

RESEARCH ARTICLE

Open Access

# Deconvolution of a multi-component interaction network using systems chemistry

Soumyadip Ghosh<sup>1</sup>, Pritam Mukhopadhyay<sup>1,2</sup>, Lyle Isaacs<sup>1\*</sup>

## Abstract

We describe the stepwise construction of an 8-component self-sorted system (**1** - **8**) by the sequential addition of components. This process occurs via a large number of states ( $2^8 = 256$ ) and even a larger number of pathways ( $8! = 40320$ ). A pathway (**5, 6, 7, 8, 4, 3, 2**, then **1**) that is self-sorted at every step along the way has been demonstrated experimentally. Another pathway (**1, 8, 3, 5, 4, 7, 2**, then **6**) resembles a game of musical chairs and exhibits interesting shuttling of guest molecules among hosts. The majority of pathways - unlike the special ones described above - proceed through several non self-sorted states. We characterized the remainder of the 40320 pathways by simulation using Gepasi and describe the influence of concentration and binding constants on the fidelity of the self-sorting pathways.

## Background

Chemical events that transform a complex system from one well defined state into a different well defined state are of critical importance in both biotic and abiotic systems [1-3]. For example, inside the cellular environment, signal transduction proceeds through a sequence of steps that transforms the system from one state to a completely different state. In single cell and multi-cellular organisms these signal transduction pathways are typically guided by various protein-protein interactions that are in turn controlled by genetic regulatory networks. A recent study revealed that the transcription regulatory networks in yeast *Saccharomyces cerevisiae* involve 4549 physical interactions between 3278 yeast proteins where as genetic regulatory network is formed 1289 directed positive or negative direct transcriptional regulations within a set of 68 proteins [4,5]. Such protein-protein interaction networks define pathways for the propagation of various signals such as phosphorylation and allosteric regulation of proteins. Another study on *Escherichia coli* was able to identify 1079 regulatory interactions out of which 741 interactions are involved in the network that regulate of amino acid biosynthesis, flagella biosynthesis, osmotic stress response, antibiotic resistance, and iron regulation [6]. The complex

network of interactions that characterizes biological systems results in remarkable emergent properties that ultimately give rise to life itself.

The success of *systems biology* [1,7-11] in the reconstruction of complex functional systems based on a fundamental understanding of the behavior of a series of biological components has served as a stimulus for chemists to begin to develop *systems chemistry* with the goal of creating functional systems by combining the behaviors of well characterized chemical building blocks [12-14]. For example, the Ghadiri group has explored the behavior of complex systems comprising peptides, enzyme, and/or DNA toward the construction of systems that display self-replication, Boolean logic, and that even are subject to evolutionary pressures [15-20]. In another line of inquiry, the development of supramolecular aggregates (e.g. rotaxanes and related structures) that undergo well defined structural changes (e.g. shuttling) in response to suitable stimuli (e.g. electrochemistry, pH, photochemistry, chemical) has been shown to form the basis for molecular machines [21-23]. Although much work has been done to create individual functional components of future molecular machines, less work has been directed toward developing methods to integrate multiple components into a larger system and to allow efficient communication between different components as described above for *Saccharomyces cerevisiae* and *Escherichia coli*.

\* Correspondence: m\_pritam@mail.jnu.ac.in

<sup>1</sup>Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA

Full list of author information is available at the end of the article

As first steps toward complex functional chemical systems, we have been developing multi-component systems that form a well-defined state consisting of a single set of supramolecular aggregates. We, and others [24-35], refer to such systems as *self-sorting* systems. For example, we selected 10 compounds from the literature well known to form supramolecular aggregates driven by H-bonding interactions and showed that this collection of aggregates forms even when all 10 components are mixed [36]. Subsequent work by our group has shown that thermodynamic self-sorting processes can be designed to occur in water using cucurbit[n]uril (CB[n]) molecular containers as hosts and can even be engineered to create systems that slowly transform from kinetic self-sorted to thermodynamic self-sorted states [37,38]. Most recently, we have begun to study the ability of such systems to respond to suitable chemical stimuli (guest addition) which resulted in artificial chaperones for folding of non-natural oligomers, pH controlled inter-aggregate molecular shuttles, metal-ion triggered folding and assembly of a heterochiral double helical structure, and as a method to control enzymatic catalytic processes [39-42].

In this paper we describe a new eight component thermodynamic self-sorted mixture comprising **1** - **8** (Figure 1) and the stepwise preparation of this eight-component mixture in eight steps. The fact that this eight-component mixture comprises a thermodynamic self-sorted state does not ensure that the precursor mixtures containing one, two, three, four, five, six or seven

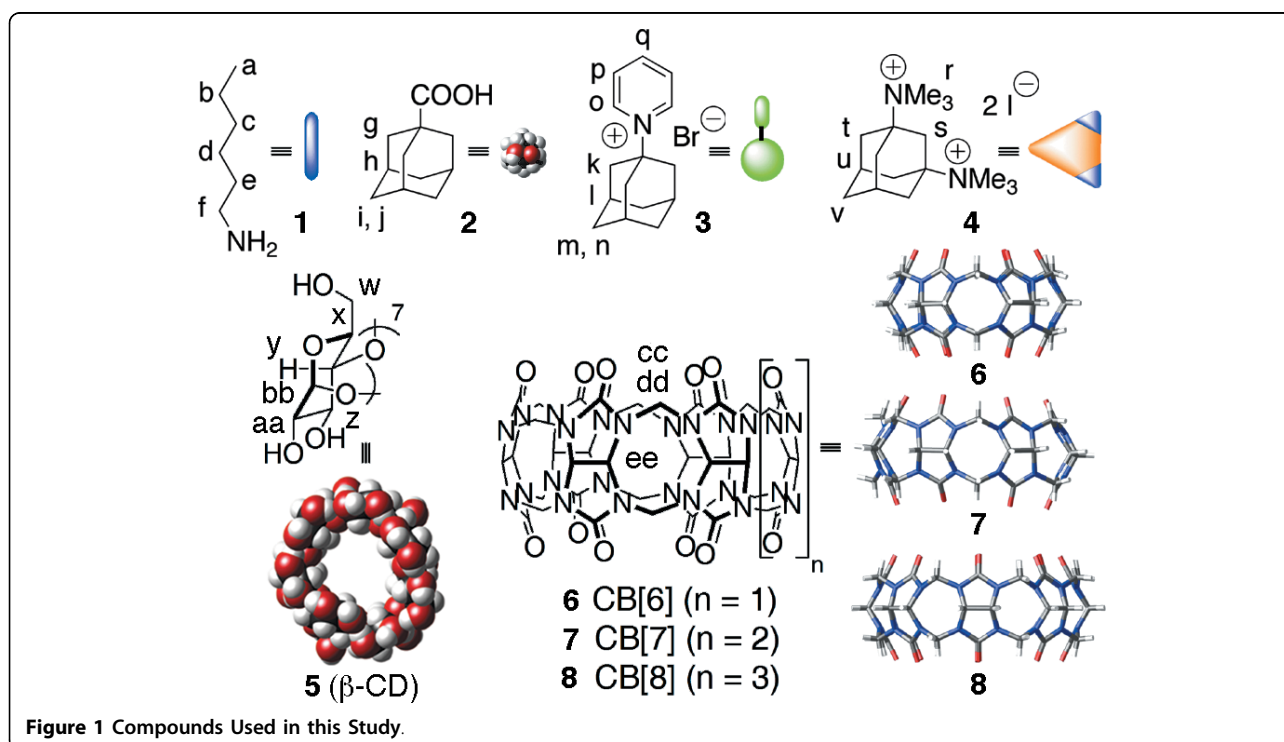
of the eight components also constitute self-sorted states. The identities of the host-guest pairs present within smaller mixtures can and do in many cases differ dramatically from those present in the *n*-component mixture. In this paper we study all of the 40320 pathways by a combination of experiment and simulation. We describe an interesting experimental pathway that is self-sorted at every step along the way that we refer to as a self-sorted pathway. We use the results obtained from simulations of the system to obtain further insights into the behavior of self-sorting systems.

## Results and Discussion

This results and discussion section is organized as follows. First, we discuss the construction of an 8-component self-sorting system comprising hosts **1** - **4** and guests **5** - **8** in water. Next, we describe three experimental pathways that involve the stepwise construction of the 8-component self-sorted mixture. Subsequently, we discuss some of the theoretical considerations involved in the stepwise construction of an *n*-component mixture and use simulations to investigate the remainder of the 40320 pathways. Finally, we discuss the implications of these results toward the use of self-sorting processes as the basis for the creation of complex functional systems.

