ORGANIC LETTERS

2009
Vol. 11, No. 22
5110−5113

Estimating Equilibrium Constants for Aggregation from the Product Distribution of a Dynamic Combinatorial Library

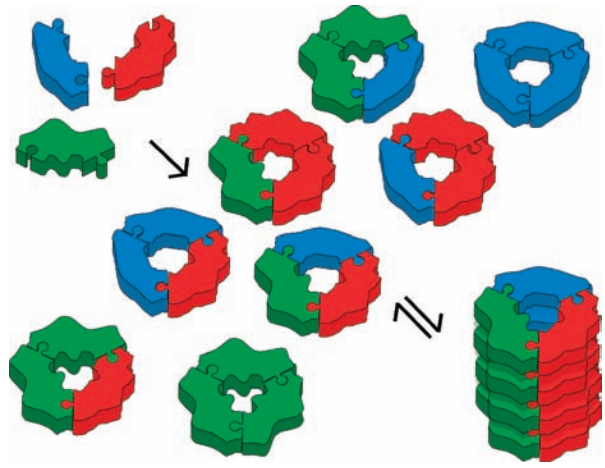
Rosemary A. R. Hunt,[†] R. Frederick Ludlow,[†] and Sijbren Otto*

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

s.otto@rug.nl

Received July 21, 2009

ABSTRACT



Multicomponent chemical systems that exhibit a network of covalent and intermolecular interactions may produce interesting and often unexpected chemical or physical behavior. The formation of aggregates is a well-recognized example and presents a particular analytical challenge. We now report the development of a numerical fitting method capable of estimating equilibrium constants for the formation of aggregates from the product distribution of a dynamic combinatorial library containing self-recognizing library members.

Dynamic combinatorial chemistry[1] has been employed in recent years mainly in the development of synthetic receptors,[2] ligands for biomolecules,[3] and catalysts,[4] where specific ligand−receptor interactions stabilize specific library members. These studies have exploited thermodynamically controlled molecular networks primarily as a means of screening for individual species with interesting properties which are then isolated. Recently a number of studies have appeared which explore the application of dynamic combinatorial libraries (DCLs) for the development of self-synthesizing

* To whom correspondence should be addressed. Fax: (+31) 50-3634296. (+31) 503638639.

[†] Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

(1) For reviews see: (a) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219. (b) Lehn, J.-M. *Chem. Soc. Rev.* **2007**, *36*, 151. (c) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *35*, 3652. (d) de Bruin, B.; Hauwert, P.; Reek, J. N. H. *Angew. Chem., Int. Ed.* **2006**, *45*, 2660.

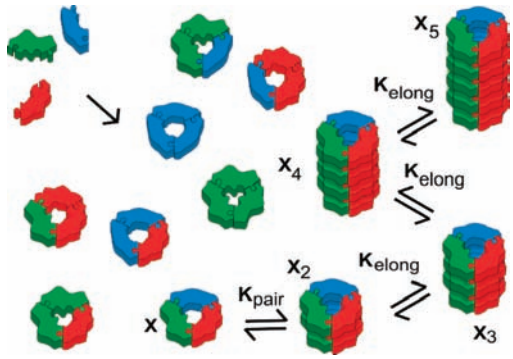
(2) (a) Au-Yeung, H. Y.; Pengo, P.; Pantos, G. D.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2009**, 419. (b) Ludlow, R. F.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 12218. (c) Chung, M.-K.; Hebling, C. M.; Jorgenson, J. W.; Severin, K.; Lee, S. J.; Gagné, M. R. *J. Am. Chem. Soc.* **2008**, *130*, 11819. (d) Besenius, P.; Cormack, P. A. G.; Liu, J. Y.; Otto, S.; Sanders, J. K. M.; Sherrington, D. C. *Chem.—Eur. J.* **2008**, *14*, 9006. (e) 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. (f) Besenius, P.; Cormack, P. A. G.; Ludlow, R. F.; Otto, S.; Sherrington, D. C. *Chem. Commun.* **2008**, 2809. (g) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem.—Eur. J.* **2008**, *14*, 2153.

