JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

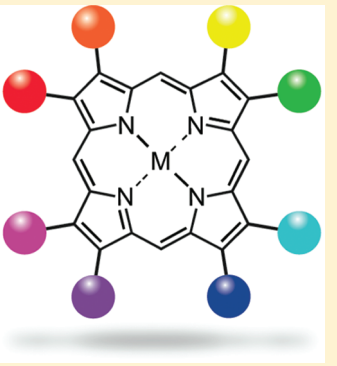
Virtual Libraries of Tetrapyrrole Macrocycles. Combinatorics, Isomers, Product Distributions, and Data Mining

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 software program (*PorphyrinViLiGe*) has been developed to enumerate the type and relative amounts of substituted tetrapyrrole macrocycles in a virtual library formed by one of four different classes of reactions. The classes include (1) 4-fold reaction of *n* disubstituted heterocycles (e.g., pyrroles or diiminoisoindolines) to form β-substituted porphyrins, β-substituted tetraazaporphyrins, or α- or β-substituted phthalocyanines; (2) combination of *m* aminoketones and *n* diones to form *m* × *n* pyrroles, which upon 4-fold reaction give β-substituted porphyrins; (3) derivatization of an 8-point tetrapyrrole scaffold with *n* reagents, and (4) 4-fold reaction of *n* aldehydes and pyrrole to form meso-substituted porphyrins. The program accommodates variable ratios of reactants, reversible or irreversible reaction (reaction classes 1 and 2), and degenerate modes of formation. Pólya's theorem (for enumeration of cyclic entities) has also been implemented and provides validation for reaction classes 3 and 4. The output includes the number and identity of distinct reaction-accessible substituent combinations, the number and identity of isomers thereof, and the theoretical mass spectrum. Provisions for data mining enable assessment of the number of products having a chosen pattern of substituents. Examples include derivatization of an octa-substituted phthalocyanine with eight reagents to afford a library of 2,099,728 members (yet only 6435 distinct substituent combinations) and reversible reaction of six distinct disubstituted pyrroles to afford 2649 members (yet only 126 distinct substituent combinations). In general, libraries of substituted tetrapyrrole macrocycles occupy a synthetically accessible region of chemical space that is rich in isomers (>99% or 95% for the two examples, respectively).



INTRODUCTION

As part of a research program to understand the possible prebiotic origins of tetrapyrrole macrocycles,[1,2] we faced a two-tiered problem of (i) enumerating the macrocycles and their relative amounts formed in a combinatorial process from a mixture of pyrroles wherein each pyrrole bears two (identical or nonidentical) β-substituents and (ii) sorting the resulting virtual library on the basis of particular substituent patterns about the perimeter of the macrocycle. Neither problem has been previously addressed. Such lacunae might seem surprising given the rich history of tetrapyrrole science, where issues of substituent patterns have been paramount from prebiological,[3−5] biosynthetic,[6] synthetic,[7−15] self-assembly,[16−19] and crystal structural[20,21] perspectives. On the other hand, a formal method of enumeration is essential given that the macrocycle symmetry (D_{4h}, D_{2h}, C_{2v}, C_{4h}, or C_s), and hence the number of products, depends on the pattern imparted by the composition of the eight β-pyrrole substituents. In this regard, the creation and data mining of a virtual library of tetrapyrrole macrocycles present challenges atypical of combinatorial libraries wherein the scaffold symmetry remains C_s regardless of the nature of substituents at multiple sites. Morever, the idiosyncratic symmetry properties of tetrapyrrole macrocycles do not appear to be readily accommodated by existing programs devoted to isomer enumeration.[22−24] Here, we have addressed the aforementioned problems and also provide general solutions for the reaction-based formation of a variety of virtual libraries of tetrapyrrole species. For perspective, we first sketch the prior treatments concerning substituent patterns of β-substituted porphyrins and then overview reported tetrapyrrole combinatorial libraries germane to the work described herein.

The issue of the number of possible porphyrin *isomers* for a given set of β-pyrrolic substituents has been well studied. An account of early work in this area is provided by Pilgrim, who manually enumerated the 60 possible isomers of the β-substituted macrocycle known as mesoporphyrin (containing four methyl, two ethyl, and two propionic acid substituents).[25] A general mathematical formalism for the enumeration of all possible species upon substitution of cyclic objects was provided by Pólya.[26] The application of Pólya's theorem to a broad variety of molecular architectures occupies a substantial literature.[27−35] Indeed, several authors employed Pólya's theorem to delineate formally the number of mesoporphyrin isomers.[36−39] Tapscott considered the broader question of the number of possible porphyrin isomers for a given number of β-pyrrole substituents:[39] for an ABCDEFGH-substituted tetrapyrrole, the eight distinct A−H substituents can be arranged in 5040 different ways.

Received: May 31, 2011
Published: August 25, 2011

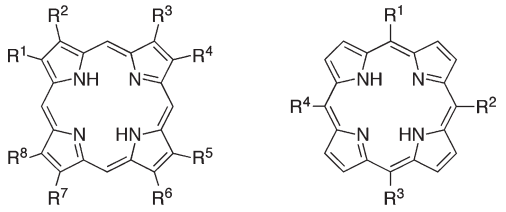


Figure 1. Substituted porphyrins.

Our objectives concern not simply *isomers* stemming from spatial arrangement of a given set of substituents, but the total number of *products* including isomers formed in combinatorial reactions. This more general problem has not been addressed. With regards to porphyrins, Pólya's theorem does enable enumeration of (i) all substituted porphyrins given the number of available sites (e.g., eight β-pyrrole positions or four meso positions; Figure 1) and the number of reactants, (ii) the number of porphyrins with a given collection of substituents, and (iii) the number of isomers of a given substituted porphyrin. On the other hand, Pólya's theorem does not describe a number of relevant chemical features of porphyrin combinatorial libraries. Indeed, the limitations of Pólya's theorem for related problems have been pointed out by others.[31] An overview of such shortcomings and three illustrative examples of Pólya's theorem are provided in the Appendix.

Libraries of porphyrins have been prepared in an approach that predates modern combinatorial chemistry by at least two decades — the reaction of two aldehydes and pyrrole.[7] The resulting mixture is composed of six meso-substituted porphyrins wherein the constituents differ by the nature of the meso-substituents (derived from the aldehyde moiety). The preparation of larger libraries has been pioneered by the groups of Richert,[40,41] Drain,[42] and Boyle.[43] Richert and Drain both employed the reaction of multiple aldehydes with pyrrole, screened the resulting libraries of meso-substituted porphyrins for biological activity, characterized the libraries by mass spectrometry, and wrote software programs to simulate the mass spectra. The program used by Richert has been described in detail.[44] A complementary approach to porphyrin-based libraries employs a meso-substituted porphyrin scaffold that is derivatized with a set of reactants.[45−47] A comprehensive review of porphyrin combinatorial libraries was recently published.[48] To our knowledge, no examples are available of libraries that contain β-pyrrole substituted porphyrins other than those containing mixtures of four isomers such as the uroporphyrin(ogen)s.[3,4]

Phthalocyanines[49−55] (and porphyrazines)[56,57] also are members of the broad class of tetrapyrrole macrocycles and have been targets of synthetic interest for many decades owing to their diverse applications (Figure 2). One method of synthesis of phthalocyanines entails cyclization of a phthalimide, diiminoisoindoline, or *o*-phthalonitrile. The reaction of a monosubstituted phthalonitrile can afford four isomers, whereas the coreaction of two different symmetrically disubstituted phthalonitriles can result in six products, two of which are isomers.[58] Several groups have recently reported the derivatization of phthalocyanine scaffolds (bearing 8 alcohols or 8 ethynes) to create diverse phthalocyanine products.[59−62] Such reports only described the use of a single derivatization reagent; however, libraries should be both accessible and desirable given the generality of the derivatization processes and the envisaged applications.

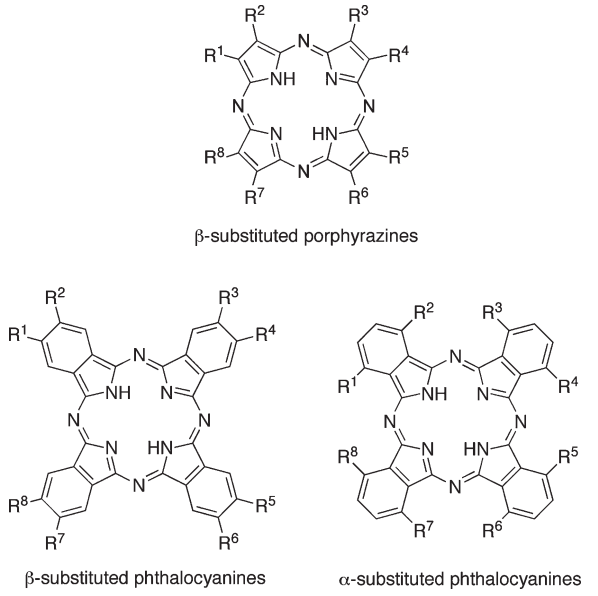


Figure 2. Substituted porphyrazines (tetraazaporphyrins) and phthalocyanines.

In this paper, we describe a program (*PorphyrinViLiGe*) for enumerating the tetrapyrrole macrocycles formed in a virtual library, identifying all isomers, and calculating the distribution (relative amounts based on statistics) of each product. The virtual library can be formed upon four types of combinatorial reactions, including the formation or derivatization of tetrapyrrole macrocycles containing four meso- or eight β-substituents. The latter case encompasses β-substituted porphyrins, β-substituted tetraazaporphyrins, β-substituted phthalocyanines, and α-substituted phthalocyanines. The program also enables data mining to assess the number of products that exhibit chosen patterns of particular substituents. Taken together, the program and results presented herein should provide a quantitative foundation for describing diverse combinatorial libraries of tetrapyrrole macrocycles.

RESULTS

Section 1. Enumeration of Combinatorial Porphyrins. *(A). Reaction Classes.* We have employed a reaction-based approach wherein a generative algorithm constructs the virtual library for a given class of reaction. Four classes of reactions are of interest (Figure 3). Note that the term tetrapyrrole macrocycle is employed in a generic manner to encompass porphyrins, phthalocyanines, and tetraazaporphyrins. The four reactions are as follows:

(1) 4-Fold reaction of *n* heterocycles each of which bears two (A, B: identical or nonidentical) substituents to form tetrapyrrole macrocycles (each composed of four heterocycles). The heterocycle can consist of a precursor to porphyrins (e.g., a pyrrole) or phthalocyanines (e.g., a diiminoisoindoline). The reaction can occur irreversibly (shown) or reversibly (not shown).