### Selection of the Chemical Components Used in this Study

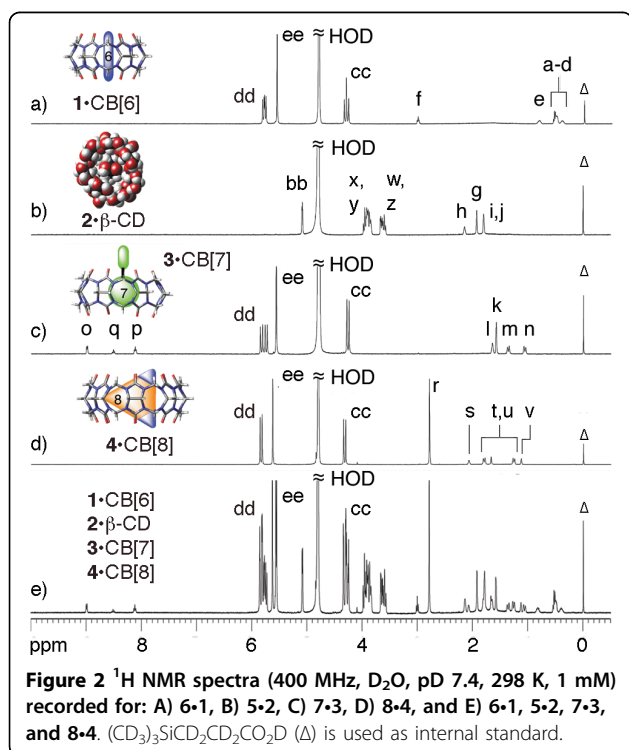
To design an 8-component social self-sorting mixture [37], we selected three members of the CB[n] family



(6 - 8, CB[6] - CB[8]) because it is well-known that CB [n] compounds bind cationic guests with high affinity and selectivity in water [43]. We choose  $\beta$ -cyclodextrin ( $\beta$ -CD) 5 as the fourth host in our study because it is commercially available and binds to a wide range of guests with low selectivity [44]. After some experimentation we selected compounds 1 - 4 as our guests. Of critical importance in the selection of 1 - 4 were: 1) their host-guest complexes undergo slow exchange on the chemical shift time scale and 2) exhibit distinct changes in chemical shift upon complexation such that the composition of the mixture can be conveniently monitored by  $^1\text{H}$  NMR spectroscopy.

### The System Comprising 1 - 8 Undergoes High Fidelity Self-Sorting

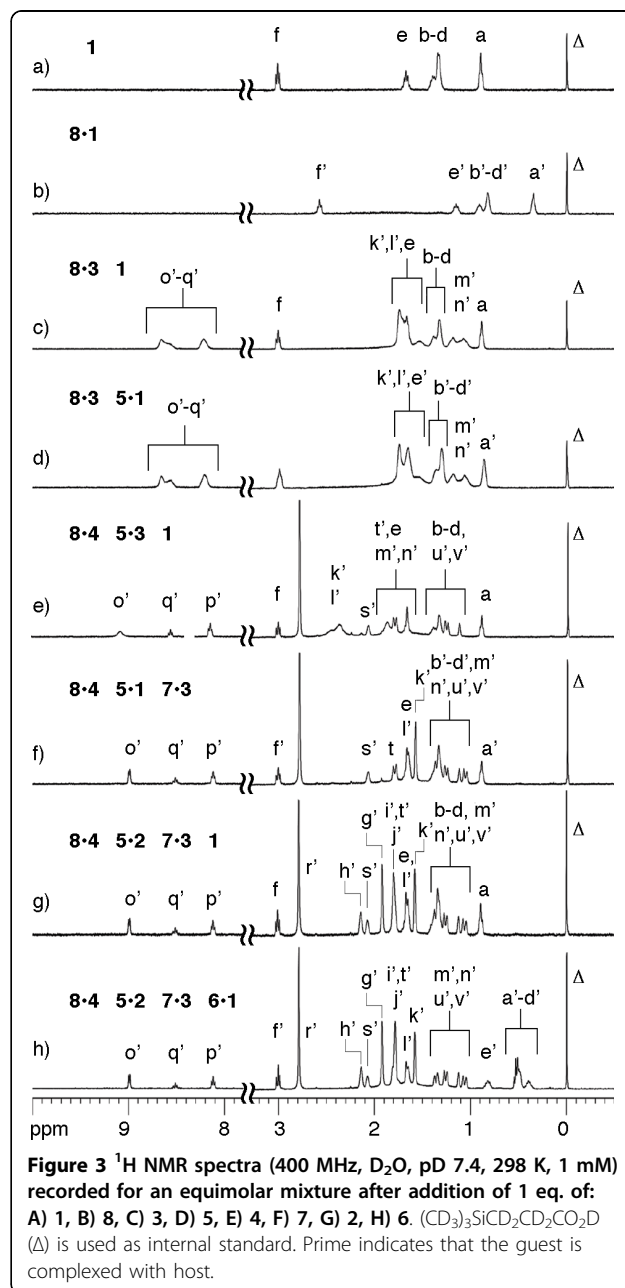
As a first step toward the build up of an  $n$ -component self-sorted mixture in  $n$ -steps by way of  $n$  self-sorted states we prepared a mixture comprising 1 - 8 and verified that the final eight component mixture was self-sorted [Additional file 1]. Figure 2 shows the  $^1\text{H}$  NMR spectra of 6•1, 5•2, 7•3, and 8•4, and a mixture of 1 - 8. The  $^1\text{H}$  NMR spectrum of the mixture (Figure 2e) is simply equal to the sum of the  $^1\text{H}$  NMR spectra of its components (Figure 2a - 2d). This spectroscopic earmark confirms that the eight-component system comprising guests 1 - 4 and hosts 5 - 8 undergoes a high fidelity self-sorting process delivering a mixture of 6•1, 5•2, 7•3, and 8•4.



### Stepwise Construction of the Eight-Component Self-Sorted Mixture

We were gratified that a mixture comprising 1 - 8 underwent high fidelity self-sorting process and decided to explore the stepwise construction of the final eight-component self-sorted state.

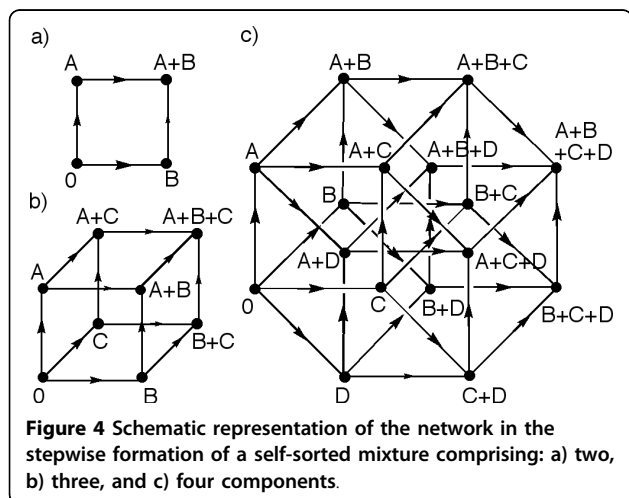
*Pathway 1: Molecular Musical Chairs.* A particularly interesting pathway for the construction of the 8-component self-sorting system comprising 6•1, 5•2, 7•3, and 8•4 involves the stepwise addition of 1, 8, 3, 5, 4, 7, 2, and then 6. The  $^1\text{H}$  NMR spectra recorded after each of the eight steps are shown in Figure 3. Remarkably, at



each of the seven intermediate steps a single well-defined set of resonances is observed by  $^1\text{H}$  NMR and each of the intermediate states is therefore self-sorted. This self-sorted pathway is particularly interesting since it proceeds by alternate addition of guest and hosts and resembles the reverse of the game of musical chairs. We whimsically refer to this as a molecular musical chair pathway. In (molecular) musical chairs,  $n$  children (guests) march to music around  $n-1$  chairs (hosts); when the music stops  $n$  children compete for  $n-1$  chairs. One of the children (guests) does not find a chair (hosts). The construction of complex self-sorting mixture resembles the microscopic reverse of this system. During this molecular musical chair pathway shuttling of guests between hosts is common. For example, in the pathway **1, 8, 3, 5, 4, 7, 2, 6** the addition of **4** to the mixture of **8•3** and **5•1** triggers the movement of **3** from **8** to **5** with a concomitant dissociation **1** from **5** (Figure 3e). Another example of guest shuttling is the behavior of compound **1** which initially forms a complex with **8** at the beginning of the pathway whereas in the final mixture it is bound to **6** and it finds its way to its final destination via **5** (Figure 3f). Similarly, guest **3** initially complexes with host **8**, then takes up residence inside **5** and finally complexes with **7** in the 8-component mixture. Host  $\beta$ -cyclodextrin **5** plays an important role in this process because it can bind to a wide range of positively charged and neutral guests with modest affinity and low selectivity [44]. As such, **5** serves as a binding depot for guests during the shuttling between high affinity binding sites in response to the addition of guests.

