JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

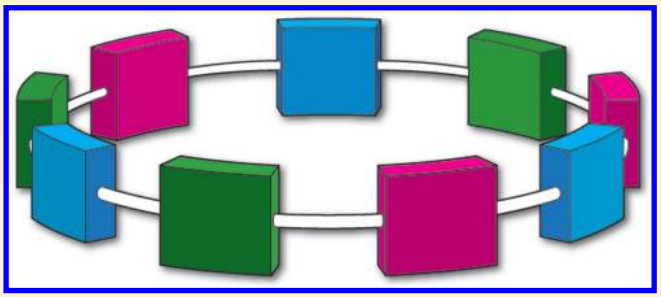
Enumeration of Virtual Libraries of Combinatorial Modular Macrocyclic (Bracelet, Necklace) Architectures and Their Linear Counterparts

Masahiko Taniguchi,* Hai Du, and Jonathan S. Lindsey*

Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, United States

Supporting Information

ABSTRACT: A wide variety of cyclic molecular architectures are built of modular subunits and can be formed combinatorially. The mathematics for enumeration of such objects is well-developed yet lacks key features of importance in chemistry, such as specifying (i) the structures of individual members among a set of isomers, (ii) the distribution (i.e., relative amounts) of products, and (iii) the effect of nonequal ratios of reacting monomers on the product distribution. Here, a software program (*Cyclaplex*) has been developed to determine the number, identity (including isomers), and relative amounts of linear and cyclic architectures from a given number and ratio of reacting monomers. The program includes both mathematical formulas and generative algorithms for enumeration; the latter go beyond the former to provide desired molecular-relevant information and data-mining features. The program is equipped to enumerate four types of architectures: (i) linear architectures with directionality (macroscopic equivalent = electrical extension cords), (ii) linear architectures without directionality (batons), (iii) cyclic architectures with directionality (necklaces), and (iv) cyclic architectures without directionality (bracelets). The program can be applied to cyclic peptides, cycloveratrylenes, cyclens, calixarenes, cyclodextrins, crown ethers, cucurbiturils, annulenes, expanded meso-substituted porphyrin(ogen)s, and diverse supramolecular (e.g., protein) assemblies. The size of accessible architectures encompasses up to 12 modular subunits derived from 12 reacting monomers or larger architectures (e.g. 13−17 subunits) from fewer types of monomers (e.g. 2−4). A particular application concerns understanding the possible heterogeneity of (natural or biohybrid) photosynthetic light-harvesting oligomers (cyclic, linear) formed from distinct peptide subunits.



INTRODUCTION

A current challenge in the field of chemoinformatics is to understand both the size of chemical space and the extent to which chemical space is filled by synthetic or virtual combinatorial libraries.[1−11] Assessing the size of chemical space is a problem that appears in diverse scientific fields ranging from drug discovery to studies of the origin of life. We recently developed a software program (*PorphyrinViLiGe*) for the enumeration of porphyrins formed upon several types of combinatorial reactions.[12] The resulting virtual libraries of porphyrins were typically found to be rich in isomers due to different patterns of substituents arrayed about the perimeter of the macrocycles.[13,14]

Macrocycles play important roles in biological systems and are found in diverse applications. Macrocycles derived from modular subunits can serve as scaffolds for structural variation by incorporation of distinct substituents with each subunit. Representative macrocycles include enterobactin,[15] cyclotriveratrylenes,[16,17] cyclens,[18,19] calixarenes,[20−22] cyclodextrins,[23] crown ethers,[24,25] cucurbiturils,[26,27] annulenes,[28] and expanded meso-substituted porphyrins[29−31] (Figure 1). More elaborate architectures that are macrocyclic in nature can be composed of peptides,[32−37] proteins,[38−41] or ribonucleic acid (RNA).[42,43]

The strategies for creation of supramolecular architectures span the range from traditional covalent chemical synthesis to reliance on self-assembly. In this regard, dynamic combinatorial chemistry (DCC) and dynamic combinatorial libraries (DCL)—where collections of molecules undergo selective chemical reactions—have become powerful approaches to generate novel biomolecules, host−guest complexes, and catalysts.[44−48] In general, the advent of new strategies for creating macrocycles[49,50] of substantial size warrants the development of quantitative methods for enumeration. Enumeration of the resulting macrocycles provides a foundation for understanding the possible composition of a given chemical sample. The virtual library thus describes the identity of all compounds and their relative amounts. Sorting the virtual library on the basis of the nature of the patterns of modular subunits (and the relative amount of a given type of subunit) can form a key element of molecular design.

Our chief interest is in the combinatorial formation of photosynthetic light-harvesting (LH) architectures. The light-harvesting antenna complexes of purple photosynthetic bacteria

Received: March 25, 2013

   dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

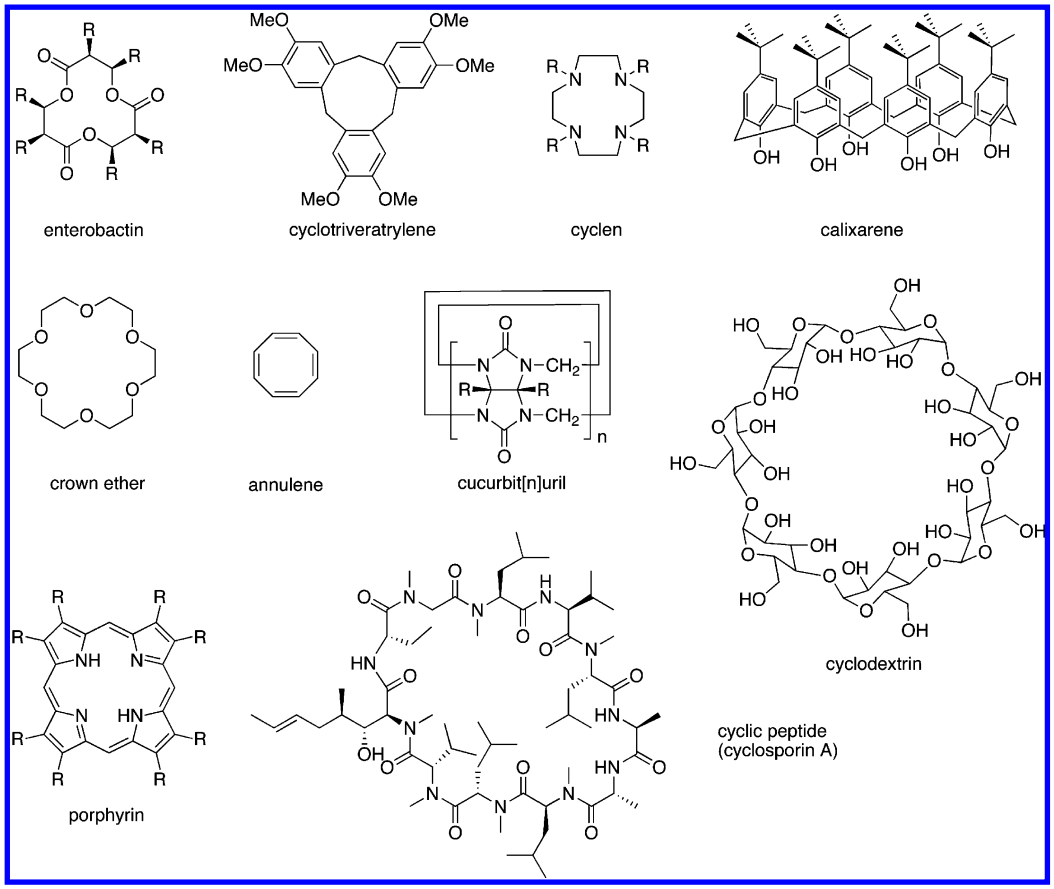


Figure 1. Representative macrocycles composed of modular subunits.

consist of oligomers of peptide dyads.[51,52] Each peptide dyad (hereafter termed a subunit) typically incorporates two or three bacteriochlorophyll *a* molecules (and zero or one carotenoid). Such subunits self-assemble in solution to give linear (i.e., acyclic) and cyclic oligomers.[53] Each peripheral light-harvesting antenna complex (LH2) typically contains 8 or 9 subunits whereas the core complex (LH1) typically contains 14−16 subunits.[54−58] Oligomers of other sizes are known, depending on bacterial strain and environmental (growth, light) conditions.[59] Heterogeneous oligomers (i.e., derived from more than one type of subunit) may also form.[60]

We recently have begun to exploit such subunits as scaffolding for the creation of biohybrid light-harvesting architectures.[61,62] In this approach, synthetic chromophores are covalently attached to one or both of the peptides that comprise the subunit, which in solution oligomerizes in a manner that resembles the natural light-harvesting system. When a single subunit type is employed in a given sample, the resulting architectures are compositionally homogeneous (although a distribution of linear and cyclic oligomers is expected). Upon consideration of the use of two, three, or more distinct peptide subunits, we were confronted with a lack of understanding of not only how many compositionally distinct cyclic oligomers could form, but also what the relative amounts of each would be. To address that specific problem, we have developed the general approaches and accompanying software program that are described herein.

In this article, we describe a program (*Cyclaplex*) for enumeration of the virtual libraries for four types of modular molecular architectures formed upon combinatorial reaction. We first describe the four architectures and highlight distinguishing features that impact enumeration. We then present established mathematical methods for enumeration (i.e., counting)[63] and calculate results concerning the effects of the number of distinct monomers and the size of the modular architecture on the number of products. The program employs such mathematical methods as well as new, generative algorithms. The generative algorithms (i.e., construction methods)[63] afford results that encompass those obtained from mathematical treatments yet are richer in chemical information (molecular identities, relative amounts of products upon combinatorial formation). The generative algorithms enable data mining to assess products with particular patterns of subunits. The chief focus here concerns the fundamentals of the program, which are illustrated with a number of examples. Use of the program should enable a quantitative description of the theoretical composition of combinatorial libraries of important linear and cyclic molecular architectures.