10.1021/ol901656x CCC: $40.75
Published on Web 10/13/2009

molecules and materials.[5] The central premise is that any library member that is capable of recognizing itself will induce its own amplification. Such systems are of considerable interest, particularly when self-recognition gives rise to larger aggregates. The resulting materials can be considered to be self-synthesizing noncovalent polymers[6] and are potentially self-replicating.[7]

We now report a numerical data fitting method that allows extraction of the equilibrium constants that define the aggregation process of self-complementary library members[8] from product distributions of DCLs. The input consists of the concentrations of the most abundant library members at various building block concentrations. Such data are routinely obtained by HPLC analysis of DCLs based on reversible covalent chemistries, including that using relatively labile imine bonds.[9] We also provide guidance on the experimental conditions best suited to obtaining the most accurate estimates. This work takes a systems chemistry[10] approach and builds on the notion that it is possible to extract useful information from the global analysis of thermodynamically controlled networks (i.e., DCLs). This enables the properties of the constituent library members to be assessed without the need for their time-consuming isolation and independent study. The unique feature of our dynamic network approach is that characterizing aggregation of a species does not require data on the proportion of this species that is in the aggregated (or nonaggregated) form—we only need to know its total concentration and that of the other covalent molecules in the network with which it is in equilibrium.

Scheme 1. A DCL in Which Library Member X Aggregates



We have previously published details of a numerical fitting program (DCLFit[11]) capable of generating estimates of equilibrium constants for host–guest association from concentrations of library members in DCLs in the presence of different concentrations of added guests. DCLFit is ill equipped to cope with extended complexes (i.e., aggregates). The problem is that all aggregates built up from different numbers of library members would have to be populated separately, which is unpractical when aggregates are polydisperse. Thus, a simplified aggregation model needs to be introduced as shown in Scheme 1. This model is the most widely used of the indefinite self-association models[12] and describes aggregation starting with a dimerization event (quantified by K_{pair}) followed by a series of stepwise extensions (each quantified by K_{elong}).[13] In many systems K_{pair} will be smaller than K_{elong} as a result of bond polarization upon association (typically observed in extended hydrogen bonding chains formed by amides and ureas[14]) and/or loss of entropy due to reduced conformational freedom in the aggregate (this happens for both interacting species in the pairing step but only for the incoming species upon subsequent elongation). Conversely, steric interactions might mean that the formation of dimer occurs readily but the subsequent formation of the trimer is extremely hindered.[15]

The experimental observables in a DCL from which K_{pair} and K_{elong} may be derived are usually the total concentration of each of the library members, which can be determined by using, for example, HPLC. For aggregating library member X (Scheme 1) the total concentration of X is given by eq 1:

$$[X]_{total} = [X] + 2[X_2] + 3[X_3] + ... \quad (1)$$

where [X] is the concentration of X that is not aggregated and $[X_n]$ is the concentration of the *n*-meric aggregate.

Given that $[X_2] = K_{pair}[X]^2$ and $[X_3] = K_{elong}[X][X_2]$ eq 1 can be rewritten as:

(3) (a) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. *J. Am. Chem. Soc.* **2008**, *130*, 16254. (b) Nielsen, M. C.; Ulven, T. *Chem.—Eur. J.* **2008**, *14*, 9487. (c) Cancilla, M. T.; He, M. M.; Viswanathan, N.; Simmons, R. L.; Taylor, M.; Fung, A. D.; Cao, K.; Erlanson, D. A. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3978. (d) Bugaut, A.; Jantos, K.; Wietor, J.-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 2677.

(4) (a) Gasparini, G.; Prins, L. J.; Scrimin, P. *Angew. Chem., Int. Ed.* **2008**, *47*, 2475. (b) Brisig, B.; Sanders, J. K. M.; Otto, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 1270.