(2) Combination of two distinct reactants to form a pyrrole, which upon 4-fold reaction forms a tetrapyrrole macrocycle. An example is given for an aminoketone and a

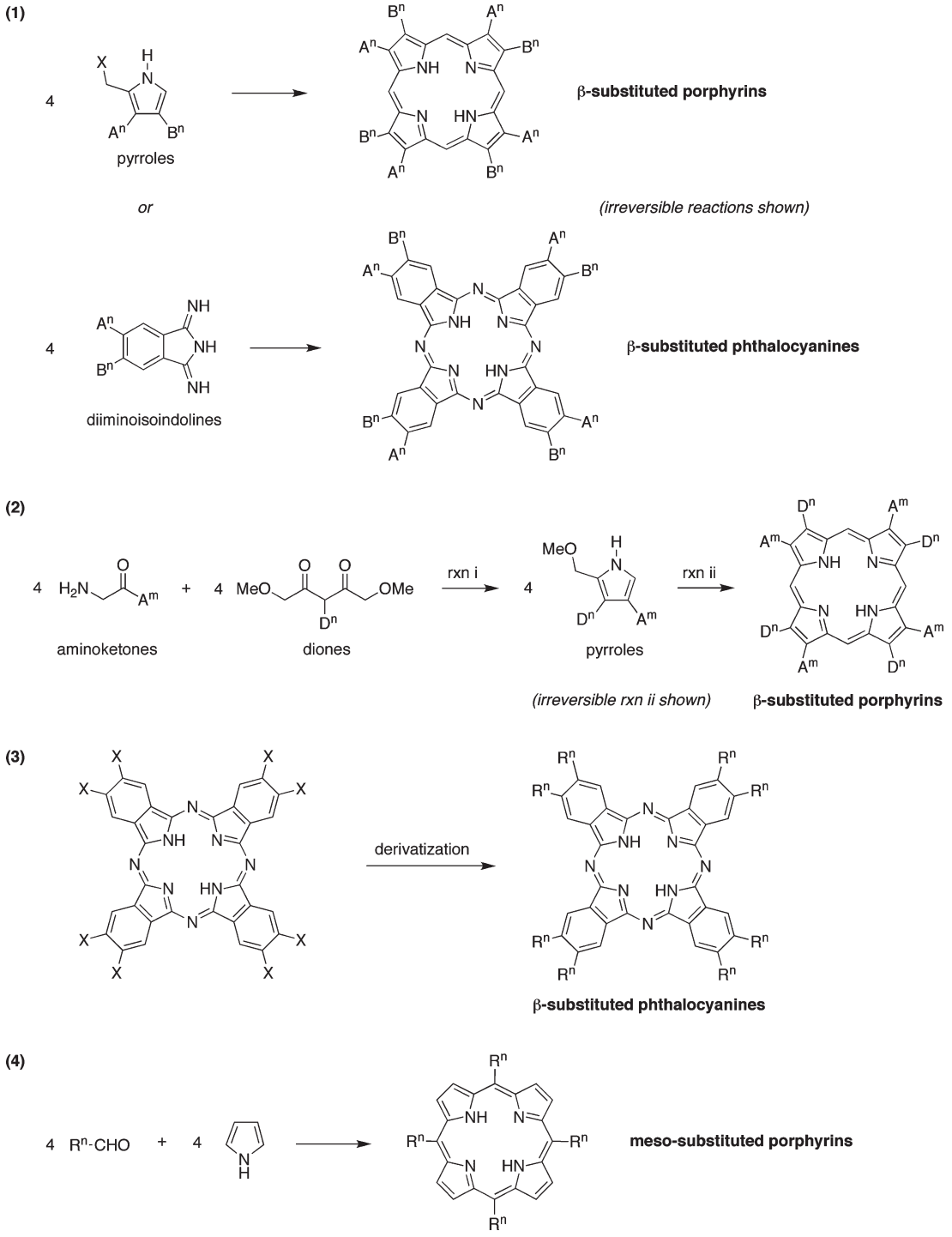


Figure 3. Four distinct reactions affording substituted tetrapyrrole macrocycles.

dione. Thus, m aminoketones X-A^m and n diones Y-D^n give $m \times n$ pyrroles. Each pyrrole contains one A substituent from reagent X and one D substituent from reagent Y. The tetrapyrrole-macrocycle forming reaction can occur irreversibly (shown) or reversibly (not shown).

(3) Derivatization of an 8-point tetrapyrrole scaffold with n reagents. An example is shown for a β-substituted phthalocyanine, yet the approach encompasses all other tetrapyrrole macrocycles.

(4) 4-Fold reaction of n aldehydes with pyrrole to form meso-substituted porphyrins.

Examples of each of these cases are provided below. The products of reaction classes 3 and 4 can be described explicitly via Pólya's theorem (see the Appendix).[26] A module has been included in *PorphyrinViLiGe* for implementation of Pólya's theorem, which provides a convenient means for validating the number of products from reaction classes 3 and 4. On the other hand, Pólya's theorem does not accommodate use of variable

dx.doi.org/10.1021/ci200240e |J. Chem. Inf. Model. 2011, 51, 2233–2247

ratios of reactants, identify the structure of any isomers, or enable data mining of the virtual library; such features are available in the reaction-based description for all reaction classes 1−4. Finally, the term "reversible" is not used here in a strict sense (to imply equilibrium between reactant and product) but rather that exchange processes enable all pyrrolic units to enter the macrocycles equally in intact (i.e., unflipped) and flipped orientations. The *PorphyrinViLiGe* program is described in the next section.

(B). *PorphyrinViLiGe*. The virtual library of tetrapyrrole macrocycles is not constructed on the basis of mathematical formulas but rather via an algorithm that generates the combination of substituents accessible for a given reaction and then counts the number of permutations thereof (i.e., the frequency). The relative amount of a given product depends on such statistics of formation, the fact that a given product can sometimes be made degenerately (i.e., from different sets of reactants), and the ratio of reactants. The generative algorithm accommodates all of these factors in determining the relative amount of each product. A further attribute of this algorithmic approach is the ability to tag substituents with regard to a given physical property (e.g., polarity) for subsequent data mining.

The workflow for the program consists of five stages:

I. Provide input.

(a) Select reaction class (1−4).

(b) Identify members in the reactant chemset,[63,64] the molecular formula of each substituent therein, and the ratio of reactants (the default is equimolar and is assumed for all examples presented herein). The substituents can be abbreviated with alphabetic letters (e.g., H, M, E, V, A for hydrogen, methyl, ethyl, vinyl, acetic acid, respectively).

(c) Specify whether the reaction is reversible or irreversible (classes 1 and 2 only).

(d) Tag substituents with markers to designate a given physical property (optional).

II. Generate all accessible product combinations and permutations.

III. Sort, count, and quantitate.

(a) Identify distinct products.

(b) Count the frequency of each product.

(c) Adjust frequencies on the basis of reactant ratios.

IV. Tabulate and provide output.

The output consists of (i) all tetrapyrrole products; (ii) their relative amounts; (iii) the number of distinct substituent combinations, and (iv) the predicted mass spectrum. The formation of any *N*-confused, expanded, or contracted porphyrins is ignored, as are the tautomers (and different conjugated forms) of free base porphyrins.[65]

The term "substituent combination" refers to the number of types of different substituents for a given macrocycle. For example, the reaction of aldehyde A and aldehyde B with pyrrole affords porphyrins of substituent combinations A_4, A_3B_1, A_2B_2, A_1B_3, and B_4. There are two isomers of the A_2B_2 composition, namely *cis*-A_2B_2 and *trans*-A_2B_2, giving a total of six porphyrin products. The identification of the number of distinct substituent combinations, and the binning of compounds with identical substituent combinations, are valuable for data mining and mass spectrometric predictions. Note that (i) all members having a given substituent combination are isomers, and (ii) the isomers differ only in the pattern of substituents about the perimeter of the macrocycle. Denoting substituent combinations is preferable to molecular formulas here because compounds with identical molecular formulas (i.e., isomers) can result coincidentally and not reflect different patterns of a given set of substituents. For example, an A_4-porphyrin (where A = ethyl) and an A_2B_2-porphyrin (A = propyl, B = methyl) have distinct substituent combinations yet the same molecular formula ($C_{28}H_{30}N_4$). This issue also has implications for the predicted mass spectrum (vide infra). In short, the use of substituent combinations has the convenience of library representations built around Markush structures[66] while retaining features relevant for consideration of the properties of tetrapyrrole macrocycles.

Further details concerning the algorithm, particularly for generating combinations and for identifying distinct products, are described in the Supporting Information (Figures S1 and S2). To our knowledge, the rich output of the *PorphyrinViLiGe* program is not available in other porphyrin-specific[42,44] or generic (nontetrapyrrole) programs.[22−24] Indeed, the standard terminology in the combinatorial chemistry field does not include provisions for denoting isomers.[63,64] This latter shortfall is debilitating for many tetrapyrrole libraries, where >99% of the members can be isomers (vide infra).

Section 2. Isomers of Tetrapyrrole Macrocycles. Distinctions among characteristic types of tetrapyrrole isomers must be drawn to understand those that can be formed in a given chemical route. We choose to focus on the four individual units (hereafter referred to as pyrroles but equally applicable to phthalocyanine or porphyrazine precursors) that compose the tetrapyrrole macrocycle. To facilitate description of the various isomers, we define the two substituents at the β-pyrrole positions as the "pyrrole constituent set" for comparison among pyrroles. The pyrrole constituent set can consist of identical (XX = homosubstituted) or nonidentical members (YZ = heterosubstituted). In the following sections, the β-substituents arrayed about the perimeter of a tetrapyrrole macrocycle are given as a linear string of four pairs of characters, where a dash represents a meso site, and the terminal right-hand dash is understood to be connected to the left-most pair of characters to close the macrocycle. Three characteristic types of β-substituted tetrapyrrole isomers are displayed in Figure 4 and have features described as follows:

(i). Distinct Locations (Exemplified by Nonidentical Homosubstituted Pyrroles). The reaction of an AA-pyrrole and a BB-pyrrole can give rise to two isomers, the *cis*- versus *trans*-substituted porphyrins.