### Statistical Considerations

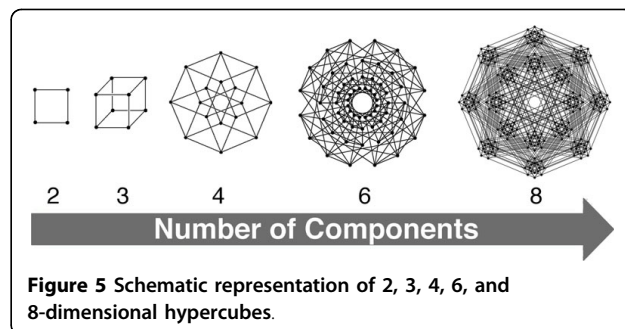
This section presents some of the statistical considerations involved in the stepwise build-up of a multi-component mixture. Figure 4 depicts the stepwise construction of two-, three-, and four-component



systems. For example, a two-component system can be made by the addition of either one of the two components in the first step followed by the addition of the other in the second step (A then B, denoted as AB; B then A, denoted as BA). We refer to these possibilities as pathways, which can be represented by a square (Figure 4a) where the vertices represent four states and arrows along the edges indicate the direction of increasing number of components. In this stepwise build-up of the two-component mixture,  $2^2$  states are conceivable (e.g. 0, A, B, and AB) and are connected by  $2!$  pathways. Similarly a three-component mixture can be built up in 6 ( $3!$ ) ways (ABC, ACB, BAC, BCA, CAB, and CBA) involving eight ( $2^3$ ) states. For a three-component mixture the various paths and states can be represented by the edges and vertices of a cube (Figure 4b). By analogy, a four-component mixture can be built up by way of 16 ( $2^4$ ) states by adding components by 24 ( $4!$ ) different pathways. This process can be depicted as a four-dimensional hypercube (Figure 4c). In general, the construction of an  $n$ -component mixture in  $n$  steps involves  $2^n$  states and there are  $n!$  different pathways to reach the final state. Such an  $n$ -component system can be represented as an  $n$  dimensional hypercube containing  $2^n$  vertices and  $2^{n-1}n$  edges. As the number of components is increased the corresponding hypercube graph reflects the enormous complexity of larger systems (Figure 5). We thank a reviewer for pointing out that a major difference between this hypercube graph model and actual protein signaling networks is that the latter are known to be scale-free whereas each state in Figure 4 is connected to exactly four other states.

### Computational Approach Towards a Global Understanding of the Stepwise Construction of a Four Component System

Given that the experimental systems described in this paper are at thermodynamic equilibrium, a detailed knowledge of all the initial concentrations and values of  $K_a$  is sufficient for the complete description of the system. These systems are, therefore, quite amenable to computational approaches. We use the program Gepasi

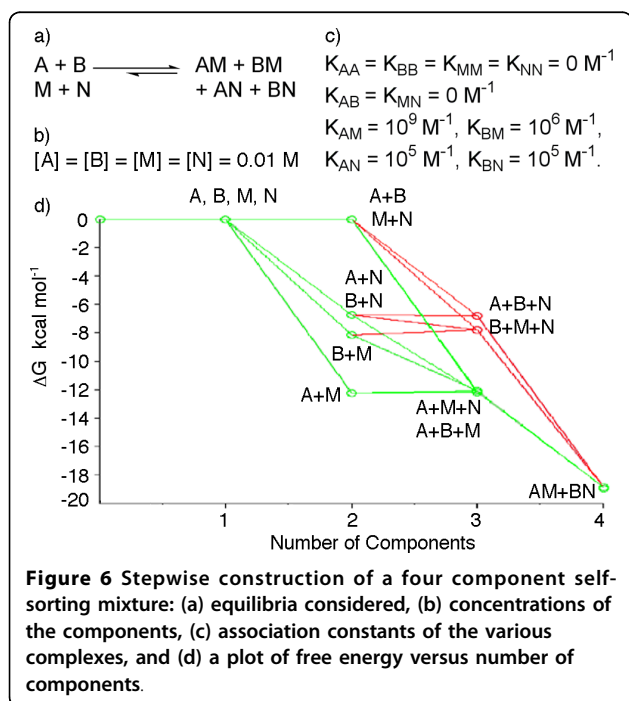




[45-47] to simulate the steady state concentrations of a multi-component system by providing an interaction model, initial concentrations, and  $K_a$  values as inputs. In this section we present the simulation of a hypothetical 4-component system that sets the stage for the complete deconvolution of the 40320 pathways to the 8-component mixture (1 - 8) described above.

Figure 6 shows a hypothetical system comprising two hosts (A and B) and two guests (M and N) subject to the constraints on concentration and values of  $K_a$  given (Figure 6b and 6c). This system is hypothetical and the chosen values of  $K_a$  are not intended to correspond to any of the experimental systems described in this paper. In processing the output of the Gepasi simulations [Additional files 2 and 3], we define a non-self-sorted state as one where one or more components or complexes have mole fractions  $0.1 \leq \chi \leq 0.9$ . Conversely, all components and complexes of self-sorted states have mole fractions either less than 0.1 or more than 0.9. This definition corresponds roughly to our ability to detect minor species by  $^1\text{H}$  NMR.

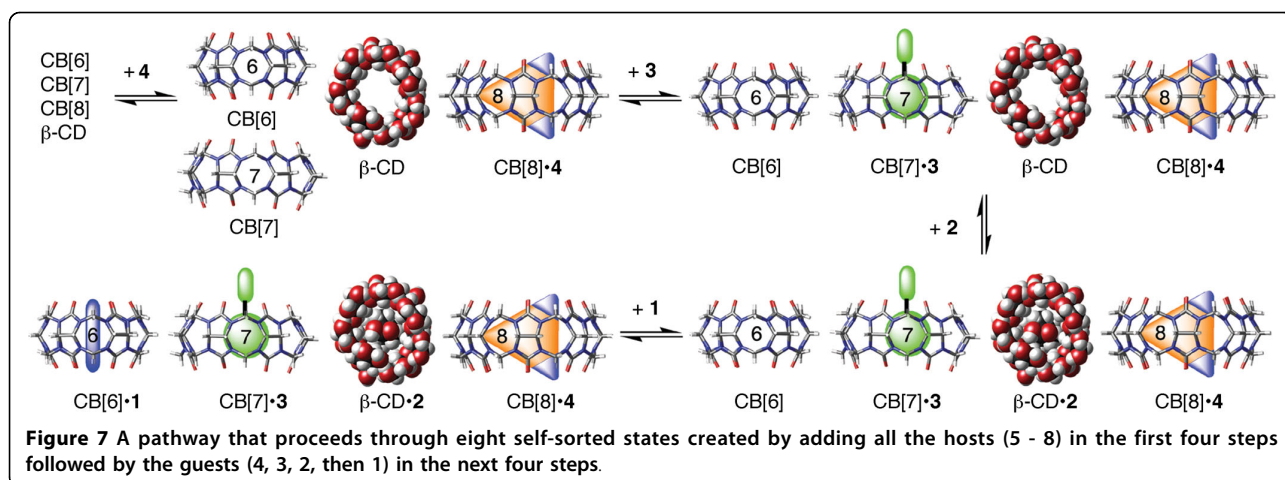
A plot of  $\Delta G^\circ$  versus number of components for each of the  $2^4$  (16) states of the system is shown in Figure 6d. Using a Matlab code [Additional file 1] we have colored self-sorted states with green dots and non-self-sorted states with red dots. Furthermore, pathways that connect two self-sorted states have been colored green; paths that connect a non-self-sorted state with a self-sorted state or two non-self-sorted states have been colored red.



sorted and even more interesting is the fact that of the 24 pathways for the construction of the four-component mixture, 12 consist entirely of self-sorted states. We refer to them as self-sorted pathways. Several other features of this system deserve comment. First, although the stepwise formation of a multi-component complex mixture is a function of path, all paths must by definition lead to an identical final state under thermodynamic control. However, changing the sequence of addition of components leads to completely different sets of complexes along the way. Second, some pathways may be trivial. For example, addition of hosts (e.g. A) followed by their most tight binding guests (e.g. M) does not lead to interesting stimuli responsive changes in composition. Third, stimuli responsive movement of guest is common. For example, the state of the system comprising A and N gives complex AN after the second step despite the fact that BN is formed after step 4. The controlled movement of N from host A to host B - driven by the free energy inherent in the 1000-fold difference in equilibrium constant between  $K_{AM}$  ( $10^9 \text{ M}^{-1}$ ) and  $K_{BM}$  ( $10^6 \text{ M}^{-1}$ ) - signals the presence of guest M. These observations made in the simulation of this four-component system are conceptually related to the experimental observations made for the 8-component system comprising 1 - 8 described above.