RESULTS

I. Four Distinct Molecular Architectures. The four molecular architectures are illustrated in Figure 2. The architectures are distinguished by the available symmetry operations.[64]

(1) Linear architectures with inherent directionality have no symmetry operations other than C_1. A macroscopic example is provided by a series of electrical extension cords (each of which has a male and female terminus; hereafter termed cords). A molecular example is given by a linear peptide, where the constituent α-amino acids

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

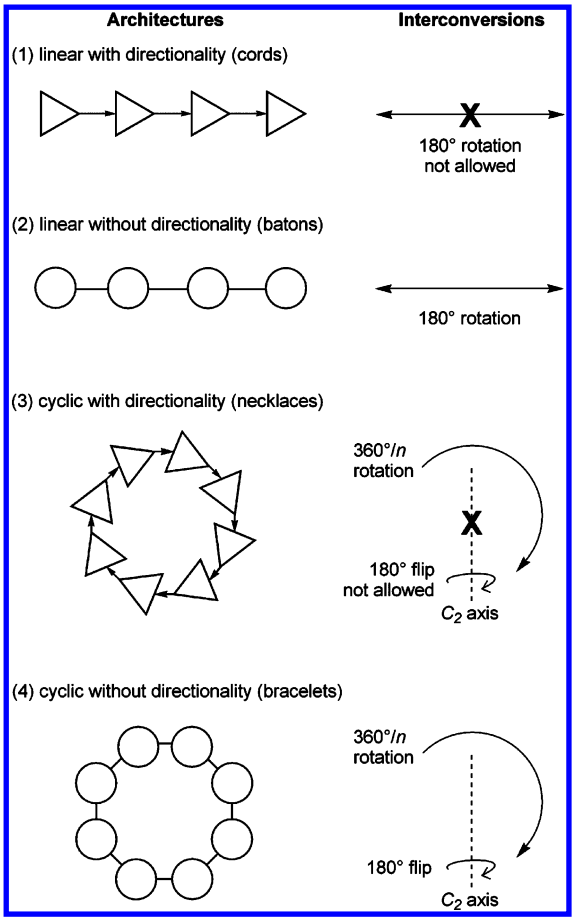


Figure 2. Illustrative molecular architectures (each composed of n modular subunits) and interconversion operations.

impart directionality and the resulting peptide has an amino terminus and a carboxylic acid terminus.

(2) Linear architectures without inherent directionality can undergo 180° rotation. A macroscopic example is a baton, whereas a molecular example consists of oligophenylenes, of which quaterphenyl is a representative member.

(3) Cyclic architectures with inherent directionality can undergo rotation in the plane by $360°/n$, where n is the number of modular subunits in the ring (i.e., the n-mer), but cannot be flipped (i.e., rotate about any of the n putative C_2 axes that bisect the molecular framework). The symmetry descriptor is C_n. A macroscopic example is a necklace (which presents only one face to the viewer), whereas a molecular example is a cyclic peptide. For example, rotation of cyclosporine (Figure 1) about a given C_2 axis would afford a stereoisomer.

(4) Cyclic architectures without inherent directionality can undergo rotation in the plane by $360°/n$ and can be rotated about any of the n C_2 axes that bisect the molecular framework. The symmetry descriptor is D_n. A macroscopic example is a bracelet (which has no preferential face and can be flipped over), whereas a molecular example includes a planar metalloporphyrin.

The self-assembled light-harvesting oligomers are composed of asymmetric subunits (each composed of an α-peptide and a β-peptide); hence, the appropriate descriptive architectures are (linear) cords and (cyclic) necklaces. By comparison, considering each dyad subunit as a linearly symmetric entity, the relevant descriptive architectures are (linear) batons and (cyclic) bracelets.

Identification of molecules in terms of the fundamental architectures shown in Figure 2 enables application of established mathematical treatments for enumeration. Our chief interest in enumeration concerns assessment of the number, identity, isomeric grouping, and relative amount of products formed in a combinatorial reaction leading to such architectures. Grouping of isomers is a critical part of enumeration, as described in the next section.

II. Handling of Isomers. Molecular libraries are known to contain large numbers of isomers,[11] which typically can be grouped in a variety of ways. An approach for grouping isomers is essential in combinatorial chemistry in general and upon use of modular subunits in particular. The necessity to handle isomers effectively stems from the fact that (i) combinatorial libraries composed of modular subunits can be exceedingly rich in isomers (e.g., >95% of the total number of products can be an isomer of at least one other product)[12] and (ii) mass spectrometry is often used as a primary tool for analysis in combinatorial chemistry and proteomics. In the following, we use the term "monomer" to refer to a reactant and "subunit" to refer to the corresponding molecular entity upon incorporation into the reaction product.

We rely on the term "condensed formula" to categorize subgroups of molecules. The definition is perhaps best illustrated with an example: peptides composed of a given set of amino acids (combination) but in different sequences (permutation) are categorized in one group. Each peptide has the same nominal mass and would give a single peak in mass spectrometry. For example, a linear tripeptide composed of two alanine (A) and one glycine (G) residues can be formed in three sequences: G-A-A, A-A-G, and A-G-A. The three peptides are distinct in sequence and hence not identical yet are isomeric given the identical condensed formula of A_2G. Reliance on such condensed formulas retains valuable chemical information that can be lost with molecular formulas owing to coincidental mass degeneracies, which are frequent upon use of modular subunits of low molecular weight and limited atomic complexity (e.g., composed solely of C, H, N, and O). Coincidental mass degeneracies can occur in at least two ways. (1) Two modular subunits are isomeric: for example, leucine (L) and isoleucine (I) give the same molecular formula, but GGAI and GALG have distinct condensed formulas (G_2AI and G_2AL). (2) A combination of modular subunits has the same molecular formula: for example, the molecular formula of (glycine + glutamic acid) is identical to that of (alanine + aspartic acid).

Enumeration of condensed formulas is valuable for determining the predicted theoretical mass distribution in combinatorial chemistry. When no coincidental mass degeneracy arises, the number of condensed formulas equals the number of molecular formulas. Mass degeneracy will decrease the number of molecular formulas; accordingly, the number of condensed formulas provides an upper bound on the number of mass peaks.

Categorization of isomers is essential; however, to date, a settled terminology has not emerged. Instead, a wide variety of terminologies are used such as "maximum number of non-isobaric compounds,"[65] "isobaric species,"[66] "multiplicity,"[67] "number of different amino acid compositions,"[67] "number of compositions,"[68] and "compositionally distinct peptides."[69] The term "condensed formula" is traditionally used to convey a structure without showing bonds (e.g., CH_2CH_2 stands for

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Number of Condensed Formulas As a Function of Architecture Size and Number of Distinct Monomers

	no. of monomers k											
size n	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	1	3	6	10	15	21	28	36	45	55	66	78
3	1	4	10	20	35	56	84	120	165	220	286	364
4	1	5	15	35	70	126	210	330	495	715	1001	1365
5	1	6	21	56	126	252	462	792	1287	2002	3003	4368
6	1	7	28	84	210	462	924	1716	3003	5005	8008	12376
7	1	8	36	120	330	792	1716	3432	6435	11440	19448	31824
8	1	9	45	165	495	1287	3003	6435	12870	24310	43758	75582
9	1	10	55	220	715	2002	5005	11440	24310	48620	92378	167960
10	1	11	66	286	1001	3003	8008	19448	43758	92378	184756	352716
11	1	12	78	364	1365	4368	12376	31824	75582	167960	352716	705432
12	1	13	91	455	1820	6188	18564	50388	125970	293930	646646	1352078

$CH_2{=}CH_2$, ethene).[70] We employed the term "condensed formula of substituents" to describe the collection of substituents appended to the perimeter of a common macrocycle skeleton,[12,13] an approach akin to denoting the amino acid side chains of a peptide. Such condensed formulas summarize the collection of relevant groups that distinguish molecular entities but are devoid of information concerning molecular connectivity. Thus, the condensed formula "A_2G" describes the composition but not the three possible peptide sequences.

The condensed formula can be calculated regardless of the position of the modular subunits. Thus, enumeration of the condensed formulas equates to enumeration of the various combinations with no need to consider any order (permutations) of the modular subunits. In other words, the number of condensed formulas is independent of the architecture (regardless of linear or cyclic, inherent or no directionality). The total number of condensed formulas M is given by eq 1, where k is the number of distinct monomers and n is the number of modular subunits that comprises the linear or cyclic architecture.

$$M = \frac{(k + n - 1)!}{(k - 1)!n!} \tag{1}$$

Hereafter, we simply refer to n as the size of the modular architecture and k as the number of monomers. Equation 1 thus describes the numbers of ways to choose architectures of "n" subunits from a set of "k" monomers. Monomers can be chosen more than once (i.e., repetition is allowed), which is known as counting multisets.[67−69,71,72] Equation 1 is contained in the mathematical module of *Cyclaplex* and was used to give the results shown in Table 1. The number of condensed formulas (M) is listed, where the number of distinct monomers encompasses k = 1−12, and the size of the architecture encompasses n = 1−12. The number of products (as opposed to condensed formulas) is provided by mathematical enumeration as described in the next section.

III. Methods of Enumeration. *A. Mathematical Description.* The program *Cyclaplex* contains a module for mathematical enumeration of each of the four types of molecular architectures. The established mathematical approaches are outlined as follows.

(1) Linear Architectures with Inherent Directionality (Cords). The enumeration of linear molecules with inherent directionality is simple. The total number of molecules (Z) depends on the length of the architecture (n subunits) and the number of distinct monomers (k) to draw from for incorporation at each position, as given in eq 2. A classic molecular example is provided by peptides, which have an amino terminus and a carboxylic acid terminus and hence are inherently directional.

$$Z = k^n \tag{2}$$

(2) Linear Architectures without Inherent Directionality (Batons). The enumeration of linear architectures without directionality is more complicated than that for molecules with inherent directionality because symmetry must be taken into consideration. Such enumeration is often described[73,74] as the "k-coloring problem of a baton (strip of cloth or rod) divided into n-cylindrical bands of equal length." An alternative expression of the underlying symmetry ramification is as follows: two overall colorings count as the same if one of them can be converted into the other by 180° rotation. Such problems can be solved by application of Pólya's theorem.[75] Pólya's theorem takes into account the symmetry and permissible motions of an object composed of modular subunits to derive a corresponding "cycle index equation." Application of a cycle index equation for a given number of monomers yields the "pattern inventory," which describes all resulting condensed formulas, and the number of members (i.e., products) that have a given condensed formula.