(ii). Distinct Orientations (Exemplified by Identical Heterosubstituted Pyrroles). The reaction of an AB-substituted pyrrole, if irreversible, affords only a type-I porphyrin with substituent pattern AB-AB-AB-AB-. The reversible reaction, on the other hand, affords four isomers that differ in the relative orientation of the pyrrole substituents. Regardless, all pyrrole entities across the entire set of isomers (types I−IV) are identically substituted, namely with one A substituent and one B substituent in the case given. This case is described in more detail in the section on examples (vide infra).

(iii). Distinct Contiguous Sites (Exemplified by Combinations of Homo- and Hetero-Substituted Pyrroles). The combination of homosubstituted pyrroles and heterosubstituted pyrroles can result in substituent patterns about the perimeter of the macrocycle that are not available with either type of pyrrole alone. For example, the reaction of an AA-pyrrole, BB-pyrrole, and AB-pyrrole afford isomers where two identical substituents (e.g., BB) are located at contiguous sites owing to location on one pyrrole or on two adjacent pyrroles flanking an intervening meso position. This case is most relevant for reaction class 2 (vide infra).

Section 3. Examples. We present examples to illustrate the program and provide results concerning virtual combinatorial libraries for each of the reaction classes shown in Figure 3.

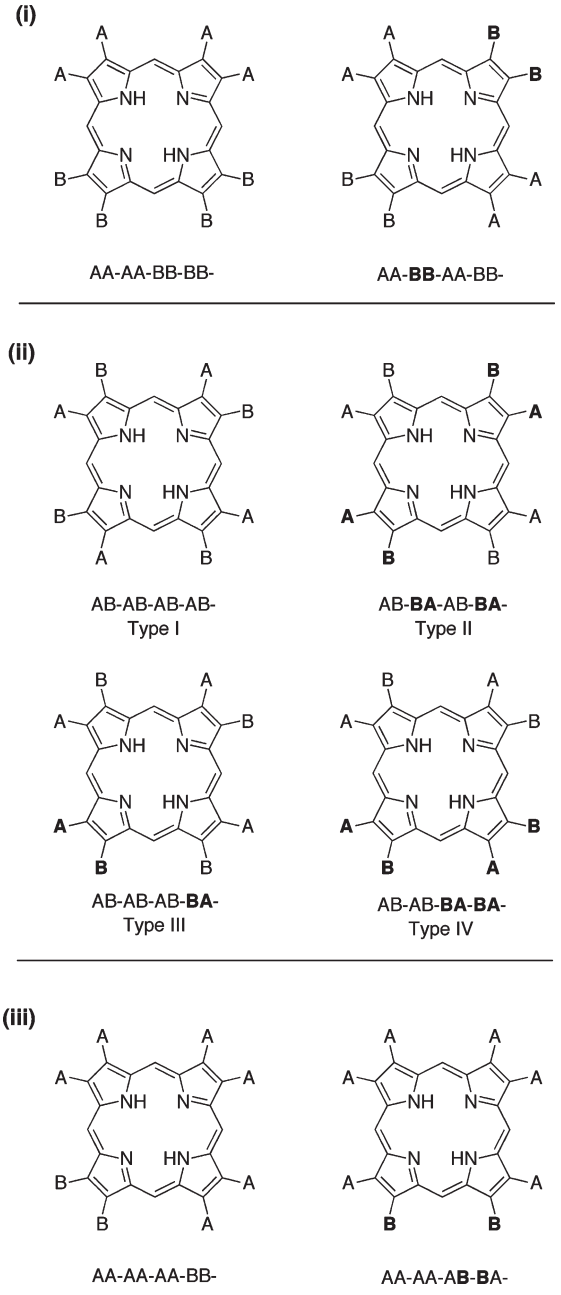


Figure 4. Three types of substituted tetrapyrrole isomers. The linear-string representation is shown beneath each structure. The substituents shown in bold emphasize the specific change in pattern that engenders the isomer.

Reaction Class 1: β-Substituted Porphyrins from Diverse Pyrroles. A heterosubstituted pyrrole upon reversible reaction affords four isomers as shown in Table 1. This case corresponds to the well-known example (and perhaps the most readily intuitive example in this domain) of porphobilinogen, which bears acetic and propionic acid substituents, converting to give the four uroporphyrinogen isomers (and upon dehydrogenation, the corresponding uroporphyrins).[3,4] For irreversible reaction (no "flipping" of the orientation of the pyrrole upon incorporation into the tetrapyrrole), of course, only the type-I isomer forms (Figure 4-ii). The relative amounts of each isomer are given by the number of permutations (i.e., the frequency) of the isomer versus the total number of permutations for all isomers.

Table 1. Porphyrins from an AB-Pyrrole

porphyrin type	permutations	frequency	relative amount in %
Type I	AB-AB-AB-AB-	2	12.5
	BA-BA-BA-BA-		
Type II	AB-BA-AB-BA-	2	12.5
	BA-AB-BA-AB-		
Type III	AB-AB-AB-BA-	8	50
	AB-AB-BA-AB-		
	AB-BA-AB-AB-		
	BA-AB-AB-AB-		
	BA-BA-BA-AB-		
	BA-BA-AB-BA-		
	BA-AB-BA-BA-		
	AB-BA-BA-BA-		
Type IV	AB-AB-BA-BA-	4	25
	AB-BA-BA-AB-		
	BA-BA-AB-AB-		
	BA-AB-AB-BA-		

Table 2. Porphyrins from an AA-Pyrrole + a BA-Pyrrole

porphyrins and permutations	relative amount (irreversible) in %	relative amount (reversible) in %
AA-AA-AA-AA-	6.25	6.25
AA-AA-AA-BA-	25.00	25.0
AA-AA-BA-BA-	25.00	12.50
AA-AA-BA-AB-		6.25
AA-AA-AB-BA-		6.25
AA-BA-AA-BA-	12.50	6.25
AA-BA-AA-AB-		6.25
AA-BA-BA-BA-	25.00	6.25
AA-BA-BA-AB-		6.25
AA-BA-AB-BA-		6.25
AA-AB-BA-BA-		6.25
BA-BA-BA-BA-	6.25	0.78
BA-BA-BA-AB-		3.13
BA-BA-AB-AB-		1.56
BA-AB-BA-AB-		0.78

The reversible reaction of two pyrroles (with AA- and BA-substituents at the pyrrole 3- and 4-positions) affords 15 porphyrins, whereas irreversible reaction affords only 6 porphyrins (Table 2). For the case of two pyrroles (with AB- and AC-substituents), reversible reaction affords 39 porphyrins, whereas irreversible reaction again affords 6 porphyrins (Table 3). We note that the results shown in Tables 2 and 3 for irreversible reaction are identical to those upon use of the binomial equation,[67] except the binomial equation does not delineate *cis*- and *trans*-isomers.

The approach taken in Tables 1-3 was extended to assess the number of β-substituted porphyrins formed from a larger number of distinct heterosubstituted pyrroles upon reversible or irreversible reaction. The results for reactions from 1 to 25 distinct pyrroles are shown in Table 4 and displayed in Figure 5. For example, the case of four pyrroles is carried out for AB-, AC-, AD, and AE-pyrroles (summarized by "AB to AE" in the table).

Table 3. Porphyrins from an AB-Pyrrole + an AC-Pyrrole

porphyrins and permutations	relative amount (irreversible) in %	relative amount (reversible) in %
AB-AB-AB-AB-	6.25	0.78
AB-AB-AB-BA-		3.13
AB-AB-BA-BA-		1.56
AB-BA-AB-BA-		0.78
AB-AB-AB-AC-	25.00	3.13
AB-AB-AB-CA-		3.13
AB-AB-AC-BA-		3.13
AB-AB-BA-AC-		3.13
AB-AB-BA-CA-		3.13
AB-AB-CA-BA-		3.13
AB-AC-AB-BA-		3.13
AB-BA-AB-CA-		3.13
AB-AB-AC-AC-	25.00	3.13
AB-AB-AC-CA-		3.13
AB-AB-CA-AC-		3.13
AB-AB-CA-CA-		3.13
AB-BA-AC-AC-		3.13
AB-BA-AC-CA-		1.56
AB-BA-CA-AC-		1.56
AB-AC-AC-BA-		3.13
AB-AC-CA-BA-		1.56
AB-CA-AC-BA-		1.56
AB-AC-AB-AC-	12.50	1.56
AB-AC-AB-CA-		3.13
AB-AC-BA-AC-		3.13
AB-AC-BA-CA-		3.13
AB-CA-AB-CA-		1.56
AB-AC-AC-AC-	25.00	3.13
AB-AC-AC-CA-		3.13
AB-AC-CA-AC-		3.13
AB-AC-CA-CA-		3.13
AB-CA-AC-AC-		3.13
AB-CA-AC-CA-		3.13
AB-CA-CA-AC-		3.13
AB-CA-CA-CA-		3.13
AC-AC-AC-AC-	6.25	0.78
AC-AC-AC-CA-		3.13
AC-AC-CA-CA-		1.56
AC-CA-AC-CA-		0.78

Note that the product distributions obtained upon reaction of one pyrrole (entry 1, Table 4) and two pyrroles (entry 2, Table 4) are presented in detail in Tables 1 and 3, respectively.

Two trends emerge from the data in Table 4 and Figure 5. First, the number of porphyrins formed via reversible reaction is nearly 8 times greater than that via irreversible reaction. Second, the number of porphyrins formed upon irreversible reaction is still larger than the number of distinct substituent combinations. As the term implies, the number of distinct substituent *combinations* is akin to a condensed formula of substituents. The *permutations* of each substituent combination (by variation in location of intact pyrroles) that accrue upon irreversible reaction increase the number of products; the further *permutations* (by variation in orientation of intact pyrroles) upon reversible reaction afford a commensurate increase in the number of products. Thus, the larger number of products versus the number of distinct substituent combinations stems entirely from the formation of isomers. For the substituent combination A_4B_2CD, for example, isomers arise owing to all of the distinct (irreversible or reversible) reaction-accessible patterns of 4A, 2B, 1C, and 1D substituents. In general, as the combinatorial library expands, the percentage of isomers increases (on the basis of substituent combinations). With eight pyrroles, for example, irreversible or reversible reaction results in a porphyrin library wherein 68% (1044) or 96% (8292) of the members are isomers, respectively, of the 330 distinct substituent combinations.

