Combinatorial Chemistry

DOI: 10.1002/anie.200700292

Host–Guest Binding Constants Can Be Estimated Directly from the Product Distributions of Dynamic Combinatorial Libraries**

*R. Frederick Ludlow, Jingyuan Liu, Hongxia Li, Sarah L. Roberts, Jeremy K. M. Sanders, and Sijbren Otto**

In the last decade, dynamic combinatorial chemistry has developed into a powerful method for the discovery of new sensor systems, catalysts, ligands for biomolecules, and synthetic receptors.[1] Until now, determining the binding strength for individual library members has required their separation from the library and subsequent analysis. We now show that it is possible to obtain most host–guest binding affinities, even in 30+ component libraries, directly from product distributions by using a numerical data-fitting procedure. We have verified our method by using a three-component experimental system and showing that the binding constants computed from the behavior of the dynamic mixture agree well with those obtained independently by binding studies on the isolated library members.

Dynamic combinatorial libraries (DCLs) of interconverting oligomers are made by mixing a set of building blocks that can reversibly combine with each other to give a dynamic equilibrium. The composition of the DCL is determined by the relative stability of all its members, so it is responsive to external stimuli that can affect these stabilities. Thus, addition of a guest molecule to a library of potential receptors should result in the stabilization and amplification of those library members that can form favorable interactions with the guest molecule. Ideally, the good receptors are amplified at the expense of the other oligomers. Quantifying their binding strength thus far entailed painstaking isolation of the selected compounds from the library. This process often requires switching off the exchange reaction through a substantial change in the pH value, or even covalent modification so that the protonation state or structure of the isolated species may no longer correspond to that of the species that was selected in the library. Also, under certain experimental conditions, the intuitive correlation between amplification and binding affinity can break down.[2] Thus, there is a strong incentive to develop a method that allows the determination of binding affinities directly from the behavior of DCLs without requiring studies on isolated library members. We reasoned that the response of a DCL of potential receptors to changes in the concentrations of building blocks and guest molecules should contain enough information about the underlying host–guest binding constants that it should be possible to extract these by using a multivariable fitting procedure.

We first tested this hypothesis on a simple set of seven libraries by using building block **1**[3] (2 mM; Figure 1) in the presence of varying amounts of guest **2** (0, 0.5, 1.0, 2.0, 4.0, 6.0, and 10 mM). The resulting DCLs contain cyclic dimer, trimer, and tetramer molecules, and the addition of acetylcholine to these libraries results in the amplification of the trimer, $\mathbf{1}_3$. We have used HPLC to determine the equilibrium concentrations of all macrocycles present in each library (Figure 2).[4] We then fitted a model of the DCL to this data set by using our program, DCLFit, based on the DCLSim software we developed previously.[2d,5] The equilibrium between the three macrocyclic products in the absence of any guest was treated in terms of (fictitious[2d]) formation constants K_f,

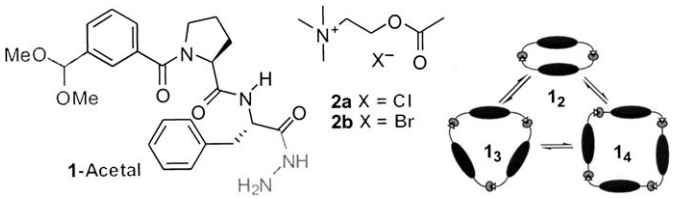

Figure 1. Pseudopeptide building block **1** (protected as dimethyl acetal), acetylcholine guests **2**, and the main products in the small DCL formed from **1**.

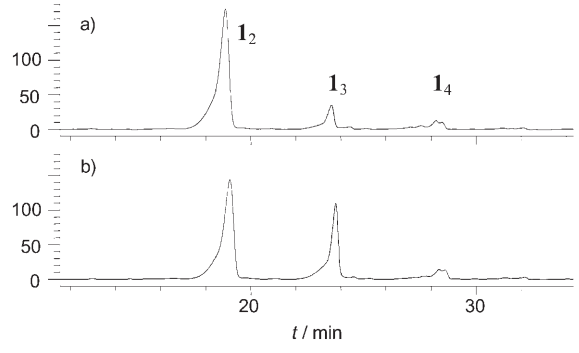

Figure 2. HPLC analyses of libraries in $CHCl_3$/MeOH (98:2) with 100 mM trifluoroacetic acid and 2 mM building block **1** containing a) no template and b) 1 mM acetylcholine chloride (**2a**).