(5) (a) Nguyen, R.; Allouche, L.; Buhler, E.; Guiseppone, N. *Angew. Chem., Int. Ed.* **2009**, *48*, 1093. (b) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nanotechnol.* **2008**, *4*, 19. (c) Otto, S. *Nat. Nanotechnol.* **2008**, *4*, 13. (d) Sadownik, J. W.; Philp, D. *Angew. Chem., Int. Ed.* **2008**, *47*, 9965. (e) Xu, S.; Giuseppone, N. *J. Am. Chem. Soc.* **2008**, *130*, 1826.

(6) (a) Abbel, R.; Grenier, C.; Pouderoijen, M. J.; Stouwdam, J. W.; Leclere, P. E. L. G.; Sijbesma, R. P.; Meijer, E. W.; Schenning, A. P. H. J. *J. Am. Chem. Soc.* **2009**, *131*, 833. (b) Bellot, M.; Bouteiller, L. *Langmuir* **2008**, *25*, 14176. (c) Bouteiller, L. In *Hydrogen Bonded Polymers*; Springer-Verlag Berlin: Berlin, Germany, 2007; Vol. 207, p 79112.

(7) (a) Vidonne, A.; Philp, D. *Eur. J. Org. Chem.* **2009**, 593. (b) Patzke, V.; von Kiedrowski, G *Arkivoc* **2007**, 293.

(8) We do not consider here aggregates composed of two different building blocks such as would be formed if the library members were charge or hydrogen bond complementary. Analysis of such systems would require the incorporation of several additional fitting parameters.

(9) Zameo, S.; Vauzeilles, B.; Beau, J. M. *Eur. J. Org. Chem.* **2006**, 5441–5444.

(10) For a reviews, see: (a) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *2*, 377. For recent examples, see: (b) Wagner, N.; Ashkenasy, G. *Chem.—Eur. J.* **2009**, *15*, 1765. (c) Sadownik, J. W.; Philp, D. *Angew. Chem., Int. Ed.* **2008**, *47*, 9965.

(11) Ludlow, R. F.; Liu, J.; Li, H. X.; Roberts, S. L.; Sanders, J. K. M.; Otto, S. *Angew. Chem., Int. Ed.* **2007**, *46*, 5762.

(12) For a review of this and other self-association models see: Martin, R. B. *Chem. Rev.* **1996**, *96*, 3043.

(13) While in the present paper only this aggregation model is implemented, the model building functionality of the DLCFit software developed for this application allows the user to specify his/her own aggregation model.

(14) Simic, V.; Bouteiller, L.; Jalabert, M. *J. Am. Chem. Soc.* **2003**, *12*, 13148.

(15) Horne, W. S.; Stout, C. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2003**, *125*, 9372.

$$[X]_{total} = [X] + 2\rho K_{elong}[X]^2 + 3\rho K_{elong}^2[X_3]^3 + ... \quad (2)$$

where ρ is a measure of the cooperativity favoring the second and subsequent binding events over the first binding event and is given by:

$$\rho = \frac{K_{pair}}{K_{elong}} \quad (3)$$

Rearranging eq 2 results in eq 4, which, using the series expansion $1 + 2x + 3x^2 + 4x^3$, etc. equals $1/(1 - x)^2$, can be simplified to eq 5.

$$[X]_{total} = [X](1 - \rho) + \rho[X](1 + 2K_{elong}[X] + 3K_{elong}^2[X]^2 + ...) \quad (4)$$

$$[X]_{total} = [X](1 - \rho) + \frac{\rho[X]}{(1 - K_{elong}[X])^2} \quad (5)$$

This expression was then incorporated into DCLFit.

DCLFit works using the following method: It starts from a set of initial guessed values for the equilibrium constants describing the covalent formation of the library members from the building blocks (the formation constants, K_f) and values for the equilibrium constants describing the noncovalent interactions between the library members (K_{pair} and K_{elong}). It simulates the library distribution that would arise if the guesses were indeed the true values. It then calculates the errors between the observed and modeled concentrations of all the species and uses a multiparameter optimization algorithm, in this case the BFGS algorithm,[16] to minimize the errors and yield a set of fitted equilibrium constants.