It warrants mention that all pyrroles shown in Table 4 carry a common substituent (A). Nonetheless, the results obtained are identical to those for the case wherein each pyrrole bears unique substituents (e.g., AB, CD, EF, GH). For example, whether 10 distinct heterosubstituted pyrroles in a chemset each bear only one common substituent (AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK) or no common substituents (AB, CD, EF, GH, IJ, KL, MN, OP, QR, ST), the number of substituent combinations and products are identical — there are 20155 porphyrins upon reversible reaction, 2530 porphyrins upon irreversible reaction, and 715 distinct substituent combinations. In addition, none of the isomers stems from coincidental degeneracies (*vide infra*).

Reaction Class 2: β-Substituted Porphyrins from Diverse Acyclic Reactants. To better understand the possible prebiogenesis of tetrapyrrole macrocycles, we recently investigated nonenzymic routes that broadly parallel the modern biosynthesis. Two pathways were discovered,[1,2] one of which is shown in Figure 3 (panel 2). In the pathway shown, an aminoketone and a 2,4-dione combine to form a pyrrole, four molecules of which then self-condense to give a porphyrinogen macrocycle. Oxidation of the latter affords the corresponding porphyrin.[2] Each of the two acyclic reactants provides one β-pyrrole substituent. The complexity of the resulting distribution of porphyrins depends on the composition of each reactant chemset and whether the members of the intermediate set of pyrroles are heterosubstituted or homosubstituted. The number of distinct porphyrins is provided in Table 5 for a variety of compositions of the two chemsets, including for reversible or irreversible self-condensation of the intermediate pyrroles. There are several interesting observations that warrant discussion.

First, the formation of the porphyrins is a consequence of two sequential combinatorial transformations: the reaction of members of an aminoketone chemset and members of a dione chemset to form a pyrrole chemset (reaction i); the self-condensation of members of the latter to form a porphyrin library (reaction ii). In a given porphyrin macrocycle, the combination of substituents at a given pyrrole is dictated by the nature of the substituents in the aminoketone and dione chemsets following reaction i, whereas the orientations of the respective set of substituents on one pyrrole unit with respect to another — and the number of resulting isomers — depends on whether reaction ii is irreversible or reversible. Accordingly, the reaction of aminoketone chemset [A,B] × dione chemset [C,D] results in diverse porphyrins, but none contains a homosubstituted pyrrole or an AB- or CD-substituted pyrrole (here the nomenclature [X,Y] refers to substituent X or Y at one site in the reactant chemset). The substituent patterns are so constrained because each acyclic chemset provides one substituent to each pyrrole.

Second, the number of porphyrins depends both on any degeneracy of substituents in the two chemsets and the

Table 4. Number of Distinct Porphyrins from Pyrroles

# of distinct pyrroles	pyrrole substituents	# of substituent combinations	# of porphyrins (irreversible)	# of porphyrins (reversible)
1	AB	1	1	4
2	AB, AC	5	6	39
3	AB, AC, AD	15	24	177
4	AB to AE	35	70	538
5	AB to AF	70	165	1290
6	AB to AG	126	336	2649
7	AB to AH	210	616	4879
8	AB to AI	330	1044	8292
9	AB to AJ	495	1665	13248
10	AB to AK	715	2530	20155
11	AB to AL	1001	3696	29469
12	AB to AM	1365	5226	41694
13	AB to AN	1820	7189	57382
14	AB to AO	2380	9660	77133
15	AB to AP	3060	12720	101595
16	AB to AQ	3876	16456	131464
17	AB to AR	4845	20961	167484
18	AB to AS	5985	26334	210447
19	AB to AT	7315	32680	261193
20	AB to AU	8855	40110	320610
21	AB to AV	10626	48741	389634
22	AB to AW	12650	58696	469249
23	AB to AX	14950	70104	560487
24	AB to AY	17550	83100	664428
25	AB to AZ	20475	97825	782200

reversibility/irreversibility of the reaction. For complete degeneracy in the substituents in the two chemsets, reversible and irreversible reaction provides equivalent numbers of porphyrins. This can be readily seen in the case of aminoketone chemset [A,B] × dione chemset [A,B], which affords pyrroles with AA, BB, AB, and BA substituents. Here, the "flipping" of β-pyrrole substituents that occurs upon reversible reaction, but not irreversible reaction, is irrelevant to the final number of tetrapyrroles given that the pool of pyrroles contains both AB and BA patterns.

Third, when there is no degeneracy among substituents in the two chemsets, the number of porphyrins increases, and the increase is most marked for reversible reactions. For example, where degeneracy exists, such as in the case of [A,B] × [A,B], 4 pyrroles [AA, AB, BA, BB] are formed but only 43 porphyrins (for reversible or irreversible reaction). By contrast, in the absence of degeneracy, such as in the case of [A,B] × [C,D], there again are 4 pyrroles [AC, AD, BC, BD] but now 70 or 538 porphyrins upon irreversible or reversible reaction, respectively. The general cases of no degeneracy among substituents in the aminoketone and dione chemsets are shown in Table 5 in the last rows of each group of [n × m]; for example, [A,B,C,D] × [E,F,G], or [A,B,C,D,E] × [F,G,H,I,J]. In the latter example, the reaction of A,B,C,D,E-aminoketones with F,G,H,I,J-diones gives 25 distinct pyrroles; the number of porphyrins is 97825 (irreversible) or 782200 (reversible). The dramatic increase upon reversible reaction stems from the large number of isomers formed. The possible prebiotic implications of such combinatorial processes will be described elsewhere.

Reaction Class 3: Derivatization of 8-Point Tetrapyrrole Scaffolds. Recently several groups have reported the derivatization of phthalocyanine scaffolds bearing 8 alcohols or 8 ethynes.[59−62] Analogous reactions have been described with porphyrin scaffolds bearing eight bromine units.[14] While such reports only describe the use of a single derivatization reagent, consideration of library preparation raises the issue of the number of products formed upon derivatization. The results upon derivatization with *n* reagents of eight sites of a tetrapyrrole macrocycle (the eight β-pyrrole positions of a porphyrin or porphyrazine or the eight inner (α) or eight outer (β) positions of a phthalocyanine) are given in Table 6. In particular, the use of merely four distinct derivatization reagents affords 8356 tetrapyrrole products, whereas eight derivatization reagents affords a library with >2 million members. The computational method and results for derivatization of phthalocyanines are identical to those for the reaction of aminoketones + diones with degenerate substituents in the two chemsets (see in Table 5 the first rows of each group of [n × m], e.g., [A,B,C,D] × [A,B,C,D], [A,B,C,D,E] × [A,B,C,D,E]).

The explicit mathematical treatment for this particular case has been described by Pólya.[26,36−39] In Pólya's treatment, the derivatization of an 8-point tetrapyrrole scaffold (e.g., β-octa-substituted porphyrins, phthalocyanines) to introduce A and B substituents affords *N* products as described by eq 1[26]

$$N = A_8 + A_7B + 6A_6B_2 + 7A_5B_3 + 13A_4B_4 + 7A_3B_5 + 6A_2B_6 + AB_7 + B_8 \quad (1)$$

Equation 1, referred to as the 'pattern inventory',[28] describes three aspects about the product distribution. First, the number of terms in the polynomial expansion equals the *number of substituent combinations*, which in this example is nine, ranging from

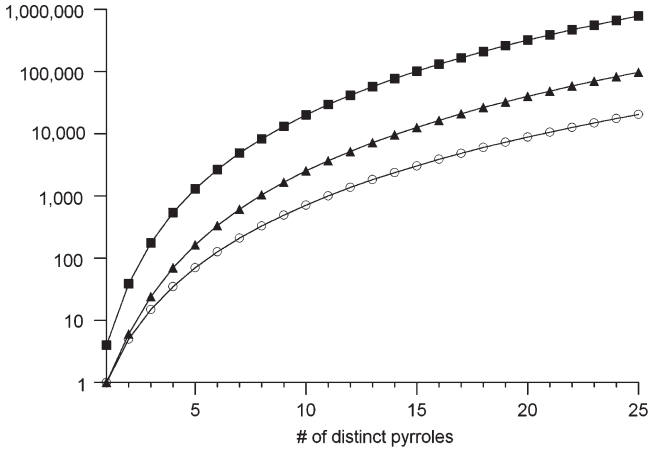


Figure 5. Number of porphyrins (solid squares, reversible reaction; solid triangles, irreversible reaction) and number of distinct substituent combinations (open circles) as a function of distinct pyrrole reactants. The increase in products above the "substituent combinations" line is due to isomers (derived from reaction-accessible permutation of the substituent combinations). The calculated data points are displayed with graphical interpolation to guide the eye.

A_8 to B_8. Second, each coefficient gives the *number of isomers* of that particular substituent combination. Thus, the A_8 species is unique (i.e., no isomers), whereas there are 6 isomers of the A_6B_2 porphyrin, 7 isomers of the A_5B_3-porphyrin, 13 isomers of the A_4B_4-porphyrin, and 6 isomers of the A_2B_6-porphyrin. Third, the sum of the coefficients gives the total *number of products (including isomers)*, which in this case is 1 + 1 + 6 + 7 + 13 + 7 + 6 + 1 + 1 = 43. A more full description of this and two other examples is provided in the Appendix.

A module has been incorporated in *PorphyrinViLiGe* that enables enumeration of the library (reaction class 3, Figure 3) on the basis of Pólya's theorem. The aforementioned three aspects of the library are readily assessed. The results in Table 6 for up to eight reagents were obtained equally with the generative algorithm and via the module for Pólya's theorem, whereas those beyond eight reagents were only obtained with the latter owing to its greater computational speed. Note that the use of 16 reagents affords a library of >0.5 billion members. While large libraries have been known for some time, the potential to create such a library on one scaffold in a one-flask process may be novel. It warrants re-emphasis that the number of permutations of ABCDEFGH substituents about the macrocycle is 5040 (see Introduction and the Appendix);[39] hence, the great size of libraries of derivatized tetrapyrroles stems in part from the numerous isomers derived by permutation of the substituent combinations.

Although enumeration via Pólya's theorem is computationally faster than via the generative algorithm, there are important results provided in *PorphyrinViLiGe* that are not available via Pólya's theorem. One example entails the relative amounts of the library members. Thus, derivatization of an 8-point phthalocyanine (1 mM) with four reagents (A, B, C, and D; 2 mM each) affords, in theory, a combinatorial mixture composed of 8356 phthalocyanines (total concentration, 1 mM). One of the least abundant species, the A_8- (or B_8-, C_8-, D_8-) phthalocyanine, is formed in 0.01526% yield (0.1526 μM), whereas one of the most abundant species, the A_7B-phthalocyanine, is formed in ~8-fold greater quantity (0.122% yield, 1.22 μM). The most abundant substituent combinations are provided by the $A_2B_2C_2D_2$-phthalocyanines, which comprise 330 isomers in 3.845% yield (38.45 μM). More extensive results for this assessment are provided in the Supporting Information (Table S1).