[*] R. F. Ludlow,[#] J. Liu,[#] Dr. H. Li, Dr. S. L. Roberts, Prof. J. K. M. Sanders, Dr. S. Otto
Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge, CB21EW (UK)
Fax: (+44) 1223-336-017
E-mail: so230@cam.ac.uk
Homepage: http://www.ch.cam.ac.uk/staff/so.html

[#] These authors contributed equally.

[**] We are grateful to Peter Corbett for advice on software development and to the EPSRC, BBSRC, and the Royal Society for financial support.

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

which describe the equilibrium between the monomer and the macrocycle (Figure 3).

Values for K_f were determined directly from the experimentally observed product distributions in the absence of guest (see the Supporting Information). We then simulated the composition of the libraries in the presence of the guest by using DCLFit. We initially used an estimated trial set of $K_{n\mathrm{mer}}$ guest values for the interaction of each oligomer with the guest. DCLFit compares the resulting simulated library composition to the experimental observations and calculates the error in the fit (see the Supporting Information for details). By varying the $K_{n\mathrm{mer}}$ guest values so as to minimize the error, the best fitting parameters can be obtained. Numerous algorithms are available that can perform this multiparameter optimization. For our system, we found that the Nelder–Mead,[6] BFGS, and Powell[7] algorithms all gave satisfactory results.

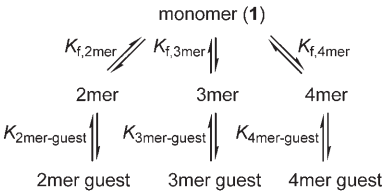

Figure 3. Model used to simulate the DCL shown in Figure 1.

For the libraries templated by acetylcholine chloride (**2a**), all three algorithms gave a fitted trimer binding energy of $-22.1\ \mathrm{kJ\,mol^{-1}}$, whereas dimer and tetramer binding energies were -14.1 and $-20.3\ \mathrm{kJ\,mol^{-1}}$, respectively. The trimer binding energy is in agreement with the value determined experimentally on isolated material by using isothermal titration calorimetry ($-22.2\ \mathrm{kJ\,mol^{-1}}$). Unfortunately, the experimental values for the dimer and tetramer were not accessible owing to the poor solubility of the dimer and the poor yield for the tetramer.

We also prepared a series of libraries by using acetylcholine bromide (**2b**) as a template. In this case, the fitted trimer binding energy ($-17.7\ \mathrm{kJ\,mol^{-1}}$) is also in agreement with the experimental value ($-18.9\ \mathrm{kJ\,mol^{-1}}$). The dependence of the affinity on the anion suggests that this may well be the first ditopic receptor to be discovered by using dynamic combinatorial chemistry; details will be reported elsewhere.

Encouraged by these examples, we set out to apply this method to more complex DCLs. We reasoned that libraries formed from three building blocks containing all combinations from dimers to tetramers represent a realistic level of complexity, capturing the majority of dynamic combinatorial receptors reported thus far.[1] However, experimentally determining the association constants for the resulting 31 oligomers is not feasible, so we decided to use simulated library distributions as the "experimental" data for the fitting process.

First we assigned a random binding energy to each of the 31 oligomers[8] and then simulated libraries with varying building block and template concentrations by using DCLSim. The calculated concentrations from these simulations formed our "experimental observations" and the original binding energies were not used in further calculations. To make this computer-generated data set resemble real experimental data, all concentrations below a designated detection limit were discarded and the remaining data were randomly modified to simulate an experimental error of 10%.[9] Finally, we applied our fitting procedure to the thus-obtained "observed concentrations" to find values for the binding energies that gave the best fit. We imposed one constraint: any oligomer for which there was less than two observations was removed from our model of the library. The challenge in such multiparameter fitting lies in identifying the global minimum, which can involve extensive computation. However, we found that for the present system, the relatively undemanding Powell algorithm gave satisfactory results when starting the iterations from a series of grossly overestimated values for the binding affinities. Repeating the minimization three times for three different initial binding affinities and taking the best fit between experimental and calculated values resulted in generally good agreements between the fitted and real binding energies (see the Supporting Information for further details). Figure 4 shows a typical example that compares the original binding energies (light gray) and the values obtained from fitting the composition of the library by using a set of 12 different experimental conditions (dark grey).[10] There is good agreement between both sets at the more interesting, higher-affinity end of the series. The poorer fitting of the weaker binding oligomers reflects the fact that their concentrations will often be close to the imposed detection limit, reducing the quality of the data for these species. Moreover the total library distribution is inevitably less affected by changes in the fitted values of the weaker binders.