By way of illustration, the cycle index equations are given in eq 3 for odd-membered batons (n is odd) and in eq 4 for even-membered batons. A more full development of the cycle index equations for batons is provided in Table S1 of the Supporting Information.

$$Z = 1/2({m_1}^n + m_1{m_2}^{(n-1)/2}) \tag{3}$$

$$Z = 1/2({m_1}^n + {m_2}^{n/2}) \tag{4}$$

The cycle index equations expand for a given number of monomers to obtain the pattern inventory. Thus, the expression for m_1 and m_2 in the cycle index equations is given by eq 5 where A, B, C, D, ..., X are the distinct monomers.

$$m_j = A^j + B^j + C^j + D^j + \ldots + X^j \tag{5}$$

For example, with two monomers (A, B), m_1 = A + B and $m_2 = A^2 + B^2$. For a baton size of n = 5, the equation given in Table 2 affords eq 6.

$$\begin{aligned} Z &= 1/2({m_1}^5 + m_1{m_2}^2) \\ &= 1/2[(A + B)^5 + (A + B)(A^2 + B^2)^2] \end{aligned} \tag{6}$$

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Cyclic Index Equations for Batons

size n	$Z(A)$
3	$1/2(m_1^3 + m_1m_2)$
4	$1/2(m_1^4 + m_2^2)$
5	$1/2(m_1^5 + m_1m_2^2)$
6	$1/2(m_1^6 + m_2^3)$
7	$1/2(m_1^7 + m_1m_2^3)$
8	$1/2(m_1^8 + m_2^4)$
9	$1/2(m_1^9 + m_1m_2^4)$
10	$1/2(m_1^{10} + m_2^5)$
11	$1/2(m_1^{11} + m_1m_2^5)$
12	$1/2(m_1^{12} + m_2^6)$

Expansion of eq 6 affords the pattern inventory (eq 7).

$$Z = A^5 + 3A^4B + 6A^3B^2 + 6A^2B^3 + 3AB^4 + B^5 \quad (7)$$

Replacement of superscripts (mathematical terminology) of eq 7 with subscripts (chemical terminology) affords the pattern inventory in a form better understood by chemists (eq 8).[12]

$$Z = A_5 + 3A_4B + 6A_3B_2 + 6A_2B_3 + 3AB_4 + B_5 \quad (8)$$

The pattern inventory (eq 8) reveals three important aspects of the enumeration:

(i) The number of the terms (A_5, A_4B, A_3B_2, A_2B_3, AB_4, B_5) equals the number of condensed formulas; thus, there are six distinct condensed formulas (i.e., six distinct isomeric subgroups).

(ii) The coefficient of each of the terms (1, 3, 6, 6, 3, 1) gives the number of isomers of that particular condensed formula. For example, A_5 and B_5 are unique (no isomers), whereas there are three isomers for each of A_4B and AB_4 and six isomers for each of A_3B_2 and A_2B_3. The members of each isomeric subgroup have a common condensed formula.

(iii) The sum of the coefficients gives the total number of products, which in this case is 1 + 3 + 6 + 6 + 3 + 1 = 20.

In summary, the pattern inventory provides a first-order description of the product distribution. If even less information is required—for example, to determine solely the total number of products with no identity provided—expansion of the polynomial equation is not required. Thus, for a baton size of n = 5, replacement of m_1 and m_2 in eq 6 by m gives eq 9, which can also be obtained by the Cauchy–Frobenius–Burnside theorem.[71,74] For the case of m = 2 monomers (A and B), Z = 20. In other words, there are 20 batons composed of 5 modular subunits when two distinct monomers (A, B) are employed.

$$Z = 1/2(m^5 + m^3) \quad (9)$$

On the other hand, in chemistry, more information is generally desired as seen by the following. For batons, if A and B are red and blue, respectively, there are 20 colored batons: 1 all red, 1 all blue, 3 red with 1 blue stripe (unspecified location), 3 blue with 1 red stripe (unspecified location), 6 with 3 red and 2 blue stripes (unspecified locations), and 6 with 2 red and 3 blue stripes (unspecified locations). With regards to chemistry, the location of the stripes—the molecular connectivity of the modular subunits—is extremely valuable. Deeper information concerning molecular connectivity of modular subunits in each isomeric subgroup (i.e., with a common condensed formula) and the relative amounts of each in a statistical reaction are not

Table 3. Number of Batons of a Given Size As a Function of Number of Distinct Monomers

	no. of monomers k											
size n	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	1	3	6	10	15	21	28	36	45	55	66	78
3	1	6	18	40	75	126	196	288	405	550	726	936
4	1	10	45	136	325	666	1225	2080	3321	5050	7381	10440
5	1	20	135	544	1625	3996	8575	16640	29889	50500	81191	125280
6	1	36	378	2080	7875	23436	58996	131328	266085	500500	886446	1493856
7	1	72	1134	8320	39375	140616	412972	1050624	2394765	5005000	9750906	17926272
8	1	136	3321	32896	195625	840456	2883601	8390656	21526641	50005000	107186761	215001216
9	1	272	9963	131584	978125	5042736	20185207	67125248	193739769	500050000	1179054371	2580014592
10	1	528	29646	524800	4884375	30236976	141246028	536887296	1743421725	5000050000	12968792826	30958806528
11	1	1056	88938	2099200	24421875	181421856	988722196	4295098368	15690795525	50000500000	142656721086	371505678336
12	1	2080	266085	8390656	122078125	1088414496	6920702425	34359869440	141215033961	500000500000	1569215074141	4458051717120

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

available from the pattern inventory but are obtained by the generative algorithms (vide infra).

Nonetheless, the first-order description of the product distribution as obtained via the approach of Pólya is valuable. The cycle index equations $[Z(A)]$ as described in eqs 3 and 4 for batons of size n = 3−12 are provided in explicit form in Table 2. The equations in Table 2 are contained in the mathematical module of *Cyclaplex* and were used to give the results shown in Table 3. The results concern the number of batons containing n = 1−12 modular subunits and derived from k = 1−12 monomers. We now turn to application of the Pólya approach to cyclic architectures with directionality (necklaces) or without directionality (bracelets).

(3) Cyclic Architectures with Inherent Directionality (Necklaces). The analysis of necklaces is similar to that of batons, although rotation of the necklaces by 360°/n, where n is the ring size, needs to be considered instead of 180° rotation (Figure 2). Necklaces by definition cannot be flipped (i.e., only one face is presented to the viewer). A trivial case for a necklace occurs where rotation also is not possible. In this case, the number of necklaces is given by eq 2, $Z = k^n$. In general, the cycle index equation of C_n is represented by eq 10, where $\phi(d)$ is Euler's totient function.[76] Thus, d takes on all integer divisors of n; for example, if n = 9, and d = 1, 3, and 9.

$$Z(C_n) = \frac{1}{n}\sum_{d|n} \phi(d) x_d^{n/d} \tag{10}$$

The cycle index equations for the enumeration of necklaces containing n = 3−12 modular subunits and derived from k monomers are listed in Table 4. The expression for each m_j in the cycle index equations is given in eq 5 above. A more full development of the cycle index equations for necklaces is provided in Supporting Information Table S2.

Table 4. Cyclic Index Equations for Necklaces

size n	$Z(C_n)$
3	$1/3(m_1^3 + 2m_3)$
4	$1/4(m_1^4 + m_2^2 + 2m_4)$
5	$1/5(m_1^5 + 4m_5)$
6	$1/6(m_1^6 + m_2^3 + 2m_3^2 + 2m_6)$
7	$1/7(m_1^7 + 6m_7)$
8	$1/8(m_1^8 + m_2^4 + 2m_4^2 + 4m_8)$
9	$1/9(m_1^9 + 2m_3^3 + 6m_9)$
10	$1/10(m_1^{10} + m_2^5 + 4m_5^2 + 4m_{10})$
11	$1/11(m_1^{11} + 10m_{11})$
12	$1/12(m_1^{12} + m_2^6 + 2m_3^4 + 2m_4^3 + 2m_6^2 + 4m_{12})$

The equations shown in Table 4 are contained in the mathematical module of *Cyclaplex* and were used to give the results shown in Table 5. The results concern the number of necklaces containing n = 1−12 modular subunits and derived from k = 1−12 monomers.

(4) Cyclic Architectures without Inherent Directionality (Bracelets). The enumeration of bracelets is similar to that of necklaces, although flip motions also need to be considered (Figure 2). Thus, enumeration of bracelets requires consideration of dihedral symmetry of D_n. These problems appear in the enumeration of regular polygons and are described[77−79] by the "k-coloring problem of vertices of n-gons." The enumeration of regular polygons has been extensively discussed,[77−86] yet for our purposes, the presentations are neither comprehensive nor suitably tailored for a chemistry perspective.

Table 5. Number of Necklaces of a Given Size As a Function of Number of Distinct Monomers

	no. of monomers k											
size n	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	1	3	6	10	15	21	28	36	45	55	66	78
3	1	4	11	24	45	76	119	176	249	340	451	584
4	1	6	24	70	165	336	616	1044	1665	2530	3696	5226
5	1	8	51	208	629	1560	3367	6560	11817	20008	32219	49776
6	1	14	130	700	2635	7826	19684	43800	88725	166870	295526	498004
7	1	20	315	2344	11165	39996	117655	299600	683289	1428580	2783891	5118840
8	1	36	834	8230	48915	210126	720916	2097684	5381685	12501280	26796726	53750346
9	1	60	2195	29144	217045	1119796	4483815	14913200	43046889	111111340	261994491	573309320
10	1	108	5934	104968	976887	6047412	28249228	107377488	348684381	1000010044	2593758618	6191761368
11	1	188	16107	381304	4438925	32981556	179756983	780903152	2852823609	9090909100	25937424611	67546215528
12	1	352	44368	1398500	20346485	181402676	1153450872	5726645688	23535840225	83333418520	261535848376	743008623292

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

In general, the cycle index equation for odd-membered rings is represented by eq 11, where again $\phi(d)$ is Euler's totient function. The cycle index equation for even-membered rings is represented by eq 12. The cycle index equations for the enumeration of bracelets of size n = 3−12 from k monomers are listed in Table 6. Again, the expression for each m_j is given in eq 5 above. A more full development of the cycle index equations is provided in Supporting Information Table S3.