#### Experimental Investigation of Other Pathways that Lead to the Eight Component Self-Sorted System

Given the successful demonstration of a molecular musical chair pathway and stimulated by the computational results on the four-component system described above, we decided to experimentally investigate some alternative pathways that might prove instructional. One such pathway involves addition of the hosts 5 - 8 in the first four steps followed by the addition of the guests 1 - 4 in the next four steps is depicted in Figure 7 (See Supporting Information for NMR spectra). Although this pathway apparently seems straightforward, there are a few things about it that are noteworthy. Although it is not important to maintain any particular order of addition of hosts (5, 6, 7, and 8) in the first four steps since they do not associate with one another, it is critical to maintain a specific sequence (4 then 3 then 2 then 1) in the addition of guests for the remaining states to be self-sorted. If compound 1 was added prior to the addition of 3 that would result in a non-self-sorted mixture because the 6•1 complex is only 10-fold tighter than the 7•1 complex [48,49]. Although 3 has 1000-fold higher affinity towards 7 than 8 [49], a six-component state containing 2, 3, and 5 - 8 is a non self-sorted mixture because 2 also binds tightly to 7. In the absence of 4 - which has high selectivity toward 8 ( $K_a = 1.11 \times 10^{11} \text{ M}^{-1}$ ) relative to 7 ( $K_a = 6.42 \times 10^4 \text{ M}^{-1}$ ) - compound 3



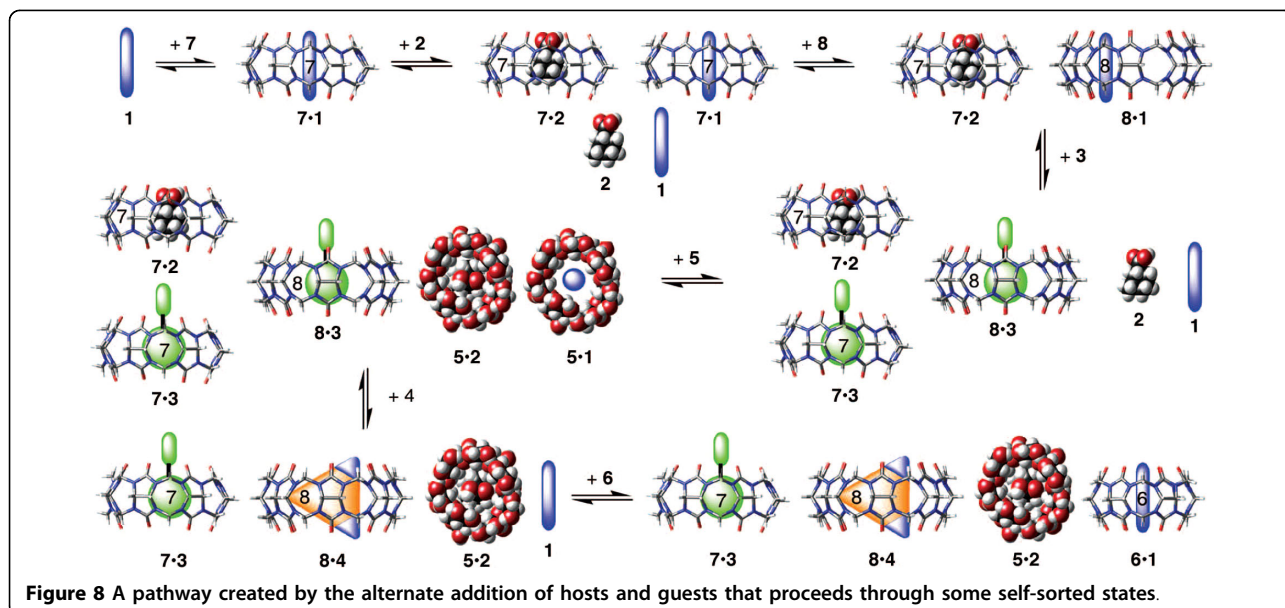
will be partially complexed with CB [8] ( $K_a = 2.00 \times 10^9 M^{-1}$ ), forcing the formation of **7·2** ( $K_a = 3.23 \times 10^8 M^{-1}$ ) in the mixture [49].

In contrast to the examples discussed so far, the majority of the pathways are non self-sorted. For example: a pathway involving the alternate addition of host and guest - **1, 7, 2, 8, 3, 5, 4** then **6** - proceeds through several non-self-sorted states (Figure 8, see Supporting Information for NMR spectra). For example, addition of **2** in the third step leads to a non self-sorted mixture because both guests (**1** and **2**) have substantial affinities towards **7**. Addition of compound **8** in the fourth step restores the state of the system from non self-sorted to self-sorted by selectively sequestering **1** to form **8·1** and **7·2**. Addition of **3** in the fifth step leads to a non self-sorted mixture and the system retains its non self-sorted state even after the addition of **5** in the sixth step.

Addition of **4** resumes self-sorted state of the system in the seventh step. The comparable affinities of guests towards hosts lead to non self-sorted mixtures in the fifth and sixth steps driven by the minimization of the overall free energy of the system. The large free energy release upon formation of **8·4** drives **3** to choose **7**, which in turn drives **2** to choose **5**. A complete NMR study of three additional pathways: A) **1, 3, 2, 4, 8, 5, 7**, and **6** B) **6, 1, 5, 2, 7, 3, 8**, and **4** and C) **7, 3, 6, 1, 5, 2, 8**, and **4** are reported in an additional material file [Additional file 1].

#### Computational Approach Towards Global Understanding of the Experimental System Comprising 1 - 8

Examination of the experimental pathways described above and in the additional material provided insights into the formation of multi-component mixtures that



were not apparent at the outset of our experiments. Rather than undertaking the unappealing prospect of investigating all 40320 (8!) pathways by experiment - we performed simulations of the system comprising **1** - **8** using the values of  $K_a$  shown in Table 1 with Gepasi [Additional files 4 and 5] [45-47]. The binding constants for **6•1**, **7•2**, **7•3**, **8•3**, **7•4**, and **8•4** were reported previously [49]. We estimated the binding constants of **1**, **2**, **3**, and **4** for **5** based on literature precedents; the binding constants of **1** with **7** and **8** were estimated by extrapolation of results from our laboratory and Mock and co-workers [48].

Figure 9a shows the free energy of all 256 ( $2^8$ ) possible states that may arise in the stepwise formation of an eight-component mixture. As we observed experimentally by  $^1\text{H}$  NMR, the final state of the simulation represents a self-sorted state. The simulation also provided mole fraction values for each component in the remaining 255 states, which were used to identify whether a particular state was self-sorted or non-self-sorted. In Figure 9a, self-sorted states are colored green and non self-sorted states are colored red. Although the initial and final states are the same, the intermediate states have distinct sets of complexes populated. This comprehensive knowledge of the compositions of all 256 states allowed us to predict the outcome of all 40320 pathways. Figure 9b represents the complete interaction network - all 40320 pathways - that was previously described by an eight dimensional hypercube.

Encouraged by the complete deconvolution of the interaction network involved in the stepwise formation of an eight-component self-sorted system, we decided to examine the behavior of various subsets of the pathways. There are four different kinds of transformations possible upon addition of a single component: a self-sorted state to another self-sorted state, a non self-sorted state to another non self-sorted state, a non self-sorted state to a self-sorted state, and a self-sorted state to a non self-sorted state (Figure 9c - f). Although the transformation from a self-sorted state to another self-sorted state is observed throughout the graph (Figure 9c), transformation from a non self-sorted state to another non self-sorted state is spatially segregated (Figure 9d) to the upper right hand corner. The formation of a highly organized self-sorted state from a disordered non

self-sorted state upon addition of a compound is associated with a substantial decrease in overall free energy of the system as indicated by the fact that the slope of the lines in Figure 9e are larger than those those in Figure 9c, d, and 9f. In contrast, some transformations from a self-sorted state to a non self-sorted state are associated with a decrease in free energy while others are slightly uphill in energy due to statistical (e.g. entropic) considerations (Figure 9f). For example, a six component mixture **1**, **2**, **3**, **5**, **6**, and **8** forms a self-sorted state consist of **1•6** ( $\chi = 0.99$ ), **2•5** ( $\chi = 0.99$ ), and **3•8** ( $\chi = 0.99$ ) ( $\Delta G^\circ = -28.9 \text{ kcal mol}^{-1}$ ). Addition of **4** results in a non-self-sorted seven component mixture containing **1•6** ( $\chi = 0.99$ ), **2** ( $\chi = 0.47$ ), **3** ( $\chi = 0.43$ ), **2•5** ( $\chi = 0.52$ ), **3•5** ( $\chi = 0.47$ ), **3•8** ( $\chi = 0.85$ ) and **4•8** ( $\chi = 0.91$ ) ( $\Delta G^\circ = -31.1 \text{ kcal mol}^{-1}$ ) and the transformation is associated with  $\Delta\Delta G^\circ = -2.2 \text{ kcal mol}^{-1}$ . The take home message is that maintaining self-sorted pathways usually requires the release of substantial amount of free energy at each step along the way (e.g. Figure 9c exhibits consistent stepwise decreases). Conversely, steps that do not result in a lowering of free energy of the system generally result in non-self-sorted states (e.g. many of the lines in Figure 9f are flat).