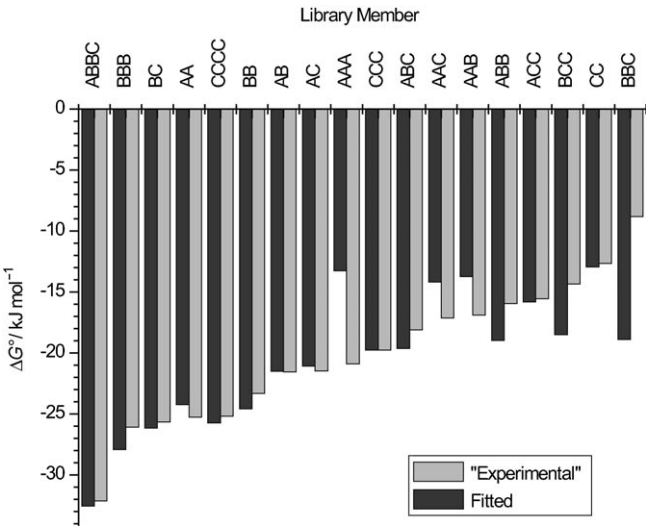

Figure 4. Comparison of the "experimental" and fitted values for the host–guest binding energies in a simulated 31-component DCL. $\Delta G°$ = host–guest binding energy.

In summary, we have developed an efficient method for the simultaneous determination of the host–guest binding

energies from the concentrations of the constituents of dynamic combinatorial libraries. We have demonstrated its effectiveness by applying it to real and simulated libraries and shown that it continues to perform well in complex libraries with many missing data points and a realistic degree of experimental error. These results show that the study of the behavior of a complex chemical system can efficiently produce a wealth of data on its components that would be much more laborious to obtain by studying all the components in isolation. Apart from providing an important new tool for dynamic combinatorial chemistry, this work represents one of the first entries into "systems chemistry", which we define, in analogy to systems biology, as the study of complex chemical networks.

Experimental Section

Dynamic combinatorial libraries were prepared by using a 100 mM solution of trifluoroacetic acid in a $CHCl_3$/MeOH (98:2) solvent mixture. The building block (2 mM) was added as the protected acetal (**1**-acetal) and deprotected in situ. For each template, **2a** and **2b**, seven libraries were prepared, with template concentrations of 0.0, 0.5, 1.0, 2.0, 4.0, 6.0, and 10 mM. Libraries were allowed to equilibrate for 7 days before analysis by HPLC and LC–MS.

HPLC and LC–MS: Analytical HPLC was carried out on an Agilent 1100 instrument coupled to a UV analyzer, set to 290 nm with 550 nm as reference. The data was processed by using HP Chemstation software. Separations were performed on a Waters Symmetry C18 column (250 × 4.6 mm, 5-μm particle size) by using a water (MilliQ) and acetonitrile (Romil) gradient at a temperature of 45 °C ($t = 0$ min: 70% water; $t = 40$ min: 20% water; $t = 45$ min: 20% water). Aliquots of 5 μL of the DCL solution were analyzed by using a flow rate of 1 mL min^{-1}.

HPLC peaks were assigned by LC–MS with an Agilent LC-MSD-Trap-XCT system. The LC column and method were identical to that described above. Mass spectra (positive mode) were acquired in ultrascan mode by using a drying temperature of 350 °C, a nebulizer pressure of 60.00 psi, a drying gas flow of 11.00 L min^{-1}, a capillary voltage of 4000 V, and a capillary current of 39 nA.