Table 6. Cyclic Index Equations for Bracelets

size n	$Z(D_n)$
3	$1/6(m_1^3 + 3m_1m_2 + 2m_3)$
4	$1/8(m_1^4 + 2m_1^2m_2 + 3m_2^2 + 2m_4)$
5	$1/10(m_1^5 + 5m_1m_2^2 + 4m_5)$
6	$1/12(m_1^6 + 3m_1^2m_2^2 + 4m_2^3 + 2m_3^2 + 2m_6)$
7	$1/14(m_1^7 + 7m_1m_2^3 + 6m_7)$
8	$1/16(m_1^8 + 4m_1^2m_2^3 + 5m_2^4 + 2m_4^2 + 4m_8)$
9	$1/18(m_1^9 + 9m_1m_2^4 + 2m_3^3 + 6m_9)$
10	$1/20(m_1^{10} + 5m_1^2m_2^4 + 6m_2^5 + 4m_5^2 + 4m_{10})$
11	$1/22(m_1^{11} + 11m_1m_2^5 + 10m_{11})$
12	$1/24(m_1^{12} + 6m_2^5m_1^2 + 7m_2^6 + 2m_3^4 + 2m_4^3 + 2m_6^2 + 4m_{12})$

$$Z(D_n) = \frac{1}{2}Z(C_n) + \frac{1}{2}x_1x_2^{(n-1)/2} = \frac{1}{2n}\left(\sum_{d|n}\phi(d)x_d^{n/d} + nx_1x_2^{(n-1)/2}\right) \quad (11)$$

$$Z(D_n) = \frac{1}{2}Z(C_n) + \frac{1}{4}x_2^{n/2} + \frac{1}{4}x_1^2x_2^{n/2-1} = \frac{1}{2n}\left(\sum_{d|n}\phi(d)x_d^{n/d} + \frac{n}{2}x_2^{n/2} + \frac{n}{2}x_1^2x_2^{n/2-1}\right) \quad (12)$$

The equations shown in Table 6 are contained in the mathematical module of *Cyclaplex* and were used to give the results shown in Table 7. The results concern the number of bracelets containing n = 1−12 modular subunits and derived from k = 1−12 monomers. We now turn to compare the numbers of products of a given architectural type (cords, batons, necklaces, bracelets) that are formed from a given number of monomers.

Example. Trimers in Four Architectures. As trimers are the smallest cycle that can be formed, and the number of products in each architecture is small enough to be manageable, it is instructive to begin with examination of the various trimers. For use of three monomers, all products are illustrated in Figure 3. The number of linear trimers with directionality is simply 3^3 = 27, whereas in the absence of directionality, the number declines to 18. For cyclic trimers with directionality, there are 11 products but only 10 products for cyclic trimers without directionality. This last distinction stems from the fact that there are two distinct tricolored products for trimers with directionality (bottom row, left panel, Figure 3) whereas there is only one such a tricolored product for trimers without directionality (bottom row, right panel, Figure 3). In other words, the two tricolored products for trimers with directionality are not interconvertible via C_2 rotation. Regardless of architecture, the number of condensed formulas is identical, namely 10. The order of size of the product distribution as seen here (linear with directionality > linear without directionality > cyclic with

Table 7. Number of Bracelets of a Given Size As a Function of Number of Distinct Monomers

	no. of monomers k											
size n	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	1	3	6	10	15	21	28	36	45	55	66	78
3	1	4	10	20	35	56	84	120	165	220	286	364
4	1	6	21	55	120	231	406	666	1035	1540	2211	3081
5	1	8	39	136	377	888	1855	3536	6273	10504	16775	25752
6	1	13	92	430	1505	4291	10528	23052	46185	86185	151756	254618
7	1	18	198	1300	5895	20646	60028	151848	344925	719290	1399266	2569788
8	1	30	498	4435	25395	107331	365260	1058058	2707245	6278140	13442286	26942565
9	1	46	1219	15084	110085	563786	2250311	7472984	21552969	55605670	131077771	286779076
10	1	78	3210	53764	493131	3037314	14158228	53762472	174489813	500280022	1297362462	3096689388
11	1	126	8418	192700	2227275	16514106	89937316	390582648	1426677525	4545954550	12969598086	33774600756
12	1	224	22913	704370	10196680	90782986	576960734	2863912668	11769248715	41669459260	130773238871	371514016094

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

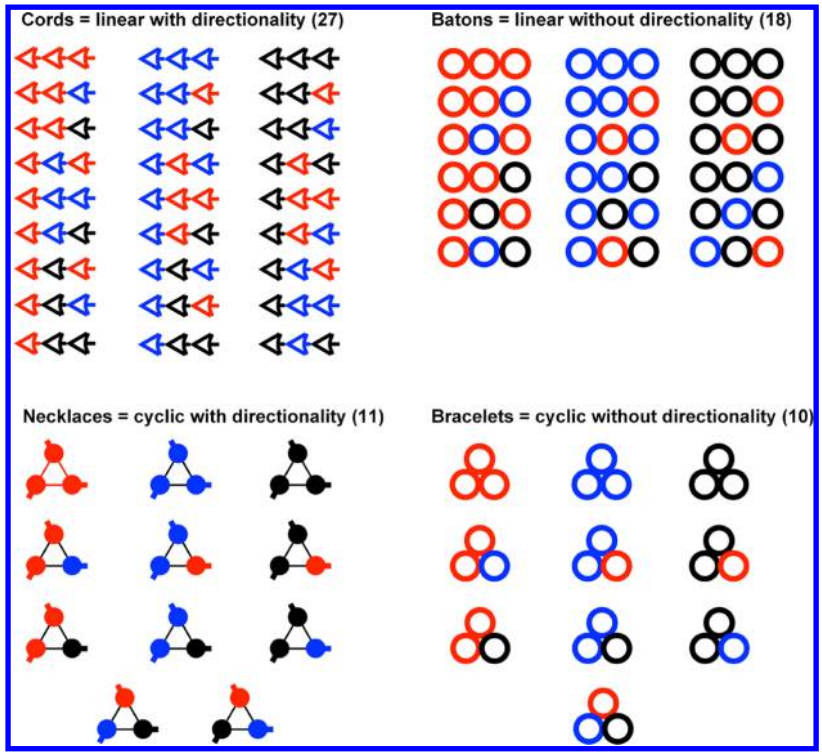


Figure 3. Trimers derived from three monomers.

directionality > cyclic without directionality) appears to be rather general, as described next.

With the expressions for enumeration in hand (eqs 2−4 and 10−12; Tables 2, 4, and 6), a series of calculations was carried out using *Cyclaplex* to compare the number of products that occur for typical numbers of monomers in the four architectures. The results are shown in Table 8 for architectures with n = 3−12 modular subunits derived from k = 2, 4, or 12 monomers. For example, the use of two monomers (A, B) to give hexamers affords 64 linear molecules with directionality but only 36 without directionality. The increased symmetry of the latter gives a diminished number versus the former. The formation of a cyclic architecture gives a further decrease in the number of products: due to rotation there are only 14 products (with directionality), and when flipping can occur, there are only 13 products (no directionality). Regardless of molecular architecture, for this case there are only seven condensed formulas: A_6, A_5B_1, A_4B_2, A_3B_3, A_2B_4, A_1B_5, and B_6. The occurrence of seven condensed formulas has two implications: (1) The maximum number of mass spectral peaks thus is seven; any coincidental mass degeneracies would result in fewer than seven peaks. (2) An architecture for which there are greater than seven products implies the existence of isomers. For example, of the 64 linear products with directionality, A_6, A_5B_1, A_1B_5, and B_6 each represent one respective product, whereas the condensed formulas of A_4B_2, A_3B_3, and A_2B_4 each contain multiple (isomeric) members. Thus, of the 64 total products, 4 are unique (A_6, A_5B_1, A_1B_5, and B_6) and 60 are isomeric with at least one other product in the set.

The number of products increases profoundly with an increase in the number of monomers. Doubling the number of monomers to 4 increases the number of hexamers to 4096, 2080, 700, and 430 for the 4 types of molecular architectures. The general trends in the number of condensed formulas, and the number of products with increasing size are shown in Figure 4 for the case of k = 12 monomers. The number of six-membered necklaces exceeds 10^6 in this case. The number of condensed formulas is shown (line with × symbols) in Figure 4. For a given architectural size, all products above the line are isomeric with at least one other product of the same size.

In summary, mathematical descriptions have been developed previously for each of the four types of molecular architectures of interest here. This approach gives rise to the pattern inventory; i.e., the number of products of a given condensed formula. The pattern inventory for the tripeptides derived by combination of A and G would be $A_3 + 3A_2G + 3AG_2 + G_3$. Yet, three key pieces of information are missing from a chemical standpoint: (i) structural information (e.g., that AGG, AGA, and GAA are examples of condensed formula A_2G); (ii) relative amounts of all eight products; and (iii) effects on the distribution depending on nonequal ratios of the A and G monomers. Accordingly, we have developed generative algorithms to create a virtual library of the products. The two approaches are complementary. The availability of mathematical equations (i.e., methods for counting)[63] enables rapid calculation to assess the number of products; the generative algorithms (i.e., construction methods)[63] are computationally slower but afford a richer chemical assessment. In addition, the virtual library derived from the generative algorithms can be data-mined to identify subsets of products with desired properties. The generative algorithms are described in the next section.

B. Generative Algorithms. The program *Cyclaplex* includes a module for generating the virtual library of products for each of the four architectural types (cords, batons, necklaces, bracelets). The module does not rely on mathematical equations but instead employs algorithms to generate all combinations and permutations of monomers for a given reaction; then, the frequency of each product containing the corresponding modular subunits is counted. In other words, each such constructed entity can be regarded as a labeled graph; all labeled

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

graphs that are isomorphic are regarded as unlabeled. In chemistry, molecules are considered as unlabeled graphs.[63] Thus, an accurate assessment of the library composition requires the ability to construct all possible molecules in all possible ways (i.e., generate all labeled graphs) and count the number of ways each distinct product can be formed (i.e., determine the number of each isomorphic type). In so doing, this approach complements the more limited description of mere groups of isomers that emerge from the mathematically derived pattern inventory (vide supra). In practice, the relative amount of each product depends not only on such statistics of formation but also on the ratio of respective monomers, all of which are accommodated by the generative algorithms. The program resembles *PorphyrinViLiGe*[12] in concept and in many aspects of software design but differs fundamentally in scientific application. Hence, we provide only an outline of the program workflow and then illustrate the program with selected applications.

The workflow of the generative module of *Cyclaplex* consists of the following stages i–v:

(i) Provide input including identification of distinct monomers (i.e., members of the reactant chemset[87]), the ratio of monomers (equimolar is the default), and tag the monomers if desired for subsequent data mining.