Reaction Class 4: meso-Substituted Porphyrin from n Aldehydes + Pyrrole. Enumeration of the distribution of meso-substituted porphyrins formed upon reaction of aldehydes and pyrrole is relatively straightforward owing to the limited number of positions of substitution (four for meso-substituted porphyrins vs eight for β-substituted porphyrins). The results are listed in Table 7 and displayed graphically in Figure 6. The results shown here exactly match those of Drain[42] and those upon implementation of Pólya's theorem (in *PorphyrinViLiGe*). For example, the reaction of 15 aldehydes + pyrrole gives 7260 porphyrins.[42] Regardless, the generative algorithm affords all of the advantages here as described for use with β-substituted porphyrins. An example of data mining of this virtual library is described below. The facile implementation of this reaction has led to the formation of a number of combinatorial libraries.[48]

In summary, the *PorphyrinViLiGe* program has been validated by agreement with the following published results: (i) the number and ratio of porphyrins upon reaction of an AB-pyrrole predicted by statistical considerations;[3,4] (ii) counting of 60 protoporphyrin isomers by manual or mathematical means;[25,36−39] (iii) the number of meso-substituted porphyrins predicted by Pólya's theorem, which also is in accord with the results of Drain;[42] and (iv) the number of octa-derivatized tetrapyrrole macrocycles predicted by Pólya's theorem (up through eight derivatization reagents).[26,36−39]

Section 4. Data Mining and Filtering. Virtual libraries have been examined by a wide variety of methods, including filtering and data mining. The chief motivations for filtering and data mining have been to reduce the size of the library, to group members on the basis of similar structure or expected properties (i.e., binning), and to identify target compounds.[66,68−71] Still, all such predictive efforts are necessarily imperfect given the discrete (noncontinuous) nature of molecular members that constitute the library, and discontinuities and outliers result.[72,73] The *PorphyrinViLiGe* program enables the virtual library to be filtered with respect to substituent combinations and mined to identify molecular products (and their percentage) with a given number and/or pattern of substituents. Two examples are provided in the following sections. The tagging of substituents to identify porphyrins having expected physical properties will be described elsewhere.

(A). Numbers of a Given Substituent. Controlling the oxidation potential of porphyrin macrocycles is essential to achieve desired properties for many applications. Consider the combinatorial distribution of meso-substituted porphyrins derived upon reaction of three aldehydes (benzaldehyde, 4-methoxybenzaldehyde, and pentafluorobenzaldehyde) in equal amounts + pyrrole. Here, data mining can be done to identify the number of porphyrins having a given number of pentafluorophenyl groups. Among 21 total porphyrins, the number of porphyrins in each subset is given in Table 8.

(B). Particular Patterns of Substituents. The number of contiguous substituents about the perimeter of the macrocycle is a pattern of interest for a variety of self-assembly or recognition processes. We define contiguous to refer to an uninterrupted tract of substituents upon circumnavigating the macrocycle, which encompasses the β-pyrrole substituents on a given pyrrole unit and/or the adjacent β-pyrrole substituents (separated by a

Table 5. Number of Distinct Porphyrins from Aminoketones + Diones

# of distinct substituents	combination of aminoketones × diones	# of distinct pyrroles	# of substituent combinations	# of porphyrins (irreversible)	# of porphyrins (reversible)
1 × 1					
1	[A] × [A]	1	1	1	1
2	[A] × [B]	1	1	1	4
2 × 1					
2	[A,B] × [A]	2	5	6	15
3	[A,B] × [C]	2	5	6	39
2 × 2					
2	[A,B] × [A,B]	4	9	43	43
3	[A,B] × [B,C]	4	25	70	322
4	[A,B] × [C,D]	4	25	70	538
3 × 2					
3	[A,B,C] × [A,B]	6	35	309	546
4	[A,B,C] × [C,D]	6	75	336	1881
5	[A,B,C] × [D,E]	6	75	336	2649
3 × 3					
3	[A,B,C] × [A,B,C]	9	45	873	873
4	[A,B,C] × [B,C,D]	9	125	1638	4893
5	[A,B,C] × [C,D,E]	9	225	1665	10557
6	[A,B,C] × [D,E,F]	9	225	1665	13248
4 × 3					
4	[A,B,C,D] × [A,B,C]	12	145	4434	6450
5	[A,B,C,D] × [C,D,E]	12	325	5199	20175
6	[A,B,C,D] × [D,E,F]	12	525	5226	35190
7	[A,B,C,D] × [E,F,G]	12	525	5226	41694
4 × 4					
4	[A,B,C,D] × [A,B,C,D]	16	165	8356	8356
5	[A,B,C,D] × [B,C,D,E]	16	425	15664	35236
6	[A,B,C,D] × [C,D,E,F]	16	825	16429	77161
7	[A,B,C,D] × [D,E,F,G]	16	1225	16456	115816
8	[A,B,C,D] × [E,F,G,H]	16	1225	16456	131464
5 × 4					
5	[A,B,C,D,E] × [A,B,C,D]	20	460	32010	41790
6	[A,B,C,D,E] × [C,D,E,F]	20	1000	39318	115878
7	[A,B,C,D,E] × [D,E,F,G]	20	1750	40083	210483
8	[A,B,C,D,E] × [E,F,G,H]	20	2450	40110	289770
9	[A,B,C,D,E] × [F,G,H,I]	20	2450	40110	320610
5 × 5					
5	[A,B,C,D,E] × [A,B,C,D,E]	25	495	49225	49225
6	[A,B,C,D,E] × [B,C,D,E,F]	25	1175	89725	167620
7	[A,B,C,D,E] × [C,D,E,F,G]	25	2275	97033	353953
8	[A,B,C,D,E] × [D,E,F,G,H]	25	3675	97798	560533
9	[A,B,C,D,E] × [E,F,G,H,I]	25	4900	97825	721525
10	[A,B,C,D,E] × [F,G,H,I,J]	25	4900	97825	782200

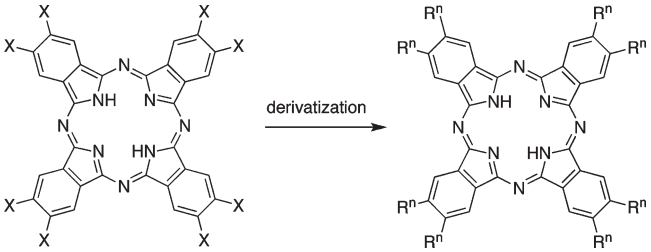
meso position) on neighboring pyrroles. One example concerns the number of contiguous vinyl groups, given that the vinyl groups can be exploited in derivatization or binding processes.

Data mining for this particular pattern of substituents is exemplified by porphyrin formation from a chemset of three diones (methyl, ethyl, vinyl) and a chemset of three aminoketones (methyl, ethyl, vinyl); all in equal amounts. The reaction of such degenerately substituted chemsets ([A,B,C] × [A,B,C]) gives 873 porphyrins (Table 5). The results of data mining to identify the number of porphyrins with a given number of contiguous vinyl groups in a single tract are shown in Table 9. (Porphyrins with more than one such tract are not counted here but could be if desired.) Thus, of 873 porphyrins in the combinatorial library, 141 porphyrins contain a pattern of at least 2 contiguous vinyl groups (with theoretical yield of 15.379%).

Section 5. Theoretical Mass Histograms for Combinatorial Distributions. Mass spectrometry can be a powerful method for assessing the members of a combinatorial library. Predicting the

dx.doi.org/10.1021/ci200240e |J. Chem. Inf. Model. 2011, 51, 2233–2247

Table 6. Number of Distinct Tetrapyrrole Macrocycles Upon Derivatization of Eight Sites



# of derivatization reagents	# of substituent combinations	# of tetrapyrroles
1	1	1
2	9	43
3	45	873
4	165	8356
5	495	49225
6	1287	210771
7	3003	722113
8	6435	2099728
9	12870	5384961
10	24310	12506275
11	43758	26804041
12	75582	53760708
13	125970	101984233
14	203490	184497691
15	319770	320393025
16	490314	536911936

Table 7. Number of Distinct meso-Substituted Porphyrins from Aldehydes and Pyrrole

# of aldehydes	# of substituent combinations	# of porphyrins
1	1	1
2	5	6
3	15	21
4	35	55
5	70	120
6	126	231
7	210	406
8	330	666
9	495	1035
10	715	1540
11	1001	2211
12	1365	3081
13	1820	4186
14	2380	5565
15	3060	7260
16	3876	9316
17	4845	11781
18	5985	14706
19	7315	18145
20	8855	22155

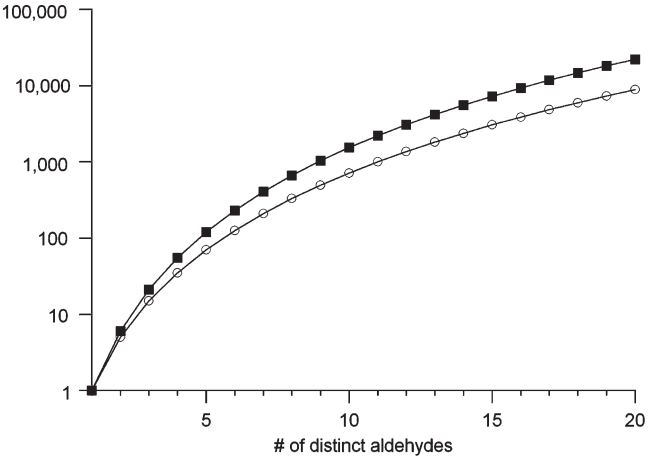


Figure 6. Number of porphyrins (solid squares) and number of substituent combinations (open circles) as a function of the number of distinct aldehydes in the reactant chemset. The calculated data points are displayed with graphical interpolation to guide the eye.