Isolation of $\mathbf{1}_3$ was performed by preparative HPLC by using a Nucleodur C18 column (250 mm × 21 mm, 100 Å, 5 μm) with a Nucleodur C18 guard column (50 mm × 21 mm, 100 Å, 5 μm). Aliquots of 2.5 mL of library solution were injected at a flow rate of 20.00 mL min^{-1}. The solvent of the collected fractions was removed in vacuo, and the residue was dried under vacuum for at least 3 h. ^{1}H NMR (500 MHz (cryoprobe), ($CDCl_3$/MeOD = 98:2): $\delta = 10.36$ (3H, br), 8.23 (3H, s), 7.94 (3H, s), 7.64 (3H, d, $J = 7.3$ Hz), 7.36 (3H, t, $J = 7.3$ Hz), 7.27–7.17 (15H, br m), 7.00 (3H, d, $J = 7.3$), 5.18 (3H, dd, $^1J = 12.0$ Hz, $^2J = 4.9$ Hz), 4.25 (3H, dd, $^1J = 11.0$ Hz, $^2J = 7.3$ Hz), 3.93 (3H, m), 3.67 (3H, dd, $^1J = 14.0$ Hz, $^2J = 4.9$ Hz), 3.49 (3H, m), 2.93 (3H, dd, $^1J = 14.0$ Hz, J2 = 12.0 Hz), 2.09 (3H, br m), 1.89 (3H, br m), 1.77 (3H, br m), 1.38 ppm (3H, br m); ^{13}C NMR (125 MHz (cryoprobe), $CDCl_3$/MeOD = 98:2): $\delta = 171.6$, 171.3, 169.0, 148.0, 137.5, 135.6, 133.4, 132.1, 130.1, 129.2, 128.8, 128.2, 126.5, 123.0, 63.8, 51.4, 50.2, 36.1, 29.5, 25.7 ppm; MS $[M+H^+]$ $C_{66}H_{67}N_{12}O_9$ requires 1172, found 1171.7; mp: decomposes at 265 °C.

Isothermal titration calorimetry was performed at 298 K by using a MCS isothermal titration calorimeter (MicroCal, Northampton, MA, USA) and $CHCl_3$/MeOH (98:2) as the solvent. The host concentration in the cell was 0.2 mM and the guest (3 mM) was titrated in 10-μL steps by using 30 injections spaced at intervals of 180 s. The binding constants were obtained by using the one-site binding model provided with the ORIGIN software (Version 2.9).

Received: January 22, 2007
Published online: June 28, 2007

Keywords: combinatorial chemistry · computer chemistry · host–guest systems · supramolecular chemistry

[1] 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.
[2] 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; c) I. Saur, K. Severin, *Chem. Commun.* **2005**, 1471–1473; d) P. T. Corbett, S. Otto, J. K. M. Sanders, *Chem. Eur. J.* **2004**, *10*, 3139–3143; e) P. T. Corbett, J. K. M. Sanders, S. Otto, *J. Am. Chem. Soc.* **2005**, *127*, 9390–9392.
[3] S. L. Roberts, R. L. E. Furlan, S. Otto, J. K. M. Sanders, *Org. Biomol. Chem.* **2003**, *1*, 1625–1633.
[4] We assumed that extinction coefficients were equal per monomer and that all material was present either as dimer, trimer, or tetramer. The concentration of each oligomer was deduced from its HPLC integration as a fraction of the total.
[5] DCLSim simulates the distribution of oligomers in the absence or presence of a template. The concentration of all oligomers in the untemplated library is simulated on the basis of the individual formation constants K_f. The product distribution of the templated library is obtained by using the template-binding constants for each oligomer as an input.
[6] J. A. Nelder, R. Mead, *Comput. J.* **1965**, *7*, 308–313.
[7] We used the implementation of these algorithms from the scientific python package: E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, **2001**-, http://www.scipy.org/.
[8] The values were drawn from a normal distribution of log(K) with a mean value of 3 and a standard deviation of 1. Repeating the simulations with a mean value of 4 and standard deviation of 1.3 gave essentially identical results (see the Supporting Information).
[9] For any individual library, the lowest detectable concentration was set as 0.1% of the concentration of the most abundant library member. Any values below this were treated as unknown. To each observed concentration, a random number was added, which was drawn from a normal distribution (mean = 0; standard deviation = 10% of the observed concentration). Repeating the simulations with an increased 20% experimental error gave essentially identical results (see the Supporting Information).
[10] Libraries were simulated by using a constant total building-block concentration of 3 mM. The building-block ratio was varied (1:1:1, 1:1:3, 1:3:1, and 3:1:1) as was the template concentration (0, 1.0, and 10 mM for each building-block ratio), giving 12 unique experimental conditions.