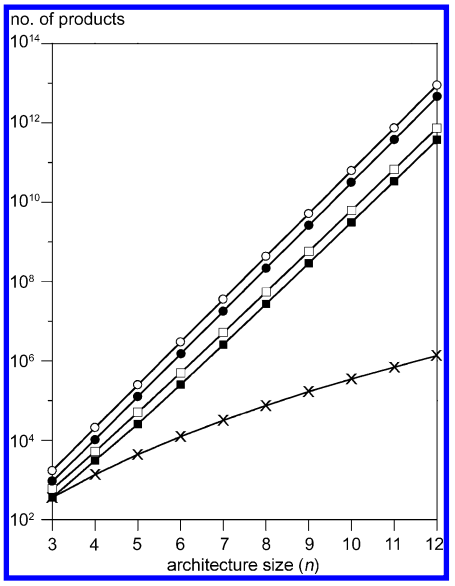


Figure 4. Number of products as a function of architecture size (n) derived from k = 12 monomers: (legend) open circles, cords; solid circles, batons; open squares, necklaces; solid squares, bracelets; × marks, number of condensed formulas.

Table 8. Number of Condensed Formulas and Architectures for k = 2, 4, and 12

size n	number of condensed formulas	number of architectures			
		cords	batons	necklaces	bracelets
		k = 2 monomers			
3	4	8	6	4	4
4	5	16	10	6	6
5	6	32	20	8	8
6	7	64	36	14	13
7	8	128	72	20	18
8	9	256	136	36	30
9	10	512	272	60	46
10	11	1024	528	108	78
11	12	2048	1056	188	126
12	13	4096	2080	352	224
		k = 4 monomers			
3	20	64	40	24	20
4	35	256	136	70	55
5	56	1024	544	208	136
6	84	4096	2080	700	430
7	120	16384	8320	2344	1300
8	165	65536	32896	8230	4435
9	220	262144	135184	29144	15084
10	286	1048576	524800	104968	53764
11	364	4194304	2099200	381304	192700
12	455	16777216	8390656	1398500	704370
		k = 12 monomers			
3	364	1728	936	584	364
4	1365	20736	10440	5226	3081
5	4368	248832	125280	49776	25752
6	12376	2985984	1493856	498004	254618
7	31824	35831808	17926272	5118840	2569788
8	75582	429981696	215001216	53750346	26942565
9	167960	5159780352	2580014592	573309320	286779076
10	352716	61917364224	30958806528	6191761368	3096689388
11	705432	743008370688	371505678336	67546215528	33774600756
12	1352078	8916100448256	4458051717120	743008623292	371514016094

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

(ii) Generate all product combinations and permutations, which are stored in several intermediate data sets.
(iii) Select one of the four architectural types and prune the intermediate data sets accordingly to obtain the final subset of products.
(iv) Implement isomer analysis to identify distinct products, count the frequency of each product, and adjust the frequencies on the basis of the ratio of initial monomers.
(v) Provide output consisting of (a) composition of all products; (b) their relative amounts; and (c) the number of condensed formulas.

The utility of the generative algorithms—and the distinctive features compared with the mathematical approach—is best conveyed by example. The following first considers the formation of heterogeneous (natural or biohybrid) light-harvesting architectures of various sizes. Then, the effects of nonequal ratios of reacting monomers on the composition of cyclic octamers are examined. Finally, the availability of information concerning each species in a virtual library enables data mining to be performed. As one example, the virtual library of linear and cyclic octamers was examined to assess the occurrence of particular tracts of subunits.

(1) Heterogeneous Light-Harvesting Architectures. The light-harvesting antenna complexes of photosynthetic bacteria consist of self-assembled oligomers of peptide−pigment subunits,[52] as mentioned in the introduction. A fertile area in the field of artificial photosynthesis is to exploit the constituents of the natural systems as chassis for light-harvesting studies. In this manner, biohybrid architectures can be created by synthetic elaboration of the natural peptides including attachment of non-natural chromophores[61,62] and introduction of provisions for surface attachment.[88]

The natural systems were initially thought to employ a single subunit or monomer type, in which case each oligomer derived therefrom would be compositionally homogeneous. More recently, the composition of light-harvesting (LH2) complexes has been found to vary with light intensity, which is attributed to the presence of different peptides that compose the constituent subunits.[60] Whether the light-harvesting system is composed of a mixture of oligomers each of which alone is homogeneous (i.e., composed of a given subunit type), or a mixture of heterogeneous oligomers (i.e., with different subunits present in a given oligomer), remains under investigation.[60] With synthetic access to modified subunits, controlled preparation of heterogeneous mixtures of oligomers also becomes possible. The question arose concerning how many cyclic nonamers are to be expected if k distinct monomers are employed? This answer is readily available.

Consider the cyclic light-harvesting oligomers to be examples of necklaces or cords. Note that deformation from a circular architecture (e.g., to give an oval or ellipse) or a linear architecture (e.g., to give an open-chain arc or curvilinear string) does not alter the topology and hence does not alter the calculation given that necklaces and cords have intrinsic directionality (i.e., flip motions cannot occur). Thus, for the cyclic nonamer (n = 9) assembled from k = 2 monomers, there are 60 products. In contrast, the linear nonamer is present as k^n = 512 products (Table 9). Upon doubling the number of monomers (k = 4), the number of products increases profoundly: there are 29 144 cyclic nonamers and 262 144 linear nonamers. Such results are available from the mathematical formulas or from the generative algorithm.

Table 9. Heterogeneity in Prototypical Light-Harvesting Architectures

no. of monomers	architecture	no. of octamers (n = 8)	no. of nonamers (n = 9)
k = 2	cords	256	512
	necklaces	36	60
	condensed formulas	9	10
k = 3	cords	6561	19683
	necklaces	834	2195
	condensed formulas	45	55
k = 4	cords	65536	262144
	necklaces	8230	29144
	condensed formulas	165	220

The natural light-harvesting complex LH1 is larger than LH2, with typically 14−16 subunits (i.e., n = 14−16) and is believed to be more pliable in terms of conformation and more fluid with regard to the presence of linear and cyclic forms.[51] Indeed, LH1 may adopt an elliptical rather than circular architecture,[89] which for computational purposes is treated as a necklace (where rotation is possible but flip motions are not permitted). The circular or elliptical architecture also may contain a gap,[89] in which case the architecture is formally linear and is necessarily treated as a cord. The number of products formed upon use of k = 2 monomers in oligomers of size n = 13−17 is provided in Table 10 for cords and necklaces (as well as batons and bracelets). For hexadecameric necklaces, there are 4116 products. More elaborate molecular assemblies are believed to form in vivo,[51] in which case the calculations here delineate the possible compositional richness of limiting forms.

Table 10. Number of Products and Computational Time for Larger Architectures[a]

	architecture			
size n	cords	batons	necklaces	bracelets
13	8192	4160	632	380
	<1 s	<1 s	∼4 s	∼2 s
14	16384	8256	1182	687
	<1 s	<1 s	19 s	11 s
15	32768	16512	2192	1224
	2 s	1 s	60 s	33 s
16	65536	32896	4116	2250
	6 s	3 s	300 s	165 s
17	131072	65792	7712	4112
	16 s	7 s	1000 s	518 s

[a]Size n = 13−17 and number of monomers k = 2. Calculations were performed on a PC equipped with a 3.0 GHz dual core CPU (Core 2 DUO, E8400) and 2 GB of RAM.

It is of interest to note the computational times required for use of the generative algorithms, which are longer than the mathematical approach. The computational time is far greater for cyclic versus linear architectures (Table 10). However, within the linear or cyclic architectures, the computational time decreases with increasing symmetry. Thus, the virtual library is composed most rapidly for batons (versus cords) and most slowly for necklaces (versus bracelets).

(2) Nonequal Ratios of Monomers. The ability to employ nonequal ratios of monomers is a distinctive feature of the generative algorithms. For example, the results for cyclic octamers (necklaces or bracelets) composed of two types of monomers (A, B) in the virtual library in 1:1 or 2:1 ratios are

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 11. Virtual Library of Cyclic Octamers (Necklace, Bracelet) Composed of A and B Monomers

Necklaces (36 Products Total)					
		A:B = 1:1		A:B = 2:1	
condensed formulas	products	frequency	yield (%)	frequency	yield (%)
A_8	A-A-A-A-A-A-A-A-	1	0.391	256	3.902
A_7B_1	A-A-A-A-A-A-A-B-	8	3.125	1024	15.607
A_6B_2	A-A-A-A-A-A-B-B-	8	3.125	512	7.804
	A-A-A-A-A-B-A-B-	8	3.125	512	7.804
	A-A-A-A-B-A-A-B-	8	3.125	512	7.804
	A-A-A-B-A-A-A-B-	4	1.563	256	3.902
A_5B_3	A-A-A-A-A-B-B-B-	8	3.125	256	3.902
	A-A-A-A-B-A-B-B-	8	3.125	256	3.902
	A-A-A-A-B-B-A-B-	8	3.125	256	3.902
	A-A-A-B-A-A-B-B-	8	3.125	256	3.902
	A-A-A-B-A-B-A-B-	8	3.125	256	3.902
	A-A-A-B-B-A-A-B-	8	3.125	256	3.902
	A-A-B-A-A-B-A-B-	8	3.125	256	3.902
A_4B_4	A-A-A-A-B-B-B-B-	8	3.125	128	1.951
	A-A-A-B-A-B-B-B-	8	3.125	128	1.951
	A-A-A-B-B-A-B-B-	8	3.125	128	1.951
	A-A-A-B-B-B-A-B-	8	3.125	128	1.951
	A-A-B-A-A-B-B-B-	8	3.125	128	1.951
	A-A-B-A-B-A-B-B-	8	3.125	128	1.951
	A-A-B-A-B-B-A-B-	8	3.125	128	1.951
	A-A-B-B-A-A-B-B-	4	1.563	64	0.975
	A-A-B-B-A-B-A-B-	8	3.125	128	1.951
	A-B-A-B-A-B-A-B-	2	0.781	32	0.488
A_3B_5	B-B-B-B-B-A-A-A-	8	3.125	64	0.975
	B-B-B-B-A-B-A-A-	8	3.125	64	0.975
	B-B-B-B-A-A-B-A-	8	3.125	64	0.975
	B-B-B-A-B-B-A-A-	8	3.125	64	0.975
	B-B-B-A-B-A-B-A-	8	3.125	64	0.975
	B-B-B-A-A-B-B-A-	8	3.125	64	0.975
	B-B-A-B-B-A-B-A-	8	3.125	64	0.975
A_2B_6	B-B-B-B-B-B-A-A-	8	3.125	32	0.488
	B-B-B-B-B-A-B-A-	8	3.125	32	0.488
	B-B-B-B-A-B-B-A-	8	3.125	32	0.488
	B-B-B-A-B-B-B-A-	4	1.563	16	0.244
A_1B_7	B-B-B-B-B-B-B-A-	8	3.125	16	0.244
B_8	B-B-B-B-B-B-B-B-	1	0.391	1	0.015
		256	100	6561	100