Table 8. Number of Pentafluorophenyl Groups (Data Mining Example A)

# of C_6F_5 groups	# of products	theoretical yield (%)
1	6	39.506
2	6	29.630
3	2	9.877
4	1	1.235

theoretical mass spectrum of a virtual library is thus of considerable value for comparison with experiment. The *PorphyrinViLiGe* program contains a mass spectral database[74] and generates the theoretical mass spectrum for a given virtual library. A polynomial-based approach is employed to calculate the isotopic distribution; the user can define the bin size (default = 0.1 mDa) and the abundance cutoff (default = 0.01% of baseline). The user also can choose to display monoisotopic (exact mass) peaks exclusively or the complete isotopic distribution.

Here, the number of "distinct substituent combinations" provides an upper bound on the number of distinct unimolecular peaks in the theoretical mass spectrum of the entire library owing to two reasons. (1) Isomers due to distinct substituent patterns. In most distributions, the number of products is far greater than the number of distinct substituent combinations due to the presence of isomers wherein the substituents occupy available sites about the perimeter of the macrocycle. (2) Coincidental molecular formulas. The actual number of molecular ion peaks will be less than the number of distinct substituent combinations in the case of coincidental degeneracies in molecular formulas. For example, octaethylporphyrin and tetramethyltetrapropylporphyrin have distinct substituent combinations (Et_8 vs Me_4Pr_4), but the compounds are isomers and obviously afford a single coincident molecular ion peak upon mass spectral analysis. Both features (1) and (2) can sharply limit the ability to unambiguously assign a given peak to a specific member of the library. A further practical impediment is that accurate mass measurements of some species in complex combinatorial libraries are often thwarted by partially overlapping peaks.[75] Of course, if the library can be fractionated on the basis of a physical property (e.g., by chromatography) other than mass, each resolved member could

Table 9. Number of Porphyrins With a Single Tract of Contiguous Vinyl Groups (Data Mining Example B)

# of contiguous vinyl groups	# of products	theoretical yield (%)	# of contiguous vinyl groups	# of products	theoretical yield (%)
= 2	72	7.804	≥2	141	15.379
= 3	32	3.902	≥3	69	7.575
= 4	20	1.951	≥4	37	3.673
= 5	8	0.975	≥5	17	1.722
= 6	6	0.488	≥6	9	0.747
= 7	2	0.244	≥7	3	0.259
= 8	1	0.015			

then be analyzed by subsequent mass analysis and thereby avoid the aforementioned problems (of isomers, coincidental molecular formulas, and partially overlapping mass spectral peaks) that accrue upon analysis of the intact (unfractionated) library. Chromatographic methods have been extensively studied for tetrapyrrole macrocycles yet to date have chiefly focused on mixtures consisting of <10 porphyrin members, and difficulties persist in separating ostensibly tractable isomers.[76−79] Similar challenges are observed in the phthalocyanine arena.[80]

In the virtual libraries described herein, the number of distinct substituent combinations often was a fraction (even <1%) of the total number of members. Such a high ratio reflects the large number of isomers typically formed and also indicates the limitation of the use of mass spectrometry to characterize the individual members of the intact library. For example, consider the porphyrins formed from a chemset of four diones [A,B,C,D] and a chemset of four aminoketones [A,B,C,D]. As described in the previous section (Table 5), 8356 porphyrin products result yet only 165 molecular ion peaks (equal to the number of distinct substituent combinations) are expected by mass spectrometry (assuming no coincidental degeneracies in molecular formulas). On the other hand, typical reactions may well result in substantial coincidental degeneracies, affording far fewer peaks. For example, when A = -H, B = -CH_3, C = -C_2H_5, and D = -C_3H_7, only 31 molecular ion peaks are expected despite the presence of 8356 porphyrin products. The profound difference in expectations for observed peaks versus actual species stems both from isomers with the same substituent combination as well as coincidental degeneracies in molecular formulas.

OUTLOOK

The ability to enumerate the types and amounts of tetrapyrrole products formed upon combinatorial reactions, and to perform subsequent data mining on the resulting virtual library, may prove useful in a broad spectrum of endeavors across porphyrin science. While most combinatorial libraries of porphyrins have been formed by reaction of a chemset of aldehydes with pyrrole,[48] the synthetic chemistry of porphyrins has now advanced to the point where a variety of other synthetic approaches can be employed.[8−15] Similarly, the extensive chemistry of pyrroles[81,82] should enable expansion beyond the existing libraries.[83−87] Our own motivation for the work described herein was to understand the diversity of tetrapyrrole macrocycles that can form via prebiotic reaction pathways. Use of the methods developed herein for such prebiotic studies will be described in due course.

A recent challenge in chemoinformatics has been to gain a deeper understanding of the size of chemical space and the extent to which chemical space is filled by synthetic and virtual combinatorial libraries.[88−98] While estimates of the size of chemical space vary wildly, an emerging view is that even for small molecules (i.e., <1000 Da), the sum of all compounds prepared to date fills a tiny fraction of the theoretical possibilities. A central finding of the results reported herein is that sizable combinatorial libraries of tetrapyrrole macrocycles occupy a region of chemical space that is rich in isomers. While this phenomenon also is expected for many other molecules including other polycyclic or macrocyclic aromatic hydrocarbons, porphyrins are likely distinguished by their facile synthesis. The existence, identity, and relative amounts of such porphyrin isomers in virtual libraries can readily be established through use of the *PorphyrinViLiGe* program. The program can be freely downloaded from the Web site www.photochemcad.com.

APPENDIX

Pólya's theorem enables enumeration of all possible substituted tetrapyrrole macrocycles given the number of available sites (e.g., eight β-pyrrole positions or four meso positions) and the number of reactants.[26] In addition, given the number of distinct substituent formulas concerning a particular macrocycle, the number of isomers can be determined. It warrants emphasis that the characteristic structure of the tetrapyrrole framework provides four pairs of β-substituents in a cyclic array wherein a given pair is separated from the adjacent pair by a meso site. In other words, a β-substituted tetrapyrrole macrocycle is not a regular octagon. Three illustrative examples of application of Pólya's theorem to tetrapyrrole combinatorial libraries are provided here.

Example 1. The derivatization of an 8-point tetrapyrrole scaffold (e.g., β-octa-substituted porphyrins, phthalocyanines) affords N possible products following eq A1[26,36,38]

$$N = 1/8(f_1^8 + 5f_2^4 + 2f_4^2) \quad \text{(A1)}$$

where f_i represent the number of different substituents. For example, upon β-substitution of a porphyrin with two ligands (A, B), the f_i expand as provided in eq A2

$$f_1 = A + B,\ f_2 = A^2 + B^2, \text{ and } f_4 = A^4 + B^4 \quad \text{(A2)}$$

Expansion of eq A1 taking into account eq A2 affords the polynomial eq A3

$$N = 1/8[(A + B)^8 + 5(A^2 + B^2)^4 + 2(A^4 + B^4)^2] \quad \text{(A3)}$$

which upon expansion affords the products described by eq A4

$$N = A^8 + A^7B + 6A^6B^2 + 7A^5B^3 + 13A^4B^4 + 7A^3B^5 + 6A^2B^6 + AB^7 + B^8 \quad \text{(A4)}$$

Here, we note that the mathematical terminology employs superscripts where chemistry employs subscripts. Rewriting eq A4 for chemists affords eq A5 (identical to eq 1 in the main text)

$$N = A_8 + A_7B + 6A_6B_2 + 7A_5B_3 + 13A_4B_4 + 7A_3B_5 + 6A_2B_6 + AB_7 + B_8 \quad (A5)$$

The resulting expansion (eq A5) describes three aspects about the product distribution. First, the number of terms in the polynomial expansion equals the *number of substituent combinations*. Thus, there are nine different patterns (i.e., substituent combinations) observed, ranging from A_8 to B_8. Second, each coefficient gives the *number of isomers* of that particular substituent pattern. Thus, the A_8 species is unique (i.e., no isomers), whereas there are 6 isomers of the A_6B_2 porphyrin, 7 isomers of the A_5B_3-porphyrin, 13 isomers of the A_4B_4-porphyrin, and 6 isomers of the A_2B_6-porphyrin. Third, the sum of all of the coefficients gives the total *number of products (including isomers)*, which in this case is 1 + 1 + 6 + 7 + 13 + 7 + 6 + 1 + 1 = 43. The latter result also can be obtained by reformulation of eq A1 as shown in eq A6

$$N = 1/8(n^8 + 5n^4 + 2n^2) \quad (A6)$$

where N = the number of products and n = the number of reagents (43 and 2, respectively, in the example given).

Example 2. The situation becomes more complex when, for example, an 8-point tetrapyrrole scaffold is substituted with eight ligands (A, B, C, D, E, F, G, and H). In this case, eq A1 is converted to give eq A7

$$N = 1/8[(A + B + C + D + E + F + G + H)^8 + 5(A^2 + B^2 + C^2 + D^2 + E^2 + F^2 + G^2 + H^2)^4 + 2(A^4 + B^4 + C^4 + D^4 + E^4 + F^4 + G^4 + H^4)^2] \quad (A7)$$

Expansion of the resulting polynomial above is available in a software module in *PorphyrinViLiGe*. Before the advent of powerful computers, Tapscott successfully assessed the number of distinct porphyrin patterns without expansion of the polynomial.[39] Tapscott identified 22 possible partitions of eight ligands on a tetrapyrrole macrocycle. The partitions range from [8], [7, 1], [6, 2], [6, 1, 1], ... [1, 1, 1, 1, 1, 1, 1, 1]. Application of this approach led to recognition that there are 5040 isomers for the partition of [1, 1, 1, 1, 1, 1, 1, 1]; in other words, an ABCDEFGH-porphyrin.[39] The Tapscott approach is a useful companion to that of Pólya but does not resolve any of the inherent shortcomings (vide infra).

Example 3. The case of meso-substituted porphyrins is substantially more tractable than that of β-substituted porphyrins. Indeed, a meso-substituted porphyrin has the symmetry of a square, for which the formula is provided in eq A8[28,30]

$$N = 1/8({f_1}^4 + 2{f_4}^1 + 3{f_2}^2 + 2{f_1}^2 * {f_2}^1) \quad (A8)$$

Expansion of this equation for three substituents (A, B, C) in the same manner leads to the result shown in eq A9

$$N = A_4 + A_3B + A_3C + 2A_2B_2 + 2A_2BC + 2A_2C_2 + AB_3 + 2AB_2C + 2ABC_2 + AC_3 + B_4 + B_3C + 2B_2C_2 + BC_3 + C_4 \quad (A9)$$

Equation A9 describes the distribution of meso-substituted porphyrins where there are three types of substituents (A, B, and C) and can be interpreted following the same manner as described for eq A5.