To test the ability of the extended version of DCLFit to determine K_{pair} and K_{elong} from library distributions we generated sets of simulated "experimental" libraries. We decided to use simulated, rather than true experimental data to validate our fitting procedure, as in this way we can challenge our method with complex DCLs, for which determining all parameters individually would be excessively time-consuming. Thus in-silico libraries were set up containing up to three building blocks A−C which were allowed to form all possible library members up to tetramers. In a first run we arbitrarily allowed library member ABC to form extended aggregates with arbitrarily chosen values for $\log(K_{pair})$ and $\log(K_{elong})$ of 2.50 and 4.00, respectively.

We then simulated library compositions for a set of 12 different building block concentrations (see the Supporting Information). The thus obtained concentrations of the library members 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 were randomly modified to simulate an experimental error of 5%, and any library member with concentrations in one or more libraries below a designated detection limit[17] was discarded. This gave a data set of a total of 16 "detectable" oligomeric species.

We then allowed DCLFit to simultaneously estimate values for all the equilibrium constants (the set of K_f values, K_{elong}, and K_{pair}).

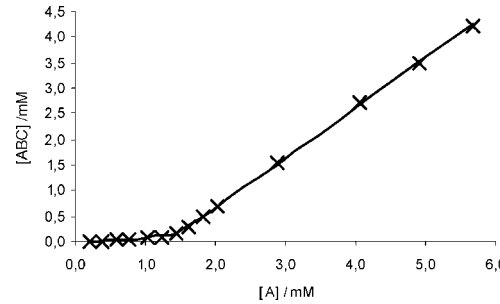


Figure 1. Experimental (×) and fitted (−) values of [ABC] as a function of building block concentrations [A] = [B] = [C].

This treatment gave estimates of 4.00 for $\log(K_{elong})$ and 2.62 for $\log(K_{pair})$, in good agreement with the actual values of 4.00 and 2.50. Figure 1 shows the "experimental" concentration data for compound ABC for different building block concentrations together with the fitted values. We have repeated this analysis for different arbitrary values of K_{elong} and K_{pair}, and also for a system that features two different aggregating species. The results are summarized in Table 1 and show that we consistently obtained estimates that give equilibrium constants within 5% of the "true" values of log K.

The above analyses demonstrate that DCLFit can indeed be used to obtain reliable estimates of K_{pair} and K_{elong} directly from library distributions. In the process of performing this analysis we obtained some important insights into how best to use DCLFit in a true experimental system in which the occurrence of aggregates is suspected. Note that aggregation will typically reveal itself in studies in which the concentrations of the building blocks are varied. At higher building block concentrations those library members that aggregate will be selectively favored over their nonaggregating counterparts. Library members rich in building blocks underrepresented in the aggregating species will also tend to increase in concentration. In libraries with relatively few members it may be difficult to distinguish these "false positives" from the aggregating library members by visual inspection of the data. However, DCLFit allows such distinctions to be made reliably. For example, in the library of entry 4 in Table 1 the aggregation of ABB also induces an increase in "bystander" CC, yet good fits of the data are only obtained with a model that includes aggregation of ABB, and not with a model that allows only CC to aggregate.

(16) We used the implementation of this algorithm from the scientific python package: Jones, E.; Oliphant, T.; Peterson, P.; et al. SciPy: Open source scientific tools for Python, 2001−, http://www.scipy.org/.

(17) For any given set of libraries, the lowest detectable concentration is taken to be 0.1% of the concentration of the most abundant species in any library in that series.