Bracelets (30 Products Total)					
		A:B = 1:1		A:B = 2:1	
condensed formulas	products	frequency	yield (%)	frequency	yield (%)
A_8	A-A-A-A-A-A-A-A-	1	0.391	256	3.902
A_7B_1	A-A-A-A-A-A-A-B-	8	3.125	1024	15.607
A_6B_2	A-A-A-A-A-A-B-B-	8	3.125	512	7.804
	A-A-A-A-A-B-A-B-	8	3.125	512	7.804
	A-A-A-A-B-A-A-B-	8	3.125	512	7.804
	A-A-A-B-A-A-A-B-	4	1.563	256	3.902
A_5B_3	A-A-A-A-A-B-B-B-	8	3.125	256	3.902
	A-A-A-A-B-A-B-B-	16	6.25	512	7.804
	A-A-A-B-A-A-B-B-	16	6.25	512	7.804
	A-A-A-B-A-B-A-B-	8	3.125	256	3.902
	A-A-B-A-A-B-A-B-	8	3.125	256	3.902
A_4B_4	A-A-A-A-B-B-B-B-	8	3.125	128	1.951
	A-A-A-B-A-B-B-B-	16	6.25	256	3.902
	A-A-A-B-B-A-B-B-	8	3.125	128	1.951
	A-A-B-A-A-B-B-B-	8	3.125	128	1.951
	A-A-B-A-B-A-B-B-	16	6.25	256	3.902
	A-A-B-A-B-B-A-B-	8	3.125	128	1.951
	A-A-B-B-A-A-B-B-	4	1.563	64	0.976
	A-B-A-B-A-B-A-B-	2	0.781	32	0.488
A_3B_5	B-B-B-B-B-A-A-A-	8	3.125	64	0.976
	B-B-B-B-A-B-A-A-	16	6.25	128	1.951
	B-B-B-A-B-B-A-A-	16	6.25	128	1.951
	B-B-B-A-B-A-B-A-	8	3.125	64	0.976
	B-B-A-B-B-A-B-A-	8	3.125	64	0.976
A_2B_6	B-B-B-B-B-B-A-A-	8	3.125	32	0.488
	B-B-B-B-B-A-B-A-	8	3.125	32	0.488
	B-B-B-B-A-B-B-A-	8	3.125	32	0.488
	B-B-B-A-B-B-B-A-	4	1.563	16	0.244
A_1B_7	B-B-B-B-B-B-B-A-	8	3.125	16	0.244
B_8	B-B-B-B-B-B-B-B-	1	0.391	1	0.015
		256	100	6561	100

listed in Table 11. When the ratio of monomers A:B = 2:1, the yield of the cyclic A_8-octamer products (i.e., A-A-A-A-A-A-A-A-) is 3.902%, which is ~10 times higher than use of a 1:1 ratio of A and B (0.391%). This result arises for both the necklace and bracelet architectures.

(3) Data Mining. The mathematical approach provides the number of products and the number of condensed formulas, but not the specific pattern of subunits in each product (vide supra). Inspection of Table 11 shows the exact composition and subunit pattern of each species present in the library. For example, there are nine isomeric subgroups (on the basis of condensed formulas) wherein each isomer contains the same number of each subunit but the isomers in a set differ in the pattern of subunits.

The availability of a complete virtual library enables data mining to glean information concerning the specific composition of the various products. Examples of data mining (for octamers composed of two types of subunits A and B) are provided as follows.

(a) How many linear octamers have no A subunits at the termini (i.e., both termini are B)? Such a question can be readily answered by the data-mining function. In total, 64 products (out of 256 possible products) have no A subunits at the termini in octameric cords, while 36 products (out of 136 possible products) have no A subunits at the termini in octameric batons (library not shown).

(b) How many linear and cyclic octamers have 2, 3, 4, 5, 6, or 7 adjacent A subunits? The results of data mining are summarized in Table 12. The counting is done first for an exact number of contiguous subunits. The case of oligomers formed by light-harvesting peptides serves for the basis of valuable comparisons. For example, if hybrid light-harvesting oligomers are prepared by an admixture of an artificial peptide (A) and a natural peptide (B), a relevant question is what fraction of the total products would contain two adjacent artificial peptides (i.e., a sequence of "A-A-"). The question is germane because of possible interactions (beneficial or deleterious) between chromophores at adjacent sites. Here the architectures are cords and necklaces.

dx.doi.org/10.1021/ci400175f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 12. Data Mining of the (A, B) Octamer Library for Products with a Single Tract of Contiguous A Subunits[a]

no. of contiguous A subunits	no. of products	theoretical yield (%)	no. of contiguous A subunits	no. of products	theoretical yield (%)
cords (256 products total)					
7	2	0.78	≥7	3	1.12
6	5	1.95	≥6	8	3.13
5	12	4.69	≥5	20	7.81
4	28	10.94	≥4	48	18.75
3	61	23.83	≥3	107	41.80
2	120	46.88	≥2	201	78.52
batons (136 products total)					
7	1	0.78	≥7	2	1.12
6	3	1.95	≥6	5	3.13
5	6	4.69	≥5	11	7.81
4	15	10.94	≥4	26	18.75
3	31	23.83	≥3	56	41.80
2	63	46.88	≥2	106	78.52
necklaces (36 products total)					
7	1	3.13	≥7	2	3.52
6	1	3.13	≥6	3	6.64
5	2	6.25	≥5	5	12.89
4	4	12.5	≥4	9	25.39
3	8	23.44	≥3	17	48.83
2	14	42.19	≥2	28	81.64
bracelets (30 products total)					
7	1	3.13	≥7	2	3.52
6	1	3.13	≥6	3	6.64
5	2	6.25	≥5	5	12.89
4	3	12.5	≥4	8	25.39
3	6	23.44	≥3	14	48.83
2	10	42.19	≥2	22	81.64

[a]The virtual library is for a 1:1 ratio of A and B monomers; all cyclic species are shown in Table 11.

For cords, there are 120 of 256 products that contain at least one "A-A-" entity. The counting is done such that sequences of >2 contiguous A subunits (e.g., "A-A-A-") are excluded. Thus, a sequence of A-A-B-B-B-A-A-A- is counted once; a sequence of A-A-B-B-B-A-A-B- is counted only once; and a sequence of A-A-A-B-B-B-B- is not counted at all. On the other hand, when the question concerns how many products have a contiguous tract of ≥2 A subunits, then sequences of A-A-B-B-B-A-A-A-, A-A-B-B-B-A-A-B-, and A-A-A-B-B-B-B- are each counted once. Thus, in cords, there are 47% with at least one set of 2 adjacent A subunits and 79% with at least one set of ≥2 adjacent A subunits.

For necklaces, there are 14 of 36 products that contain at least one "A-A-" entity, and 28 of 36 that contain ≥2 adjacent A subunits. In summary, there are 42% with at least one set of exactly two adjacent A subunits and 82% with at least one set of ≥2 adjacent A subunits.

In summary, the percentage (i.e., theoretical yield) of such A-A- containing entities is relatively fixed regardless of linear versus cyclic architecture (although the number is substantially larger in the former versus latter). These insights readily emerge upon data mining of the virtual library formed by the generative algorithms available in *Cyclaplex*.

OUTLOOK

The ability to enumerate macrocyclic architectures (and their linear counterparts) is invaluable for understanding the possible diversity formed upon combinatorial reactions. The generation of a virtual library, and data mining thereof, enables consideration of the properties of a chemical sample in terms of the underlying molecular diversity. For example, prior modeling of the light-harvesting performance of modular architectures relied on homogeneous samples and examined linear, branched, or cyclic architectures;[90] the results described herein should enable analogous calculations of heterogeneous samples, including those containing distributions of compositionally rich molecular architectures. A chief challenge to such modeling may be the underdetermination of the observables with respect to the distinct components of the sample. Regardless, predictions of the composition of such samples, particularly those containing macrocycles, was previously quite difficult. Macrocycles populate chemical space in myriad forms yet the inherent rotational and reflection symmetries typically place enumeration of these important molecules—at least for any reasonable size and composition—outside the bounds of intuition or simple calculation. For architectures that resemble cords, batons, necklaces, or bracelets, the program *Cyclaplex* fills this lacuna and can be freely downloaded from the Web site www.photochemcad.com.

ASSOCIATED CONTENT

Supporting Information

Overview of mathematical enumeration and tables of cycle index equations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jlindsey@ncsu.edu (J.S.L.), mtanigu@ncsu.edu (M.T.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0001035.

REFERENCES

(1) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3−50.

(2) Weininger, D. Combinatorics of Small Molecular Structures. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Schreiner, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P., Schaefer, H. F., III, Eds.; John Wiley & Sons, Ltd.: Chichester, 1998; Vol. *1*, pp 425−430.

(3) Ertl, P. Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374−380.

(4) Dobson, C. M. Chemical Space and Biology. *Nature* **2004**, *432*, 824−828.

(5) Fink, T.; Bruggesser, H.; Reymond, J.-L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Da. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 1504−1508.

(6) Kerber, A.; Laue, R.; Meringer, M.; Rücker, C. Molecules in Silico: Potential Versus Known Organic Compounds. *MATCH−Commun. Math. Comp. Chem.* **2005**, *54*, 301−312.

