binding detectable library members versus the overall best binder present, and experimental recommendations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) For recent reviews, see: (a) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219–226. (b) Lehn, J. M. *Chem. Soc. Rev.* **2007**, *36*, 151–160. (c) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652–3711. For recent publications, see: (d) Au-Yeung, H. Y.; Pantos, G. D.; Sanders, J. K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10466–10470. (e) Au-Yueng, H. Y.; Pengo, P.; Pantos, G. D.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2009**, 419–421. (f) Barboiu, M.; Dumitru, F.; Legrand, Y. M.; Petit, E.; van der Lee, A. *Chem. Commun.* **2009**, 2192–2194. (g) Cacciapaglia, R.; Di Stefano, S.; Ercolani, G.; Mandolini, L. *Macromolecules* **2009**, *42*, 4077–4083. (h) Coskun, N.; Aksoy, C. *Tetrahedron Lett.* **2009**, *50*, 3008–3012. (i) Gasparini, G.; Bettin, F.; Scrimin, P.; Prins, L. J. *Angew. Chem., Int. Ed.* **2009**, *48*, 4546–4550. (j) Inzgerman, L. A.; Waters, M. L. *J. Org. Chem.* **2009**, *74*, 111–117. (k) Nguyen, R.; Allouche, L.; Buhler, E.; Giuseppone, N. *Angew. Chem., Int. Ed.* **2009**, *48*, 1093–1096. (l) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nanotechnol.* **2009**, *4*, 19–24. (m) Otto, S. *Nat. Nanotechnol.* **2009**, *4*, 13–14. (n) Perez-Fernandez, R.; Pittelkow, M.; Belenguer, A. M.; Lane, L. A.; Robinson, C. V.; Sanders, J. K. M. *Chem. Commun.* **2009**, 3708–3710. (o) Saggiomo, V.; Luning, U. *Chem. Commun.* **2009**, 3711–3713. (p) Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science* **2009**, *325*, 73–77.

(2) (a) Gasparini, G.; Prins, L. J.; Scrimin, P. *Angew. Chem., Int. Ed.* **2008**, *47*, 2475–2479. (b) Vial, L.; Sanders, J. K. M.; Otto, S. *New J. Chem.* **2005**, *29*, 1001–1003. (c) Brisig, B.; Sanders, J. K. M.; Otto, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 1270–1273.

(3) (a) Xu, S.; Giuseppone, N. *J. Am. Chem. Soc.* **2008**, *130*, 1826–1827. (b) Sadownik, J. W.; Philp, D. *Angew. Chem., Int. Ed.* **2008**, *47*, 9965–9970. (c) Carnall, J. M. A.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C. A.; Peyralans, J. J.-P.; Otto, S. *Science* **2010**, *327*, 1502–1506.

(4) (a) West, K. R.; Ludlow, R. F.; Corbett, P. T.; Besenius, P.; Mansfeld, F. M.; Cormack, P. A. G.; Sherrington, D. C.; Goodman, J. M.; Stuart, M. C. A.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 10834–10835. (b) Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667–669.

(5) Amplification factors are defined as the concentration of a given library member in the presence of the template relative to its concentration in the absence of template.

(6) (a) Saur, I.; Severin, K. *Chem. Commun.* **2005**, 1471–1473. (b) Severin, K. *Chem.—Eur. J.* **2004**, *10*, 2565–2580.

(7) (a) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem.—Eur. J.* **2008**, *14*, 2153–2166. (b) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 9390–9392. (c) Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Chem.—Eur. J.* **2004**, *10*, 3139–3143.

(8) Using a large excess of template gives the highest amplification factors, while using close-to-stoichiometric amounts of template gives the best selectivities.

(9) Moore, J. S.; Zimmerman, N. W. *Org. Lett.* **2000**, *2*, 915–918.

(10) Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Org. Lett.* **2004**, *6*, 1825–1827.

(11) These models assume that any library member can directly interconvert into any other library member, which is only true for libraries made from a single building block.

(12) McNaughton, B. R.; Gareiss, P. C.; Miller, B. L. *J. Am. Chem. Soc.* **2007**, *129*, 11306–11307.

(13) Ludlow, R. F.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 12218–12219.

(14) Two more reports of large DCLs exist: (a) Otto, S.; Kubik, S. *J. Am. Chem. Soc.* **2003**, *125*, 7804–7805. (b) Hochgurtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3382−3387. In (a) the screening method relies on the chromophore of only one of the building blocks, so that all library members not containing this building block remain invisible. In (b) a difunctional scaffold is used, but the binding pocket of the protein target can only accommodate singly functionalized scaffold molecules. Moreover, the imine equilibrium lies on the side of the starting materials, so that the concentration of the difunctionalized library members remains very low.

(15) In these simulations we do not consider any chemical structure, only template binding affinities and equilibrium constants describing the formation of library members from the building blocks.

(16) Normal distributions appear to be a good approximation of affinity distributions observed for receptors within a given class. See: (a) Conners, K. A. *Chem. Rev.* **1997**, *97*, 1325–1357. (b) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Y. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872–4897.

(17) Note that the size of the data set needed for obtaining statistically significant guidance on the issue of library size presents a huge challenge to any experimental approach of this subject.

(18) This implies a lower absolute detection threshold for larger libraries. However, within identical experimental conditions (same template and total building block concentration) the absolute threshold for a 16 building block library differs from that of a 4 building block library only by a factor of 7 (Figure S1). Hence, injecting a 7-fold larger volume of sample for the 16 building block library would give effectively the same absolute detection threshold.

(19) Repeating the analysis with a detection threshold of amplification effects of 10% of the concentration of the most abundant library member gave comparable results. See Figure S2.

(20) While strong binders will generally have higher amplification factors[5] these may still not be sufficient to make these compounds detectable in cases where the concentration of the particular library member in the absence of the template was very low (as is, for example, the case for a higher oligomer containing several copies of the same building block).

(21) The average was taken of 100 separate simulations that differed only in the randomly assigned template binding affinities.

(22) Given the less than perfect correlation between amplification factors and binding affinities in most DCLs[6,7] it is advisable to follow up on several of the most amplified compounds. Here we arbitrarily decided to consider the three most amplified compounds.

(23) The statistically expected affinity refers to the highest affinity based on n random draws from a normal distribution (mean log $K = 2$; standard deviation $= 1$), where $n =$ the number of detectable library members. See: Teichroew, D. *Ann. Math. Stat.* **1956**, *27*, 410–426.

(24) More detailed experimental recommendations are provided in the Supporting Information.

(25) Our simulations are based on the assumption that binding constants of the library members obey a log normal distribution. This assumption may not hold when combining structurally very different building blocks.

JA1013689