#### Effect of Number of Components on Self-Sorted States

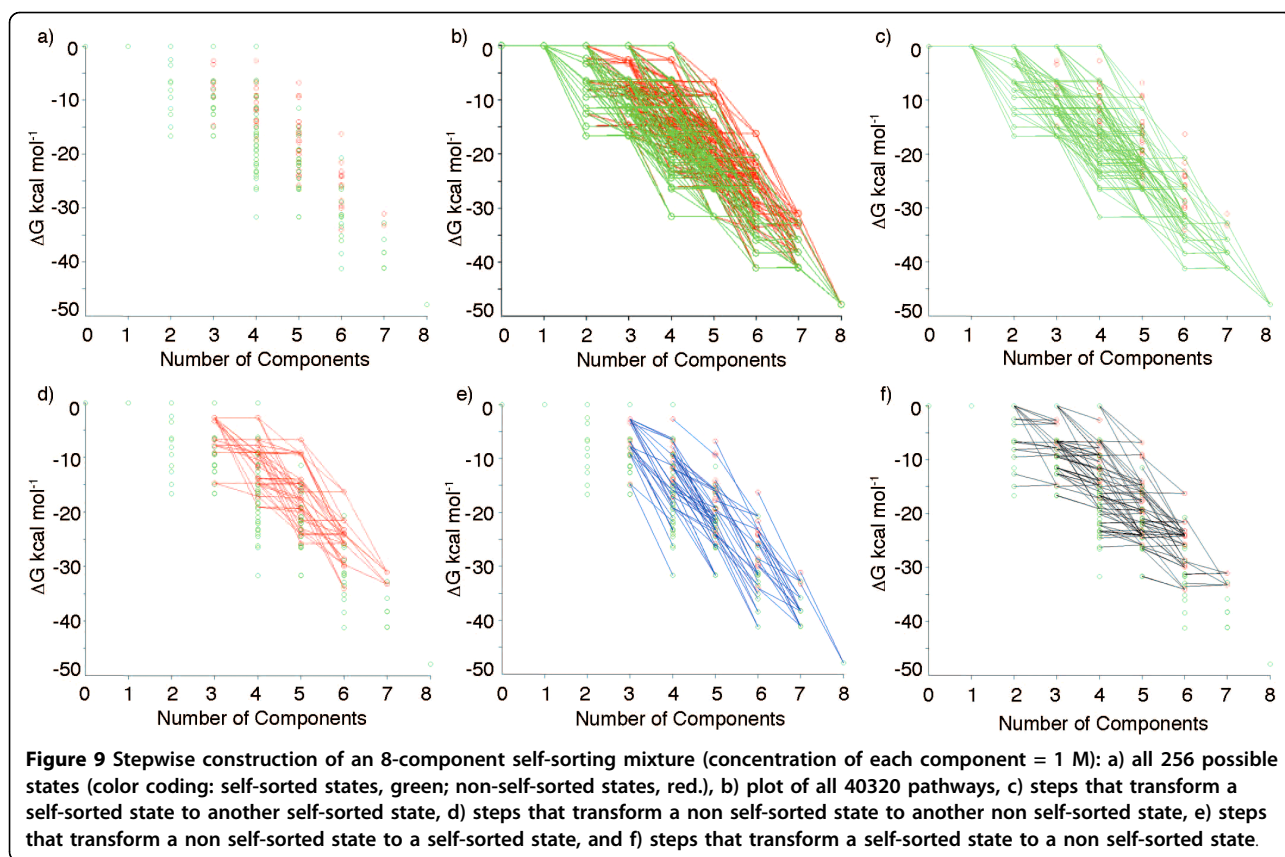
We observed several hidden patterns as well as interesting behaviors in the simulation. First, more non self-sorted states emerge when there are an intermediate number of components (e.g. four or five) due to the increase of potentially competing interactions with the increase in the number of components (Table 2). Those states that are self-sorted generally consist of high affinity and highly selective host•guest pairs because the differences in free energy due to selective binding is what drives the formation of a single set of complexes under thermodynamic control (Figure 9a).

#### Effect of Concentration of Components on Self-Sorted States and Pathways

We were interested to quantify how concentration affects the stepwise formation of the self-sorted system since Nature uses concentration as the primary method to control self-assembly processes. As the component concentration is increased from 0.01  $\mu\text{M}$  to 1 M, we observed an initial decrease in the number of self-sorted states followed by an increase above 1  $\mu\text{M}$  (Table 2 and Figure 10). This is due to the fact that at 0.01  $\mu\text{M}$  - much below their value of  $K_d$  - many complexes are fully dissociated which is self-sorted according to our definition. At concentrations comparable to  $K_d$  (e.g. 1  $\mu\text{M}$  concentration) the complexes are partially formed which represents a non-self-sorting situation according to our definition which results in fewer self-sorted

**Table 1 Values of  $K_a$  ( $\text{M}^{-1}$ ) of Different Host-Guest Complexes Used In the Simulation.**

	$\beta$ -CD	CB[6]	CB[7]	CB[8]
<b>1</b>	500	$1 \times 10^7$	$1 \times 10^6$	$1 \times 10^5$
<b>2</b>	$1 \times 10^5$	100	$3.23 \times 10^8$	100
<b>3</b>	$1 \times 10^5$	100	$1.98 \times 10^{12}$	$2.0 \times 10^9$
<b>4</b>	100	100	$6.42 \times 10^4$	$1.11 \times 10^{11}$



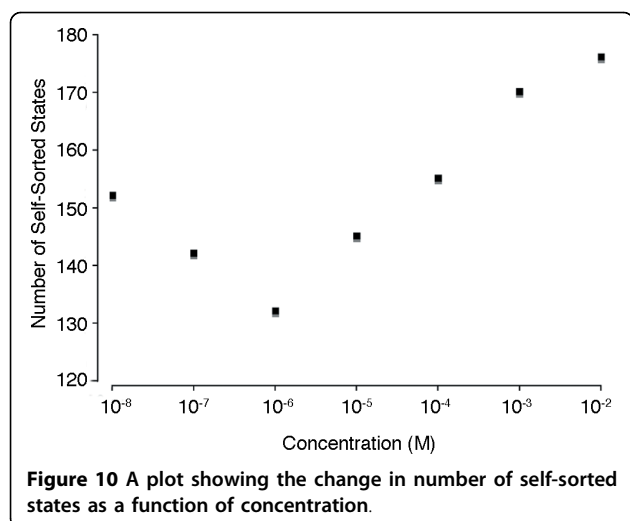
states. At higher concentrations of the components those host-guest pairs are forced to complex, which result in an increase in the number of self-sorted states. We also observed a substantial effect of concentration on the number of self-sorted pathways. As such completely self-sorted pathways are relatively uncommon. Only 5858 and 3376 pathways are completely self-sorted at 1 M and 1 mM concentration, respectively. The total number of self-sorted pathways can be calculated from

Table 2. The total number of self-sorted pathways = total number of pathways  $\times$  probability of achieving a self-sorted state at each step. When the concentration of the components is 1 M, there are total 5858 pathways that are self-sorted. ( $40320 \times 45/56 \times 49/70 \times 36/56 \times 15/28 \times 6/8 = 5858$ ). Total number of non self-sorted pathways =  $40320 - 5858 = 34462$ . Somewhat surprisingly there is no self-sorted pathway at 1  $\mu$ M concentration due to fact that the final state itself becomes non

**Table 2 Total Number of Self-Sorted and Non Self-Sorted States at Different Concentration.**

Number of Components	Number of States	Conc. (1 M)		Conc. (1 mM)		Conc. (1 $\mu$ M)	
		Self-sorted states	Non self-sorted states	Self-sorted states	Non self-sorted states	Self-sorted states	Non self-sorted states
0	1	1	0	1	0	1	0
1	8	8	0	8	0	8	0
2	28	28	0	25	3	25	3
3	56	45	11	39	17	41	15
4	70	49	21	44	26	35	35
5	56	36	20	32	24	16	40
6	28	15	13	14	14	6	22
7	8	6	2	6	2	0	8
8	1	1	0	1	0	0	1
Total	256	189	67	170	86	132	124





self-sorted at that concentration (Table 3). These results indicate that by changing the concentration - just as Nature does - we can influence the behavior of the interaction network in a complex multi-component system.