Table 1. Comparison of Values for log K_{pair} and log K_{elong} Used in Simulating the Libraries with Those Obtained by Fitting

		log K_{pair}		log K_{elong}		ρ	
	aggregating species	"true"	fitted	"true"	fitted	"true"	fitted
1	ABC	2.50	2.62	4.00	4.00	−1.50	−1.38
2	ABC	4.50	4.42	4.50	4.45	0.00	−0.04
3	ABC	4.50	4.52	3.50	3.50	1.00	1.02
4	ABB	4.00	4.08	4.50	4.50	−0.50	−0.42
5	AAB	3.50	3.62	4.50	4.52	−1.00	−0.89
	ABB	3.20	3.20	4.20	4.18	−1.00	−0.98

The accuracy of the fitted values of K_{elong} and K_{pair} depends critically on the appropriate choice of the experimental conditions, most notably on the range of the building block concentrations. This range needs to be focused on the regime where the system changes from a state in which little aggregation occurs to one where aggregates dominate the mixture. Note that, perhaps counterintuitively, the latter regime alone is not suitable for quantifying K_{elong} and K_{pair} since in this regime an increase in building block concentrations results mostly in a shift in the composition of the mixture from smaller to larger aggregates. Such a shift is not captured in the analysis of the mixture in which only the total concentrations of the aggregating (and non-aggregating) library members are measured.

How can the concentration range capturing the onset of aggregation be identified without prior knowledge of the values of K_{elong} and K_{pair}? The answer can be obtained from a qualitative analysis of the concentration of the aggregating library member as a function of the building block concentration. Figure 2 shows this relationship for different values of ρ. The two linear regimes observed for small values of ρ (the bottom traces) correspond to the aforementioned cases where aggregates are either not present (at low building block concentrations) or dominant (at high building block concentrations). The curved region is where most experimental data need to be gathered. Note that this region becomes increasingly narrow as the value of ρ decreases (i.e., as the aggregation process becomes more cooperative), making accurate determination of ρ increasingly difficult.

In conclusion, we have developed a numerical fitting tool that is able to determine equilibrium constants for formation and aggregation of species in a dynamic combinatorial library from the concentrations of library members. Our method complements existing methods for quantitative characterization of aggregating species, which require a means of determining the concentration of the aggregating species that is free in solution. These methods rely on being able to distinguish between free and aggregated species with a technique that does not perturb the noncovalent assemblies. Our method is not limited by these requirements: we can determine the equilibrium constants for *noncovalent* assembly based on the total concentrations of the most abundant *covalent* species in solution.[18] Importantly, as the covalent composition of the mixture reflects the noncovalent interactions, the integrity of the noncovalent assembly need not be maintained during analysis. This method therefore constitutes a useful tool for the dynamic combinatorial discovery of self-synthesizing self-assembling materials from complex mixures, where chromotography is essential. Its first application to a dynamic combinatorial library that was made from a self-recognizing urea containing building block will be reported elsewhere.[19]

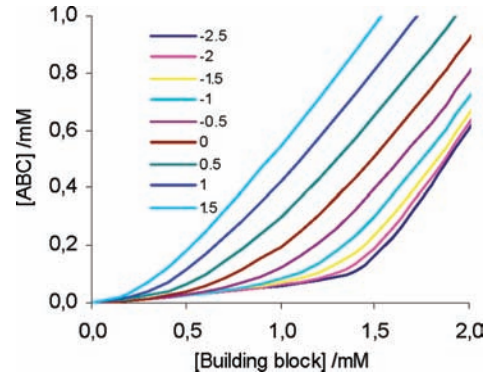


Figure 2. Typical example, log (K_{elong}) = 4.0, of the dependence of the total concentration of an aggregating library member (in this case ABC) on the building block concentration for a set of values of log ρ.

Acknowledgement. We thank the Royal Society, EPSRC, and COST CM0703 for financial support.

Supporting Information Available: Simulation and fitting details. This material is available free of charge via the Internet at http://pubs.acs.org.

OL901656X

(18) This is because the free concentration of an aggregating library member is related by a network of equilibrium constants to the concentrations of the other library members. Thus the free concentration need not be measured directly as it is still reported on by the changes in concentrations of other library members with which it shares one or more building blocks.

(19) Hunt, R. A. R.; West K. R.; Ludlow, R. F.; Camp, C.; Otto, S. *Submitted for publication.*