(7) Gorse, A.-D. Diversity in Medicinal Chemistry Space. *Curr. Top. Med. Chem.* **2006**, *6*, 3−18.
(8) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952−2963.
(9) Reymond, J.-L.; van Deursen, R.; Blum, L. C.; Ruddigkeit, L. Chemical Space as a Source for New Drugs. *Med. Chem. Commun.* **2010**, *1*, 30−38.
(10) Varnek, A.; Baskin, I. I. Cheminformatics as a Theoretical Chemistry Discipline. *Mol. Inf.* **2011**, *30*, 20−32.
(11) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.
(12) Taniguchi, M.; Du, H.; Lindsey, J. S. Virtual Libraries of Tetrapyrrole Macrocycles. Combinatorics, Isomers, Product Distributions, and Data Mining. *J. Chem. Inf. Model.* **2011**, *51*, 2233−2247.
(13) Taniguchi, M.; Lindsey, J. S. Diversity, Isomer Composition, and Design of Combinatorial Libraries of Tetrapyrrole Macrocycles. *J. Porphyrins Phthalocyanines* **2012**, *16*, 1−13.
(14) Taniguchi, M.; Lindsey, J. S. Enumeration of Isomers of Substituted Tetrapyrrole Macrocycles: From Classical Problems in Biology to Modern Combinatorial Libraries. In *Handbook of Porphyrin Science*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; World Scientific: Singapore, 2012, Vol. *23*, pp 1−80.
(15) Lin, H.; Fischbach, M. A.; Liu, D. R.; Walsh, C. T. In Vitro Characterization of Salmochelin and Enterobactin Trilactone Hydrolases IroD, IroE, and Fes. *J. Am. Chem. Soc.* **2005**, *127*, 11075−11084.
(16) Luz, Z.; Poupko, R.; Wachtel, E. J.; Zheng, H.; Friedman, N.; Cao, X.; Freedman, T. B.; Nafie, L. A.; Zimmermann, H. Structural and Optical Isomers of Nonamethoxy Cyclotriveratrylene: Separation and Physical Characterization. *J. Phys. Chem. A* **2007**, *111*, 10507−10516.
(17) Hardie, M. J. Recent Advances in the Chemistry of Cyclotriveratrylene. *Chem. Soc. Rev.* **2010**, *39*, 516−527.
(18) Bruce, J. I.; Dickins, R. S.; Govenlock, L. J.; Gunnlaugsson, T.; Lopinski, S.; Lowe, M. P.; Parker, D.; Peacock, R. D.; Perry, J. J. B.; Aime, S.; Botta, M. The Selectivity of Reversible Oxy-Anion Binding in Aqueous Solution at a Chiral Europium and Terbium Center: Signaling of Carbonate Chelation by Changes in the Form and Circular Polarization of Luminescence Emission. *J. Am. Chem. Soc.* **2000**, *122*, 9674−9684.
(19) Yoo, J.; Reichert, D. E.; Welch, M. J. Comparative in Vivo Behavior Studies of Cyclen-Based Copper-64 Complexes: Regioselective Synthesis, X-ray Structure, Radiochemistry, log *P*, and Biodistribution. *J. Med. Chem.* **2004**, *47*, 6625−6637.
(20) Ikeda, A.; Shinkai, S. Novel Cavity Design Using Calix[*n*]arene Skeletons: Toward Molecular Recognition and Metal Binding. *Chem. Rev.* **1997**, *97*, 1713−1734.
(21) Coquière, D.; Le Gac, S.; Darbost, U.; Sénèque, O.; Jabin, I.; Reinaud, O. Biomimetic and Self-Assembled Calix[6]arene-based Receptors for Neutral Molecules. *Org. Biomol. Chem.* **2009**, *7*, 2485−2500.
(22) Morohashi, N.; Narumi, F.; Iki, N.; Hattori, T.; Miyano, S. Thiacalixarenes. *Chem. Rev.* **2006**, *106*, 5291−5316.
(23) Engeldinger, E.; Armspach, D.; Matt, D. Capped Cyclodextrins. *Chem. Rev.* **2003**, *103*, 4147−4173.
(24) Krakowiak, K. E.; Bradshaw, J. S.; Zamecka-Krakowiak, D. J. Synthesis of Aza-Crown Ethers. *Chem. Rev.* **1989**, *89*, 929−972.
(25) Gokel, G. W.; Leevy, W. M.; Weber, M. E. Crown Ethers: Sensors for Ions and Molecular Scaffolds for Materials and Biological Models. *Chem. Rev.* **2004**, *104*, 2723−2750.
(26) Isobe, H.; Sato, S.; Nakamura, E. Synthesis of Disubstituted Cucurbit[6]uril and Its Rotaxane Derivative. *Org. Lett.* **2002**, *4*, 1287−1289.
(27) Day, A. I.; Arnold, A. P.; Blanch, R. J. A Method for Synthesizing Partially Substituted Cucurbit[*n*]uril. *Molecules* **2003**, *8*, 74−84.
(28) Spitler, E. L.; Johnson, C. A., II; Haley, M. M. Renaissance of Annulene Chemistry. *Chem. Rev.* **2006**, *106*, 5344−5386.
(29) Krivokapic, A.; Anderson, H. L. Synthesis and Crystal Structure of a *meso*-Trialkynyl [28]hexaphyrin. *Org. Biomol. Chem.* **2003**, *1*, 3639−3641.
(30) Sankar, J.; Mori, S.; Saito, S.; Rath, H.; Suzuki, M.; Inokuma, Y.; Shinokubo, H.; Kim, K. S.; Yoon, Z. S.; Shin, J.-Y.; Lim, J. M.; Matsuzaki, Y.; Matsushita, O.; Muranaka, A.; Kobayashi, N.; Kim, D.; Osuka, A. Unambiguous Identification of Möbius Aromaticity for *meso*-Aryl-Substituted [28]Hexaphyrins(1.1.1.1.1.1). *J. Am. Chem. Soc.* **2008**, *130*, 13568−13579.
(31) Misra, R.; Chandrashekar, T. K. Structural Diversity in Expanded Porphyrins. *Acc. Chem. Res.* **2008**, *41*, 265−279.
(32) Clark, T. D.; Buehler, L. K.; Ghadiri, M. R. Self-Assembling Cyclic β^3-Peptide Nanotubes as Artificial Transmembrane Ion Channels. *J. Am. Chem. Soc.* **1998**, *120*, 651−656.
(33) Davies, J. S. The Cyclization of Peptides and Depsipeptides. *J. Peptide Sci.* **2003**, *9*, 471−501.
(34) Kofoed, J.; Reymond, J.-L. A General Method for Designing Combinatorial Peptide Libraries Decodable by Amino Acid Analysis. *J. Comb. Chem.* **2007**, *9*, 1046−1052.
(35) Fluxa, V. S.; Reymond, J.-L. On-Bead Cyclization in a Combinatorial Library of 15,625 Octapeptides. *Bioorg. Med. Chem.* **2009**, *17*, 1018−1025.
(36) Hwang, H.; Schatz, G. C.; Ratner, M. A. Coarse-Grained Molecular Dynamics Study of Cyclic Peptide Nanotube Insertion into a Lipid Bilayer. *J. Phys. Chem. A* **2009**, *113*, 4780−4787.
(37) Choi, S.-J.; Jeong, W.-J.; Kang, S.-K.; Lee, M.; Kim, E.; Ryu, D. Y.; Lim, Y.-B. Differential Self-Assembly Behaviors of Cyclic and Linear Peptides. *Biomacromolecules* **2012**, *13*, 1991−1995.
(38) Onoda, A.; Takahashi, A.; Oohora, K.; Onuma, Y.; Hayashi, T. Fibrous Supramolecular Hemoprotein Assemblies Connected with Synthetic Heme Dimer and Apohemoprotein Dimer. *Chem. Biodivers.* **2012**, *9*, 1684−1692.
(39) Nakanishi, T.; Naito, M.; Takeoka, Y.; Matsuura, K. Versatile Self-assembled Hybrid Systems with Exotic Structures and Unique Functions. *Curr. Opin. Colloid Interface Sci.* **2011**, *16*, 482−490.
(40) Schulz, G. E. The Structures of General Porins. In *Bacterial and Eukaryotic Porins*; Benz, R., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2004, pp 25−40.
(41) Reedy, C. J.; Gibney, B. R. Heme Protein Assemblies. *Chem. Rev.* **2004**, *104*, 617−649.
(42) Carlson, J. C. T.; Jena, S. S.; Flenniken, M.; Chou, T.-F.; Siegel, R. A.; Wagner, C. R. Chemically Controlled Self-Assembly of Protein Nanorings. *J. Am. Chem. Soc.* **2006**, *128*, 7630−7638.
(43) Grabow, W. W.; Zakrevsky, P.; Afonin, K. A.; Chworos, A.; Shapiro, B. A.; Jaeger, L. Self-Assembling RNA Nanorings Based on RNAI/II Inverse Kissing Complexes. *Nano Lett.* **2011**, *11*, 878−887.
(44) Poulsen, S.-A.; Gates, P. J.; Cousins, G. R. L.; Sanders, J. K. M. Electrospray Ionisation Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry of Dynamic Combinatorial Libraries. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 44−48.
(45) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. Dynamic Combinatorial Libraries of Macrocyclic Disulfides in Water. *J. Am. Chem. Soc.* **2000**, *122*, 12063−12064.
(46) 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.
(47) Roy, L.; Case, M. A. Recursively Enriched Dynamic Combinatorial Libraries for the Self-Selection of Optimally Stable Proteins. *J. Phys. Chem. B* **2011**, *115*, 2454−2464.
(48) Chakrabarty, R.; Mukherjee, P. S.; Stang, P. J. Supramolecular Coordination: Self-Assembly of Finite Two- and Three-Dimensional Ensembles. *Chem. Rev.* **2011**, *111*, 6810−6918.
(49) Laurent, B. A.; Grayson, S. M. Synthetic Approaches for the Preparation of Cyclic Polymers. *Chem. Soc. Rev.* **2009**, *38*, 2202−2213.
(50) Endo, K. Synthesis and Properties of Cyclic Polymers. *Adv. Polym. Sci.* **2008**, *217*, 121−183.
(51) Bullough, P. A.; Qian, P.; Hunter, C. N. Reaction Center-Light Harvesting Core Complexes of Purple Bacteria. In *The Purple Phototrophic Bacteria*; Hunter, C. N., Daldal, F., Thurnauer, M. C.,