#### Effect of Sequence of Addition of Components on Pathways

Next, we decided to look at the effect of the order of addition of components on the number of self-sorted states and pathways. We conjectured that the outcome of the pathways would be very different depending on the order of addition of components (addition of all hosts followed by all guests versus alternate addition of hosts and guests). We investigated the following order of additions by simulation: 1) 576 pathways ( $4 \times 3 \times 2 \times 1 \times 4 \times 3 \times 2 \times 1$ ) for each of the following addition sequences involving addition of all four hosts followed by the addition of all four guests (HHHHGGGG) and

vice versa (GGGGHHHH), 2) 576 pathways ( $4 \times 4 \times 3 \times 3 \times 2 \times 2 \times 1 \times 1$ ) for each of the two different alternate addition sequence of hosts and guests (GHGHGHGH and HGHHGHGH)[Additional files 6, 7, 8 and 9]. Interestingly, we observed that pathways originating from the different of addition sequences set themselves apart from each other in the free energy landscape (Figure 11 a - d). In the sequence HHHHGGGG and GGGGHHHH, there is no change in the free energy observed in the first four steps followed by sharp decrease in free energy for the remaining four steps. In contrast, the alternate host-guest addition sequences (HGHHGHGH or GHGHGHGH) result in a more uniform decrease in overall free energy.

#### The Effect of Equilibrium Constants

In the previous section we demonstrated that concentration and sequence of addition of components alter the number of self-sorted states and pathways in the step-wise construction of 8-component mixture. We next used simulations to explore another key variable - equilibrium constant - that Nature uses to control its assembly processes. Table 4 shows a selection of the mean  $\pm$  standard deviations of  $K_a$  values for biomolecule•guest interactions reported by Houk [50]. Table 4 shows the values of  $\log K_a$  (mean  $\pm$  standard deviation) that we used as inputs for Gepasi simulations [Additional files 10 and 11] of hypothetical eight component systems based on  $\beta$ -CD, CB [6], CB [7], CB [8], and guests. We used the random number generator within Gepasi (using  $\log K_a$  mean  $\pm$  standard deviation values) to assign a random set of 16 ( $4 \times 4$ )  $K_a$  values. In a similar manner to that described above, we used Gepasi to simulate the behavior of the hypothetical eight-component system; we used the repetition function in Gepasi to simulate a total of 67 randomly generated systems. Figure 12 shows a plot of  $\Delta G^\circ$  versus number of

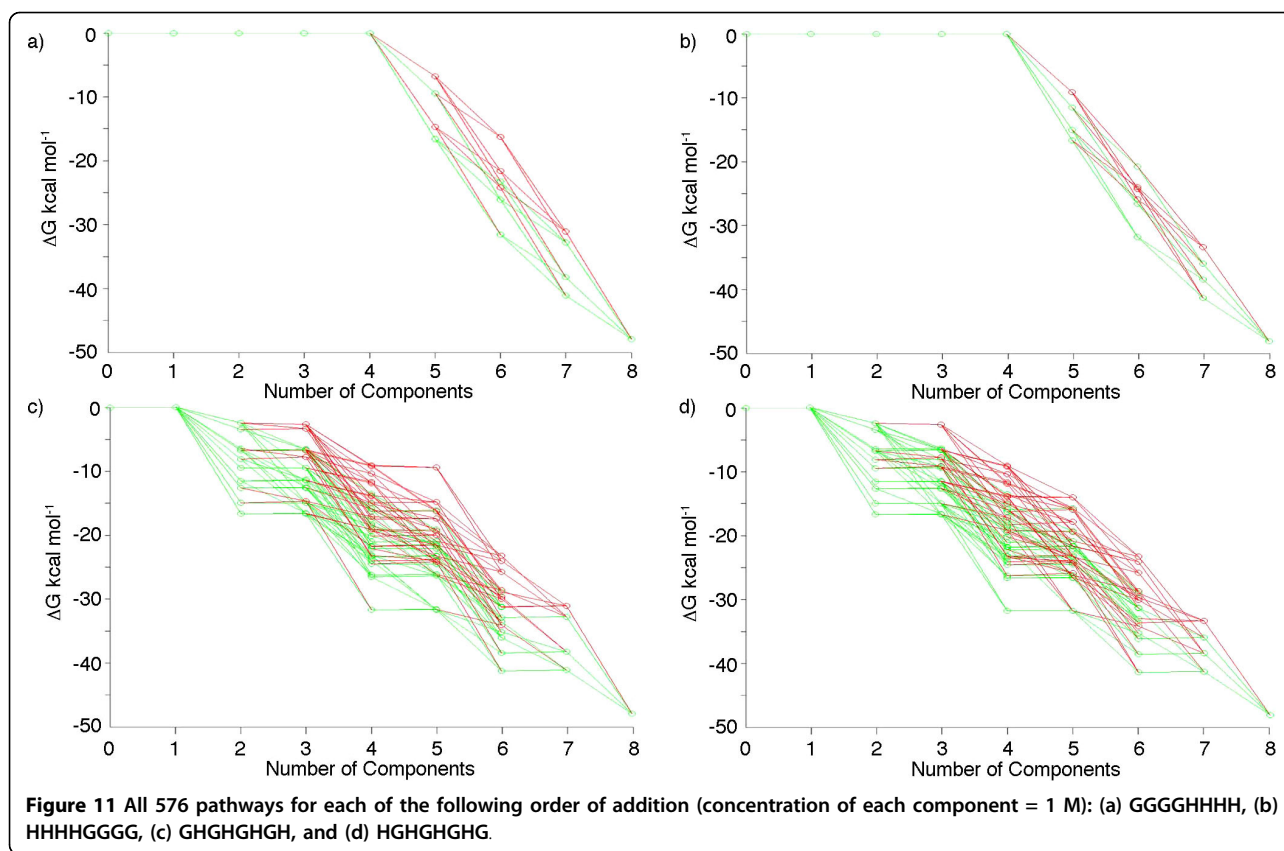
**Table 3** Effect of concentration of components and sequence of addition on the number of self-sorted pathways.

Total Number of pathways	Conc. (1 M)		Conc. (1 mM)		Conc. (1 $\mu$ M)			
	Self-sorted pathways	Non self-sorted pathways	Self-sorted pathways	Non self-sorted pathways	Self-sorted pathways	Non self-sorted pathways		
40320	5858	34462	3376	36944	0	40320		
	14.53%	85.47%	8.38%	91.62%	0%	100%		

Sequence HHHHGGGG <sup>a</sup>		Sequence GGGGHHHH <sup>a</sup>		Sequence GHGHGHGH <sup>a</sup>		Sequence HGHHGHGH <sup>a</sup>		Random sequence <sup>a</sup>	
Self-sorted paths	Non self-sorted paths	Self-sorted paths	Non self-sorted paths	Self-sorted paths	Non self-sorted paths	Self-sorted paths	Non self-sorted paths	Self-sorted paths	Non self-sorted paths
162	414	108	468	68	508	86	490	5434	32582
28.1%	71.9%	18.8%	81.2%	11.8%	88.2%	14.9%	85.1%	14.2%	85.8%

<sup>a</sup> Concentration of each component = 1 M

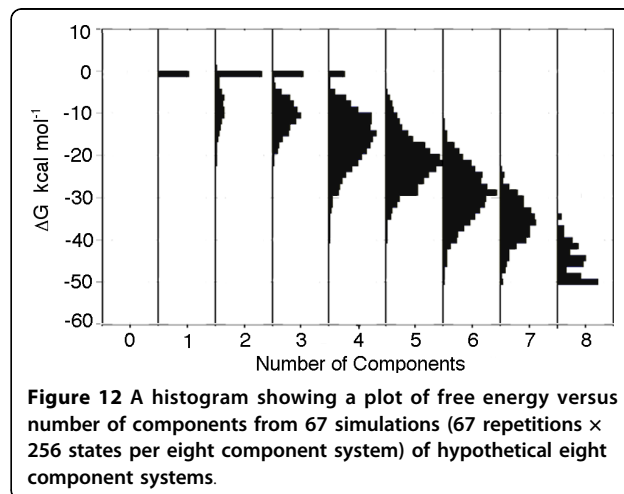


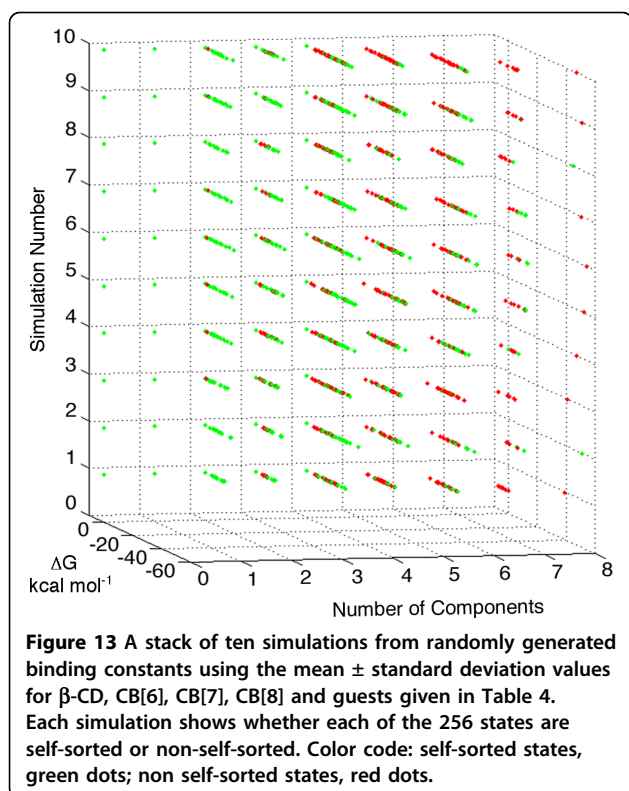
components for the 67 randomly generated systems. Figure 12 shows a steady decrease in the overall free energy of the system as the number of components increases; similar behavior was noted above in Figure 9b during the simulation using the experimentally derived  $K_a$  values of the eight component system comprising **1 - 8**. In Figure 12, the average decrease in free energy per step is controlled by the mean value ( $\log K_a$ ) of randomly generated values of  $K_a$  whereas the width of the distribution at each step is controlled by the standard

deviation of the mean; both parameters control the prevalence of self-sorting states and self-sorting pathways. Further analysis of the output of these 67 simulations (10 systems are shown in Figure 13) showed that a full 20% of the final states and approximately 3% of the total pathways are self-sorted. Given that the mean  $\pm$  standard deviation values that we employed in our simulations for  $\beta$ -CD and CB [6] - CB [8] are similar to those reported for biomolecular interactions (Table 4) we