In summary, Pólya's theorem can be applied to good effect to a number of combinatorial problems concerning both β-substituted and meso-substituted tetrapyrrole macrocycles. Pólya's theorem is effective in affording (i) the total number of products, (ii) the number of distinct substituent patterns, and (iii) the number of isomers of each substituent pattern. On the other hand, the powerful mathematical formalism of Pólya has significant shortcomings, and was never intended, for describing combinatorial libraries presented herein. The shortcomings include the following:

(i) The enumeration does not delineate the ratio of various products.
(ii) The enumeration does not accommodate nonequal ratios of reactants.
(iii) The enumeration does not identify the structure of any isomers.
(iv) The enumeration is not suited to describe the library of products wherein the groupings of substituents is restricted by the nature of the reaction. A key example is provided by the reaction of an aminoketone chemset and a dione chemset to form a pyrrole chemset, the members of which then self-react to form a porphyrin library (see text). Pólya's theorem describes the library built around a target molecular scaffold and does not easily accommodate the constraints that can accrue by reactions leading to such scaffolds.
(v) There are no provisions for data mining of the reaction products.

■ ASSOCIATED CONTENT

Supporting Information. Expanded description of several features of the *PorphyrinViLiGe* program including the internal representation of molecular species, the methods for identification of distinct products, and the frequency of the tetrapyrrole macrocycles. Also, the 165 substituent combinations are described upon 8-point derivatization of a phthalocyanine with four reagents. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: jlindsey@ncsu.edu.

■ ACKNOWLEDGMENT

This work was supported by a grant from the NSF Chemistry of Life Processes Program (NSF CHE-0953010).

■ REFERENCES

(1) Lindsey, J. S.; Ptaszek, M.; Taniguchi, M. Simple Formation of an Abiotic Porphyrinogen in Aqueous Solution. *Origins Life Evol. Biospheres* **2009**, *39*, 495–515.
(2) Lindsey, J. S.; Chandrashaker, V.; Taniguchi, M.; Ptaszek, M. Abiotic Formation of Uroporphyrinogen and Coproporphyrinogen from Acyclic Reactants. *New J. Chem.* **2011**, *35*, 65–75.
(3) Mauzerall, D. The Thermodynamic Stability of Porphyrinogens. *J. Am. Chem. Soc.* **1960**, *82*, 2601–2605.
(4) Mauzerall, D. The Condensation of Porphobilinogen to Uroporphyrinogen. *J. Am. Chem. Soc.* **1960**, *82*, 2605–2609.

(5) Mauzerall, D. Chlorophyll and Photosynthesis. *Philos. Trans. R. Soc., B* **1976**, *273*, 287–294.
(6) Layer, G.; Jahn, D.; Deery, E.; Lawrence, A. D.; Warren, M. J. Biosynthesis of Heme and Vitamin B_{12}. In *Comprehensive Natural Products II Chemistry and Biology*; Mander, L., Lui, H.-W., Eds.; Elsevier: Oxford, 2010; Vol. 7, pp 445−499.
(7) Lindsey, J. S. Synthesis of *meso*-Substituted Porphyrins. In *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; Academic Press: San Diego, CA, 2000; Vol. 1, pp 45−118.
(8) Smith, K. M. Strategies for the Synthesis of Octaalkylporphyrin Systems. In *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; Academic Press: San Diego, CA, 2000; Vol. 1, pp 1−43.
(9) Vicente, M. G. H.; Smith, K. M. Porphyrins and Derivatives: Synthetic Strategies and Reactivity Profiles. *Curr. Org. Chem.* **2000**, *4*, 139–174.
(10) Senge, M. O. Nucleophilic Substitution as a Tool for the Synthesis of Unsymmetrical Porphyrins. *Acc. Chem. Res.* **2005**, *38*, 733–743.
(11) Lindsey, J. S. Synthetic Routes to *Meso*-Patterned Porphyrins. *Acc. Chem. Res.* **2010**, *43*, 300–311.
(12) Senge, M. O.; Shaker, Y. M.; Pintea, M.; Ryppa, C.; Hatscher, S. S.; Ryan, A.; Sergeeva, Y. Synthesis of *meso*-Substituted ABCD-Type Porphyrins by Functionalization Reactions. *Eur. J. Org. Chem.* **2010**, 237–258.
(13) Ono, N.; Yamada, H.; Okujima, T. Synthesis of Porphyrins Fused with Aromatic Rings. In *Handbook of Porphyrin Science*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; World Scientific: Singapore, 2010; Vol. 2, pp 1−102.
(14) Cavaleiro, J. A. S.; Tomé, A. C.; Neves, M. G. P. M. S. *meso*-Tetraarylporphyrin Derivatives: New Synthetic Methodologies. In *Handbook of Porphyrin Science*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; World Scientific: Singapore, 2010; Vol. 2, pp 193−294.
(15) Senge, M. O. Stirring the Porphyrin Alphabet Soup—Functionalization Reactions for Porphyrins. *Chem. Commun.* **2011**, *47*, 1943–1960.
(16) Chambron, J.-C.; Heitz, V.; Sauvage, J.-P. Noncovalent Multiporphyrin Assemblies. In *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; Academic Press: San Diego, CA, 2000; Vol. 6, pp 1−42.
(17) Balaban, T. S.; Tamiaki, H.; Holzwarth, A. R. Chlorins Programmed for Self-Assembly. *Top. Curr. Chem.* **2005**, *258*, 1–38.
(18) Medforth, C. J.; Wang, Z.; Martin, K. E.; Song, Y.; Jacobsen, J. L.; Shelnutt, J. A. Self-assembled Porphyrin Nanostructures. *Chem. Commun.* **2009**, 7261–7277.
(19) Balaban, T. S. Self-Assembling Porphyrins and Chlorins as Synthetic Mimics of the Chlorosomal Bacteriochlorophylls. In *Handbook of Porphyrin Science*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; World Scientific: Singapore, 2010; Vol. 1, pp 221−306.
(20) Senge, M. O. Database of Tetrapyrrole Crystal Structure Determinations. In *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; Academic Press: San Diego, CA, 2000; Vol. 10, pp 1−218.
(21) Engel, M. K. Single-Crystal Structures of Phthalocyanine Complexes and Related Macrocycles. In *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; Elsevier Science: San Diego, CA, 2003; Vol. 20, pp 1−242.
(22) Gugisch, R.; Kerber, A.; Laue, R.; Meringer, M.; Weidinger, J. MOLGEN-COMB, A Software Package for Combinatorial Chemistry. *MATCH-Comm. Math. Comp. Chem.* **2000**, *41*, 189–203.
(23) Schüller, A.; Hähnke, V.; Schneider, G. SmiLib v2.0: A Java-Based Tool for Rapid Combinatorial Library Enumeration. *QSAR Comb. Sci.* **2007**, *26*, 407–410.
(24) Trinajstić, N.; Jeričević, Ž.; Knop, J. V.; Müller, W. R.; Szymanski, K. Computer Generation of Isomeric Structures. *Pure Appl. Chem.* **1983**, *55*, 379–390.
(25) Pilgrim, R. L. C. The Number of Possible Isomers in the Porphyrins. *J. Chem. Educ.* **1974**, *51*, 316–318.
(26) Pólya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*; Springer-Verlag: New York, 1987.
(27) Rouvray, D. H. Isomer Enumeration Methods. *Chem. Soc. Rev.* **1974**, *3*, 355–372.
(28) Tucker, A. Polya's Enumeration Formula by Example. *Math. Mag.* **1974**, *47*, 248–256.
(29) Balaban, A. T. Applications of Graph Theory in Chemisty. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334–343.
(30) Nemba, R. M.; Ngouhouo, F. On the Enumeration of Chiral and Achiral Skeletons of Position Isomers of Homosubstituted Monocyclic Cycloalkanes with a Ring Size n (Odd or Even). *Tetrahedron* **1994**, *50*, 6663–6670.
(31) Shao, Y.; Wu, J.; Jiang, Y. Enumeration and Symmetry of Substitution Isomers. *J. Phys. Chem.* **1996**, *100*, 15064–15067.
(32) Wieland, T. Mathematical Simulations in Combinatorial Chemistry. *MATCH-Comm. Math. Comp. Chem.* **1996**, *34*, 179–206.
(33) Shao, Y.; Wu, J.; Jiang, Y. Isomer Counting and Isomer Permutation Representation. *Chem. Phys. Lett.* **1996**, *248*, 366–372.
(34) Fujita, S. Benzene Derivatives with Achiral and Chiral Substituents and Relevant Derivatives Derived from D_{6h} Skeletons. Symmetry-Itemized Enumeration and Symmetry Characterization by the Unit-Subduced Cycle Index Approach. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 151–163.
(35) Faulon, J.-L.; Visco, D. P., Jr.; Roe, D. Enumerating Molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K., Ed.; Wiley-VCH: 2005; Vol. 21, pp 209−286.
(36) Balaban, A. T. Chemical Graphs. XXV Isomers of the Porphyrins—An Exercise Using Pólya's Counting Theorem. *Rev. Roum. Chim.* **1975**, *20*, 227–232.
(37) Aronoff, S. The Number of Biologically Possible Porphyrin Isomers. *Ann. N.Y. Acad. Sci.* **1975**, *244*, 327–333.
(38) Alspach, B.; Aronoff, S. Enumeration of Structural Isomers in Alicyclic Hydrocarbons and in Porphyrins. *Can. J. Chem.* **1977**, *55*, 2773–2777.
(39) Tapscott, R. E.; Marcovich, D. Enumeration of Permutational Isomers: The Porphyrins. *J. Chem. Educ.* **1978**, *55*, 446–447.
(40) Berlin, K.; Jain, R. K.; Tetzlaff, C.; Steinbeck, C.; Richert, C. Spectrometrically Monitored Selection Experiments: Quantitative Laser Desorption Mass Spectrometry of Small Chemical Libraries. *Chem. Biol.* **1997**, *4*, 63–77. Erratum *Chem. Biol.* **1997**, *4*, 237–238.
(41) 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.
(42) 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.
(43) Elgie, K. J.; Scobie, M.; Boyle, R. W. Application of Combinatorial Techniques in the Synthesis of Unsymmetrically Substituted 5,15-Diphenylporphyrins. *Tetrahedron Lett.* **2000**, *41*, 2753–2757.
(44) Steinbeck, C.; Berlin, K.; Richert, C. MASP—A Program Predicting Mass Spectra of Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 449–457.
(45) Dombi, K. L.; Richert, C. Relative Reactivities of Activated Carboxylic Acids in Amide-Forming Reactions Employed for the Synthesis of Tetraphenylporphyrin Libraries. *Molecules* **2000**, *5*, 1265–1280.
(46) Shi, B.; Scobie, M.; Boyle, R. W. Parallel Synthesis of Unsymmetrically Substituted Tetraphenyl Porphyrins on Wang Resin. *Tetrahedron Lett.* **2003**, *44*, 5083–5086.
(47) Samaroo, D.; Vinodu, M.; Chen, X.; Drain, C. M. *meso*-Tetra(pentafluorophenyl)porphyrin as an Efficient Platform for Combinatorial Synthesis and the Selection of New Photodynamic Therapeutics using a Cancer Cell Line. *J. Comb. Chem.* **2007**, *9*, 998–1011.
(48) Drain, C. M.; Singh, S. Combinatorial Libraries of Porphyrins: Chemistry and Applications. In *Handbook of Porphyrin Science*; Kadish, K. M., Smith, K. M., Guilard, R., Eds.; World Scientific: Singapore, 2010; Vol. 3, pp 485−537.