Beatty, J. T., Eds.; Advances in Photosynthesis and Respiration, Vol. *28*; Springer: Dordrecht, The Netherlands, 2009; pp 155−179.
(52) Gabrielsen, M.; Gardiner, A. T.; Cogdell, R. J. Peripheral Complexes of Purple Bacteria. In *The Purple Phototrophic Bacteria*; Hunter, C. N., Daldal, F., Thurnauer, M. C., Beatty, J. T., Eds.; Advances in Photosynthesis and Respiration, Vol. *28*; Springer: Dordrecht, The Netherlands, 2009; pp 135−153.
(53) Westerhuis, W. H. J.; Sturgis, J. N.; Ratcliffe, E. C.; Hunter, C. N.; Niederman, R. A. Isolation, Size Estimates, and Spectral Heterogeneity of an Oligomeric Series of Light-Harvesting 1 Complexes from *Rhodobacter sphaeroides*. *Biochemistry* **2002**, *41*, 8698−8707.
(54) McDermott, G.; Prince, S. M.; Freer, A. A.; Hawthornthwaite-Lawless, A. M.; Papiz, M. Z.; Cogdell, R. J.; Isaacs, N. W. Crystal Structure of an Integral Membrane Light-Harvesting Complex from Photosynthetic Bacteria. *Nature* **1995**, *374*, 517−521.
(55) Koepke, J.; Hu, X.; Muenke, C.; Schulten, K.; Michel, H. The Crystal Structure of the Light-Harvesting Complex II (B800−850) from *Rhodospirillum molischianum*. *Structure* **1996**, *4*, 581−597.
(56) Jamieson, S. J.; Wang, P.; Qian, P.; Kirkland, J. Y.; Conroy, M. J.; Hunter, C. N.; Bullough, P. A. Projection Structure of the Photosynthetic Reaction Centre−Antenna Complex of *Rhodospirrilum rubrum* at 8.5 Å Resolution. *EMBO J.* **2002**, *21*, 3927−3935.
(57) Roszak, A. W.; Howard, T. D.; Southall, J.; Gardiner, A. T.; Law, C. J.; Isaacs, N. W.; Cogdell, R. J. Crystal Structure of the RC-LH1 Core Complex from *Rhodopseudomonas palustris*. *Science* **2003**, *302*, 1969−1972.
(58) Papiz, M. Z.; Prince, S. M.; Howard, T.; Cogdell, R. J.; Isaacs, N. W. The Structure and Thermal Motion of the B800-850 LH2 Complex from *Rps. acidophila* at 2.0 Å Resolution and 100K: New Structural Features and Functionally Relevant Motions. *J. Mol. Biol.* **2003**, *326*, 1523−1538.
(59) Niedzwiedzki, D. M.; Bina, D.; Picken, N.; Honkanen, N.; Blankenship, R. E.; Holten, D.; Cogdell, R. J. Spectroscopic Studies of Two Spectral Variants of Light-Harvesting Complex 2 (LH2) from the Photosynthetic Purple Sulfur Bacterium *Allochromatium vinosum*. *Biochim. Biophys. Acta* **2012**, *1817*, 1576−1587.
(60) Brotosudarmo, T. H. P.; Collins, A. M.; Gall, A.; Roszak, A. W.; Gardiner, A. T.; Blankenship, R. E.; Cogdell, R. J. The Light Intensity under which Cells Are Grown Controls the Type of Peripheral Light-Harvesting Complexes that Are Assembled in a Purple Photosynthetic Bacterium. *Biochem. J.* **2011**, *440*, 51−61.
(61) Springer, J. W.; Parkes-Loach, P. S.; Reddy, K. R.; Krayer, M.; Jiao, J.; Lee, G. M.; Niedzwiedzki, D. M.; Harris, M. A.; Kirmaier, C.; Bocian, D. F.; Lindsey, J. S.; Holten, D.; Loach, P. A. Biohybrid Photosynthetic Antenna Complexes for Enhanced Light-Harvesting. *J. Am. Chem. Soc.* **2012**, *134*, 4589−4599.
(62) Reddy, K. R.; Jiang, J.; Krayer, M.; Harris, M. A.; Springer, J. W.; Yang, E.; Jiao, J.; Niedzwiedzki, D. M.; Pandithavidana, D.; Parkes-Loach, P. S.; Kirmaier, C.; Loach, P. A.; Bocian, D. F.; Holten, D.; Lindsey, J. S. Palette of Lipophilic Bioconjugatable Bacteriochlorins for Construction of Biohybrid Light-Harvesting Architectures. *Chem. Sci.* **2013**, *4*, 2036−2053.
(63) Faulon, J.-L.; Visco, Jr. D. P.; Roe, D. Enumerating Molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Larter, R., Cundari, T. R., Eds.; John Wiley and Sons, Inc.: Hoboken, NJ, 2005, Vol. *21*, pp 209−286.
(64) Cotton, F. A. *Chemical Applications of Group Theory*; John Wiley & Sons: New York, 1990.
(65) Berlin, K.; Jain, R. K.; Richert, C. Are Porphyrin Mixtures Favorable Photodynamic Anticancer Drugs? A Model Study with Combinatorial Libraries of Tetraphenylporphyrins. *Biotechnol. Bioeng. (Comb. Chem.)* **1998**, *61*, 107−118.
(66) Drain, C. M.; Gong, X.; Ruta, V.; Soll, C. E.; Chicoineau, P. F. Combinatorial Synthesis and Modification of Functional Porphyrin Libraries: Identification of New, Amphipathic Motifs for Biomolecule Binding. *J. Comb. Chem.* **1999**, *1*, 286−290.
(67) Demirev, P. A.; Zubarev, R. A. Probing Combinatorial Library Diversity by Mass Spectrometry. *Anal. Chem.* **1997**, *69*, 2893−2900.
(68) Nefedov, A. V.; Sadygov, R. G. A Parallel Method for Enumerating Amino Acid Compositions and Masses of All Theoretical Peptides. *BMC Informatics* [Online] **2011**, *12*, Article 432; http://www.biomedcentral.com/1471-2105/12/432.
(69) He, F.; Emmett, M. R.; Håkansson, K.; Hendrickson, C. L.; Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *J. Proteome Res.* **2004**, *3*, 61−67.
(70) Roberts, J. D.; Caserio, M. C. *Basic Principles of Organic Chemistry*, 2nd ed.; W. A. Benjamin, Inc.: Menlo Park, CA, 1977, p 32.
(71) Mazur, D. R. *Combinatorics. A Guided Tour*; The Mathematical Association of America: Washington, DC, 2010; pp 187−224.
(72) Nefedov, A. V.; Mitra, I.; Brasier, A. R.; Sadygov, R. G. Examining Troughs in the Mass Distribution of All Theoretically Possible Tryptic Peptides. *J. Proteome Res.* **2011**, *10*, 4150−4157.
(73) Edelberg, M. *Solutions to Problems in Introduction to Combinatorial Mathematics*; McGraw Hill Book Co.: New York, 1968; p 103.
(74) Davis, T. *Pólya's Counting Theory*. http://www.geometer.org/mathcircles (accessed November 10, 2012).
(75) Pólya, G.; Tarjan, R. E.; Woods, D. R. *Notes on Introductory Combinatorics*; Birkhäuser: Boston, 1983.
(76) Dineva, R. *The Euler Totient, the Mobius and the Divisor Functions*; Mount Holyoke College: South Hadley, MA, 2005; pp 1−22; http://www.mtholyoke.edu/~robinson/reu/reu05/rdineva1.pdf.
(77) Nieman, J. *Symmetries, Coloring, and Polyanumeration*. M. S. Thesis, School of Mathematical Sciences, Rochester Institute of Technology, 2007.
(78) Bjørge, A. N. *Counting and Coloring with Symmetry. A Presentation of Polya's Enumeration Theorem with Applications*. M. S. Thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2009.
(79) Badar, M.; Iqbal, A. *Polya's Enumeration Theorem*. M. S. Thesis, School of Computer Science, Physics and Mathematics, Linnaeus University, 2010.
(80) Kennedy, B. A.; McQuarrie, D. A.; Brubaker, C. H., Jr. Group Theory and Isomerism. *Inorg. Chem.* **1964**, *3*, 265−268.
(81) Freudenstein, F. The Basic Concepts of Polya's Theory of Enumeration, with Application to the Structural Classification of Mechanisms. *J. Mechan.* **1967**, *3*, 275−290.
(82) Riordan, J. *An Introduction to Combinatorial Analysis*; Princeton University Press: Princeton, NJ, 1980; pp 147−151.
(83) Economou, A. The Symmetry of the Equal Temperament Scale. In *Mathematics & Design 98, Proceedings of the Second International Conference*; Barrallo, J., Ed.; The University of Basque Country: Spain, 1998; pp 557−566.
(84) Baraldi, I.; Fiori, C.; Vanossi, D. On the Cycle Index of Point Groups. *J. Math. Chem.* **1999**, *25*, 23−30.
(85) Tucker, A. C. Burnside/Pólya Counting Formula. In *Handbook of Discrete and Combinatorial Mathematics*; Rosen, K. H., Michaels, J. G., Gross, J. L., Grossman, W. J., Shier, D. R., Eds.; CRC Press: New York, 2000; pp 120−127.
(86) Mossinghoff, M. J. Enumerating Isodiametric and Isoperimetric Polygons. *J. Combin. Theory Ser. A* **2011**, *118*, 1801−1815.
(87) Maclean, D.; Baldwin, J. J.; Ivanov, V. T.; Kato, Y.; Shaw, A.; Schneider, P.; Gordon, E. M. Glossary of Terms Used in Combinatorial Chemistry. *Pure Appl. Chem.* **1999**, *71*, 2349−2365.
(88) Escalante, M.; Zhao, Y.; Ludden, M. J. W.; Vermeij, R.; Olsen, J. D.; Berenschot, E.; Hunter, C. N.; Huskens, J.; Subramanian, V.; Otto, C. Nanometer Arrays of Functional Light Harvesting Antenna Complexes by Nanoimprint Lithography and Host−Guest Interactions. *J. Am. Chem. Soc.* **2008**, *130*, 8892−8893.
(89) Böhm, P. S.; Southall, J.; Cogdell, R. J.; Köhler, J. Single-Molecule Spectroscopy on RC-LH1 Complexes of *Rhodopseudomonas acidophila* Strain 10050. *J. Phys. Chem. B* **2013**, *117*, 3120−3126.
(90) Van Patten, P. G.; Shreve, A. P.; Lindsey, J. S.; Donohoe, R. J. Energy-Transfer Modeling for the Rational Design of Multiporphyrin Light-Harvesting Arrays. *J. Phys. Chem. B* **1998**, *102*, 4209−4216.