**Table 4** Assigned mean  $\pm$  standard deviation values of  $K_a$  for the synthetic hosts used in the simulation and summary of binding constants for various interactions observed experimentally in biological systems as tabulated by Houk [50]

Host type	Guest type	Mean ( $\log K_a$ )
$\beta$ -cyclodextrin	Organic molecule	$4 \pm 2$
CB[6]	Organic molecule	$6 \pm 2$
CB[7]	Organic molecule	$8 \pm 3$
CB[8]	Organic molecule	$8 \pm 3$
Catalytic antibody	Transition state	$6.6 \pm 2.0$ [50]
Receptor	Drug	$7.3 \pm 1.5$ [50]
Antibody	Antigen	$8.1 \pm 2.0$ [50]
Enzyme	Inhibitor	$8.6 \pm 2.1$ [50]





believe that our simulations have some relevance to related biological systems. For example, although it seems that entirely self-sorted pathways are rare (3%) it also suggests that self-sorted states are relatively common (20%) when tight binding and selective hosts are present.

## Conclusions

In summary, we showed that a mixture comprising 1 - 8 forms an eight-component self-sorting mixture consisting of 6•1, 5•2, 7•3, and 8•4. We investigated selected pathways by  $^1\text{H}$  NMR and the rest by simulations for the formation of the eight-component mixture. The formation of this eight-component self-sorted system can occur by way of 256 ( $2^8$ ) states and 40320 (8!) pathways; the self-sorting nature of the final state has no bearing on whether the intermediate states are self-sorted or non-self-sorted. A particularly interesting pathway that resembles a game of musical chairs that is self-sorted at every step along the way was demonstrated experimentally. We performed simulations of the experimental system using the program Gepasi to gain further insight into the system (number of components, sequence of host/guest addition, concentration, and binding constant values). Of particular interest was the segregation of the self-sorted (non-self-sorted) pathways to the lower left (upper right) corner of plots of  $\Delta G^\circ$  versus number of components (Figure 9b) which we trace to the need for

a steady decrease in the overall value of  $\Delta G^\circ$  to maintain self-sorted states. Finally, we used simulations of hypothetical eight-component systems based on input  $\log K_a$  values (mean  $\pm$  standard deviation) and observed that self-sorting is rather common (20%) but that self-sorted pathways are not. Given that the  $\log K_a$  mean  $\pm$  standard deviation values for CB[n] hosts are similar to those of biomolecular interactions bodes well for their use in complex functional systems in the future.

Several aspects of this study may be of interest beyond the system specific considerations described above. First, similar to the molecular networks operating inside living cells, our designed experimental system are based on an intricate web of molecular recognition events. The fact that both experiments and simulations show that self-sorting is a relatively common behavior (20%) suggests that the use of self-sorting systems - particularly ones with segregated network topologies [51] - as the basis for the further development of non-natural functional complex systems is justified. One strategy that Nature uses to control the connectivity of its networks is compartmentalization. Compartmentalization serves to segregate incompatible chemical reactions and interactions and thereby greatly simplifies the overall network topology. The development of compartmentalized self-sorting systems represents a further step toward their integration with biological systems. Finally, we have begun to take steps toward using self-sorting systems to control enzymatic catalytic processes [40] which will be very important for the development of feedback loops and adaptability that are so critical in biological systems. When such approaches can be extended to more complex biological media it may be possible to use self-sorting system to interface with and exert control over portions of the interaction network of the biological system. Second, we have shown how the addition of a new component can dramatically change the composition of the multi-component mixture which suggests that self-sorting systems will be useful in sensing applications [52]. Lastly, this study highlights the power of combinations of simulation and experiment in systems chemistry. For example, we were able to use Gepasi simulation to fully explore 40320 pathways in a time comparable to that needed to investigate six pathways experimentally. These simulations highlighted that the successful generation of complex self-sorting systems relies on the availability of synthetic or biomolecular hosts that display both high affinity and high selectivity toward their guests.

## Experimental

### Materials

Compounds 1, 2, 3 and  $\beta$ -CD were obtained from commercial sources. Compound 4 and CB [6] - CB [8] were

synthesized according to the literature procedures [49,53].

#### Sample Preparation

The mixtures described in this paper were prepared as follows: 1) the calculated amounts of each component were weighed out separately and transferred to a 5 mL screw cap vial, 2) D<sub>2</sub>O (2 mL) was added, 3) the mixture was sonicated or vortexed for several minutes, 4) the pD was adjusted using conc. KOD or DCl solution, 5) the solution was stirred at room temperature overnight, 6) the solution was centrifuged, and 6) the solution was transferred to an NMR tube for analysis.

#### NMR Experiments

<sup>1</sup>H NMR spectra were measured on spectrometers operating at 400 or 500 MHz. Temperature was controlled to 298 ± 0.5 K with a temperature control module that had been calibrated using separation of resonances of methanol. All spectra were measured in D<sub>2</sub>O unless mentioned and referenced relative to external (CD<sub>3</sub>)<sub>3</sub>SiCD<sub>2</sub>CD<sub>2</sub>CO<sub>2</sub>D.

#### Simulations

Simulations were performed using Gepasi 3.30 running on a Windows XP workstation. The Gepasi output files were processed using Microsoft Office Excel 2003 and MatLab running on a Windows XP workstation. The Gepasi model files and MatLab codes used in these simulations are deposited in the Supporting Information.

#### Additional material

**Additional file 1: Supporting Information.** This file contains the <sup>1</sup>H NMR spectra for all guests **1 - 4**, hosts **5 - 8**, all host-guest complexes, and selected pathways as well as the MatLab codes used for processing the simulation output.

**Additional file 2: Gepasi Model File.** This file is the Gepasi model file used to run the simulation reported in Figure 6.

**Additional file 3: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 6.

**Additional file 4: Gepasi Model File.** This file is the Gepasi model file used to run the simulation reported in Figure 9.

**Additional file 5: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 9a.

**Additional file 6: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 11a.

**Additional file 7: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 11b.

**Additional file 8: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 11c.

**Additional file 9: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 11d.

**Additional file 10: Gepasi Model File.** This file is the Gepasi model file used to run the simulations reported in Figure 12 and Figure 13.

**Additional file 11: Excel Spreadsheet.** This file is an Excel spreadsheet that contains the data for Figure 12.

#### Acknowledgements

We thank the National Science Foundation (CHE-0615049 and CHE-0914745) for financial support.

#### Author details

<sup>1</sup>Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. <sup>2</sup>School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110067, India.