(49) McKeown, N. B.; Chambrier, I.; Cook, M. J. Synthesis and Characterization of Some 1,4,8,11,15,18,22,25-Octa-alkyl- and 1,4,8,11,15,18-Hexa-alkyl-22,25-bis(carboxpropyl)phthalocyanines. *J. Chem. Soc., Perkin Trans. 1* **1990**, 1169–1177.

(50) Linssen, T. G.; Hanack, M. Synthesis, Separation and Characterization of Unsymmetrically Substituted Phthalocyanines. *Chem. Ber.* **1994**, *127*, 2051–2057.

(51) de la Torre, G.; Claessens, C. G.; Torres, T. Phthalocyanines: The Need for Selective Synthetic Approaches. *Eur. J. Org. Chem.* **2000**, 2821–2830.

(52) McKeown, N. B. Phthalocyanine-Containing Polymers. *J. Mater. Chem.* **2000**, *10*, 1979–1995.

(53) de la Torre, G.; Claessens, C. G.; Torres, T. Phthalocyanines: Old Dyes, New Materials. Putting Color in Nanotechnology. *Chem. Commun.* **2007**, 2000–2015.

(54) Dumoulin, F.; Durmus, M.; Ahsen, V.; Nyokong, T. Synthetic Pathways to Water-soluble Phthalocyanines and Close Analogs. *Coord. Chem. Rev.* **2010**, *254*, 2792–2847.

(55) Nemykin, V. N.; Lukyanets, E. A. Synthesis of Substituted Phthalocyanines. *ARKIVOC* **2010**, *i*, 136–208.

(56) Rodríguez-Morgade, M. S.; Stuzhin, P. A. The Chemistry of Porphyrazines. An Overview. *J. Porphyrins Phthalocyanines* **2004**, *8*, 1129–1165.

(57) Fuchter, M. J.; Zhong, C.; Zong, H.; Hoffman, B. M.; Barrett, A. G. M. Porphyrazines: Designer Macrocycles by Peripheral Substituent Change. *Aust. J. Chem.* **2008**, *61*, 235–255.

(58) Dumoulin, F.; Zorlu, Y.; Ayhan, M. M.; Hirel, C.; Isci, U.; Ahsen, V. A First ABAC Phthalocyanine. *J. Porphyrins Phthalocyanines* **2009**, *13*, 161–165.

(59) Chen, X.; Thomas, J.; Gangopadhyay, P.; Norwood, R. A.; Peyghambarian, N.; McGrath, D. V. Modification of Symmetrically Substituted Phthalocyanines Using Click Chemistry: Phthalocyanine Nanostructures by Nanoimprint Lithography. *J. Am. Chem. Soc.* **2009**, *131*, 13840–13843.

(60) Juricek, M.; Kouwer, P. H. J.; Rehák, J.; Sly, J.; Rowan, A. E. A Novel Modular Approach to Triazole-Functionalized Phthalocyanines Using Click Chemistry. *J. Org. Chem.* **2009**, *74*, 21–25.

(61) Berthold, H. J.; Franke, S.; Thiem, J.; Schotten, T. Ex Post Glycoconjugation of Phthalocyanines. *J. Org. Chem.* **2010**, *75*, 3859–3862.

(62) Berthold, H. J.; Schotten, T.; Hoffmann, F.; Thiem, J. A Highly Versatile Octasubstituted Phthalocyanine Scaffold for *ex post* Chemical Diversification. *Synthesis* **2010**, 741–748.

(63) 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.

(64) Maclean, D.; Martin, E. J. On the Representation of Combinatorial Libraries. *J. Comb. Chem.* **2004**, *6*, 1–11.

(65) Senge, M. O. Extroverted Confusion—Linus Pauling, Melvin Calvin, and Porphyrin Isomers. *Angew. Chem., Int. Ed.* **2011**, *50*, 4272–4277.

(66) Barnard, J. M.; Downs, G. M.; von Scholley-Pfab, A.; Brown, R. D. Use of Markush Structure Analysis Techniques for Descriptor Generation and Clustering of Large Combinatorial Libraries. *J. Mol. Graphics Modell.* **2000**, *18*, 452–463.

(67) Staley, S. W.; Eliasson, B.; Kearney, P. C.; Schreiman, I. C.; Lindsey, J. S. Synthesis and Electron-Transfer Studies of Conjugated π-Electron Systems. Model Systems for Molecular Computation Studies. In *Molecular Electronic Devices*; 3rd ed.; Carter, F. L., Siatkowski, R. E., Wohltjen, H., Eds.; Elsevier: North Holland, 1988; pp 543—554.

(68) Oprea, T. I. Rapid Estimation of Hydrophobicity for Virtual Combinatorial Library Analysis. *SAR QSAR Environ. Res.* **2001**, *12*, 129–141.

(69) Charifson, P. S.; Walters, W. P. Filtering Databases and Chemical Libraries. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 311–323.

(70) Barbosa, F.; Horvath, D. Molecular Similarity and Property Similarity. *Curr. Top. Med. Chem.* **2004**, *4*, 589–600.

(71) Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216.

(72) Maggiora, G. M. On Outliers and Activity Cliffs—Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.

(73) Sisay, M. T.; Peltason, L.; Bajorath, J. Structural Interpretation of Activity Cliffs Revealed by Systematic Analysis of Structure—Activity Relationships in Analog Series. *J. Chem. Inf. Model.* **2009**, *49*, 2179–2189.

(74) http://www.sisweb.com/referenc/source/exactmaa.htm (accessed May 20, 2011).

(75) Palmblad, M.; Drijfhout, J. W.; Deelder, A. M. High Resolution Mass Spectrometry for Rapid Characterization of Combinatorial Peptide Libraries. *J. Comb. Chem.* **2010**, *12*, 65–68.

(76) Falk, J. E. Chromatography of Porphyrins and Metalloporphyrins. *J. Chromatogr.* **1961**, *5*, 277–299.

(77) Lim, C. K.; Li, F.; Peters, T. J. High-Performance Liquid Chromatography of Porphyrins. *J. Chromatogr.* **1988**, *429*, 123–153.

(78) Grishina, L. E.; Brykina, G. D.; Shpigun, O. A. High-Performance Liquid Chromatography of Porphyrins: Ligands and Metal Complexes. *J. Anal. Chem.* **1995**, *50*, 826–835.

(79) Danton, M.; Lim, C. K. Porphyrin Profiles in Blood, Urine and Faeces by HPLC/Electrospray Ionization Tandem Mass Spectrometry. *Biomed. Chromatogr.* **2006**, *20*, 612–621.

(80) Jiang, Z.; He, W.; Yao, H.; Wang, J.; Chen, N.; Huang, J. Isomeric Separation and Identification of Tetra-, Tri-, and Di-β-sulphonic Phthalocyanine Zinc Complexes. *J. Porphyrins Phthalocyanines* **2011**, *15*, 140–148.

(81) *Pyrroles Part One: The Synthesis and the Physical and Chemical Aspects of the Pyrrole Ring*; Jones, A. R., Ed.; John Wiley & Sons: New York, 1990.

(82) *Pyrroles Part Two: The Synthesis, Reactivity, and Physical Properties of Substituted Pyrroles*; Jones, A. R., Ed.; John Wiley & Sons: New York, 1992.

(83) Strocker, A. M.; Keating, T. A.; Tempest, P. A.; Armstrong, R. W. Use of a Convertible Isocyanide for Generation of Ugi Reaction Derivatives on Solid Support: Synthesis of α-Acylaminoesters and Pyrroles. *Tetrahedron Lett.* **1996**, *37*, 1149–1152.

(84) Mjalli, A. M. M.; Sarshar, S.; Baiga, T. J. Solid Phase Synthesis of Pyrroles Derived from a Four Component Condensation. *Tetrahedron Lett.* **1996**, *37*, 2943–2946.

(85) Hansford, K. A.; Zanzarova, V.; Dörr, A.; Lubell, W. D. Three-Step Solution-Phase Combinatorial Access to 1,2-Disubstituted and 1,2,5-Trisubstituted Pyrroles from Carboxylic Esters. *J. Comb. Chem.* **2004**, *6*, 893–898.

(86) Bianchi, I.; Forlani, R.; Minetto, G.; Peretto, I.; Regalia, N.; Taddei, M.; Raveglia, L. F. Solution Phase Synthesis of a Library of Tetrasubstituted Pyrrole Amides. *J. Comb. Chem.* **2006**, *8*, 491–499.

(87) Bremner, W. S.; Organ, M. G. Formation of Substituted Pyrroles via an Imine Condensation/Aza-Claisen Rearrangement/Imine—Allene Cyclization Process by MAOS. *J. Comb. Chem.* **2008**, *10*, 142–147.

(88) 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.

(89) 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.

(90) 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.

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

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

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

(94) Gorse, A.-D. Diversity in Medicinal Chemistry Space. *Curr. Top. Med. Chem.* **2006**, *6*, 3–18.
(95) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952–2963.
(96) Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732–8733.
(97) 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.
(98) Varnek, A.; Baskin, I. I. Cheminformatics as a Theoretical Chemistry Discipline. *Mol. Inf.* **2011**, *30*, 20–32.