#### Authors' contributions

LI, SG, and PM designed the research. SG and PM collected the experimental data. SG performed the simulations. LI, SG, and PM analyzed and interpreted the data. LI, SG, and PM wrote the paper. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 23 March 2010 Accepted: 18 August 2010

Published: 18 August 2010

#### References

1. Bonneau R: **Learning biological networks: from modules to dynamics.** *Nat Chem Biol* 2008, **4**:658-664.
2. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402**:C47-C52.
3. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network Motifs: Simple Building Blocks of Complex Networks.** *Science* 2002, **298**:824-827.
4. Costanzo MC, Crawford ME, Hirschman JE, Kranz JE, Olsen P, Robertson LS, Skrzypek MS, Braun BR, Hopkins KL, Kondu P, et al: **YPD, PombePD and WORMPD: model organism volumes of the BioKnowledge Library, an integrated resource for protein information.** *Nucleic Acids Res* 2001, **29**:75-79.
5. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98**:4569-4574.
6. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: **Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.** *PLoS Biol* 2007, **5**:54-66.
7. Aloy P, Russell RB: **Structural systems biology: modelling protein interactions.** *Nat Rev Mol Cell Biol* 2006, **7**:188-197.
8. Butcher EC: **Innovation: Can cell systems biology rescue drug discovery?** *Nat Rev Drug Discovery* 2005, **4**:461-467.
9. Kitano H: **Systems biology: A brief overview.** *Science* 2002, **295**:1662-1664.
10. Oltvai ZN, Barabasi A-L: **Perspectives: Systems biology: Life's complexity pyramid.** *Science* 2002, **298**:763-764.
11. Westerhoff HV, Palsson BO: **The evolution of molecular biology into systems biology.** *Nat Biotechnol* 2004, **22**:1249-1252.
12. Ludlow RF, Otto S: **Systems chemistry.** *Chem Soc Rev* 2008, **37**:101-108.
13. Nitschke JR: **Systems chemistry: Molecular networks come of age.** *Nature* 2009, **462**:736-738.
14. Wagner N, Ashkenasy G: **Systems chemistry: logic gates, arithmetic units, and network motifs in small networks.** *Chem Eur J* 2009, **15**:1765-1775.
15. Ashkenasy G, Ghadiri MR: **Boolean Logic Functions of a Synthetic Peptide Network.** *J Am Chem Soc* 2004, **126**:11140-11141.
16. Ashkenasy G, Jagasia R, Yadav M, Ghadiri MR: **Design of a directed molecular network.** *Proc Natl Acad Sci USA* 2004, **101**:10872-10877.
17. Frezza BM, Cockroft SL, Ghadiri MR: **Modular Multi-Level Circuits from Immobilized DNA-Based Logic Gates.** *J Am Chem Soc* 2007, **129**:14875-14879.
18. Gianneschi NC, Ghadiri MR: **Design of molecular logic devices based on a programmable DNA-regulated semisynthetic enzyme.** *Angew Chem, Int Ed* 2007, **46**:3955-3958.
19. Saghatelian A, Yokobayashi Y, Soltani K, Ghadiri MR: **A chiroselective peptide replicator.** *Nature* 2001, **409**:797-801.
20. Ura Y, Beierle JM, Leman LJ, Orgel LE, Ghadiri MR: **Self-Assembling Sequence-Adaptive Peptide Nucleic Acids.** *Science* 2009, **325**:73-77.



21. Balzani V, Credi A, Raymo FM, Stoddart JF: **Artificial molecular machines.** *Angew Chem, Int Ed* 2000, **39**:3348-3391.
22. Kay ER, Leigh DA, Zerbetto F: **Synthetic molecular motors and mechanical machines.** *Angew Chem, Int Ed* 2007, **46**:72-191.
23. Ko YH, Kim E, Hwang I, Kim K: **Supramolecular assemblies built with host-stabilized charge-transfer interactions.** *Chem Commun* 2007, 1305-1315.
24. Kramer R, Lehn JM, Marquis-Rigault A: **Self-recognition in helicate self-assembly: Spontaneous formation of helical metal complexes from mixtures of ligands and metal ions.** *Proc Natl Acad Sci USA* 1993, **90**:5394-5398.
25. Nitschke JR: **Construction, Substitution, and Sorting of Metallo-organic Structures via Subcomponent Self-Assembly.** *Acc Chem Res* 2007, **40**:103-112.
26. Rowan SJ, Hamilton DG, Brady PA, Sanders JKM: **Automated Recognition, Sorting, and Covalent Self-Assembly by Predisposed Building Blocks in a Mixture.** *J Am Chem Soc* 1997, **119**:2578-2579.
27. Barboiu M, Dumitru F, Legrand Y-M, Petit E, van der Lee A: **Self-sorting of equilibrating metallosupramolecular DCLs via constitutional crystallization.** *Chem Commun* 2009, 2192-2194.
28. Bilgicer B, Xing X, Kumar K: **Programmed Self-Sorting of Coiled Coils with Leucine and Hexafluoro-leucine Cores.** *J Am Chem Soc* 2001, **123**:11815-11816.
29. Braekers D, Peters C, Bogdan A, Rudzевич Y, Boehmer V, Desreux JF: **Self-Sorting Dimerization of Tetraurea Calix[4]arenes.** *J Org Chem* 2008, **73**:701-706.
30. Burd C, Weck M: **Self-Sorting in Polymers.** *Macromolecules* 2005, **38**:7225-7230.
31. Jiang W, Winkler HDF, Schalley CA: **Integrative Self-Sorting: Construction of a Cascade-Stoppered Hetero[3]rotaxane.** *J Am Chem Soc* 2008, **130**:13852-13853.
32. Kamada T, Aratani N, Ikeda T, Shibata N, Higuchi Y, Wakamiya A, Yamaguchi S, Kim KS, Yoon ZS, Kim D, Osuka A: **High Fidelity Self-Sorting Assembling of meso-Cinchononimide Appended meso-meso Linked Zn(II) Diporphyrins.** *J Am Chem Soc* 2006, **128**:7670-7678.
33. Mahata K, Schmittel M: **From 2-Fold Complete to Integrative Self-Sorting: A Five-Component Supramolecular Trapezoid.** *J Am Chem Soc* 2009, **131**:16544-16554.
34. Moffat JR, Smith DK: **Controlled self-sorting in the assembly of 'multi-gelator' gels.** *Chem Commun* 2009, 316-318.
35. Taylor PN, Anderson HL: **Cooperative Self-Assembly of Double-Strand Conjugated Porphyrin Ladders.** *J Am Chem Soc* 1999, **121**:11538-11545.
36. Wu A, Isaacs L: **Self-Sorting: The Exception or the Rule?** *J Am Chem Soc* 2003, **125**:4831-4835.
37. Mukhopadhyay P, Wu A, Isaacs L: **Social Self-Sorting in Aqueous Solution.** *J Org Chem* 2004, **69**:6157-6164.
38. Mukhopadhyay P, Zavalij PY, Isaacs L: **High Fidelity Kinetic Self-Sorting in Multi-Component Systems Based on Guests with Multiple Binding Epitopes.** *J Am Chem Soc* 2006, **128**:14093-14102.
39. Chakrabarti S, Mukhopadhyay P, Lin S, Isaacs L: **Reconfigurable Four-Component Molecular Switch Based on pH-Controlled Guest Swapping.** *Org Lett* 2007, **9**:2349-2352.
40. Ghosh S, Isaacs L: **Biological Catalysis Regulated by Cucurbit[7]uril Molecular Containers.** *J Am Chem Soc* 2010, **132**:4445-4454.
41. Huang W-H, Zavalij PY, Isaacs L: **Metal-Ion-Induced Folding and Dimerization of a Glycoluril Decamer in Water.** *Org Lett* 2009, **11**:3918-3921.
42. Liu S, Zavalij PY, Lam Y-F, Isaacs L: **Refolding Foldamers: Triazene-Arylene Oligomers That Change Shape with Chemical Stimuli.** *J Am Chem Soc* 2007, **129**:11232-11241.
43. Isaacs L: **Cucurbit[n]urils: from mechanism to structure and function.** *Chem Commun* 2009, 619-629.
44. Rekharsky MV, Inoue Y: **Complexation Thermodynamics of Cyclodextrins.** *Chem Rev* 1998, **98**:1875-1917.
45. Mendes P: **GEPASI: a software package for modeling the dynamics, steady states and control of biochemical and other systems.** *Comput Appl Biosci* 1993, **9**:563-571.
46. Mendes P: **Biochemistry by numbers: simulation of biochemical pathways with GEPASI 3.** *Trends Biochem Sci* 1997, **22**:361-363.
47. Mendes P, Kell DB: **Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.** *Bioinformatics* 1998, **14**:869-883.
48. Mock WL, Shih NY: **Structure and selectivity in host-guest complexes of cucurbituril.** *J Org Chem* 1986, **51**:4440-4446.
49. Liu S, Ruspic C, Mukhopadhyay P, Chakrabarti S, Zavalij PY, Isaacs L: **The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems.** *J Am Chem Soc* 2005, **127**:15959-15967.
50. Houk KN, Leach AG, Kim SP, Zhang X: **Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes.** *Angew Chem, Int Ed* 2003, **42**:4872-4897.
51. Saur I, Scopelliti R, Severin K: **Utilization of self-sorting processes to generate dynamic combinatorial libraries with new network topologies.** *Chem Eur J* 2006, **12**:1058-1066.
52. Buryak A, Pozdnoukhov A, Severin K: **Pattern-based sensing of nucleotides in aqueous solution with a multicomponent indicator displacement assay.** *Chem Commun* 2007, 2366-2368.
53. Day A, Arnold AP, Blanch RJ, Snushall B: **Controlling Factors in the Synthesis of Cucurbituril and Its Homologues.** *J Org Chem* 2001, **66**:8094-8100.

doi:10.1186/1759-2208-1-6

**Cite this article as:** Ghosh *et al.*: Deconvolution of a multi-component interaction network using systems chemistry. *Journal of Systems Chemistry* 2010 **1**:6.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
<http://www.chemistrycentral.com/manuscript/>

