Configurational Analysis of Tetracyclic Dimeric Pyrrole−Imidazole Alkaloids Using a Floating Chirality Approach

Matthias Köck,*,[†] Gesine Schmidt,[†] Ian B. Seiple,[‡] and Phil S. Baran[‡]

[†]Alfred-Wegener-Institut für Polar- und Meeresforschung in der Helmholtz-Gemeinschaft, Am Handelshafen 12, D-27570 Bremerhaven, Germany

[‡]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

ABSTRACT: The structure elucidation of the palau'amine congener tetrabromostyloguanidine (**1**), which used interproton distances from ROESY spectra as restraints in a computational approach, the so-called fc-rDG/DDD method, led to a revision of the relative configuration of palau'amine (**2**) and its congeners in 2007. The recent total synthesis of (±)-palau'amine (**2**) subsequently confirmed the computed structural revision of the relative configuration. In order to test a broader application range of the fc-rDG/DDD method, the present study investigated two additional dimeric pyrrole−imidazole alkaloids, axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**). These calculations allowed the simultaneous assignment of the relative configuration for all eight stereogenic centers of compounds **3** and **4** without using any information from the reported configurations. In contrast to the palau'amine congeners, the fc-rDG/DDD method confirmed the relative configuration originally described for axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**).
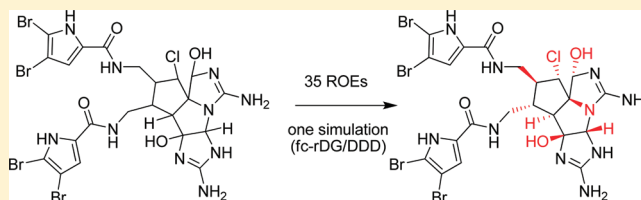
The configurational assignment of natural products (relative and absolute configuration) is essential to understand their biological activity on a molecular level and to allow their procurement through total synthesis. When crystalline products cannot be obtained, structural elucidations can be very difficult, especially when the targets do not lend themselves to standard 2D NMR techniques and may not readily be synthesized. Herein, we discuss how NOESY/ROESY data in combination with computational methods can be applied to the assignment of relative configurations of natural products. Traditionally, the NOE/ROE effects have been used qualitatively or as restraints in energy minimization (EM) and molecular dynamics (MD) simulations. The application of both approaches is problematic with a large number of unknown stereogenic centers. EM/MD methods, for example, require one calculation for each possible relative configuration, resulting in extremely long calculation times. The structural elucidation of amorphous molecules with several unknown stereogenic centers would benefit greatly from a method that could simultaneously analyze all configurations. Distance geometry (DG)[1] in combination with distance-bounds-driven dynamics (DDD)[2] calculations using interproton distances and floating chirality[3] is discussed herein for the determination of the relative configuration of natural products (the method is called "floating chirality restrained DG/DDD" or "fc-rDG/DDD"). This method has been applied to small organic molecules since the 1990s.[4]

One of the most investigated complex natural products over the past decade was palau'amine (**2**). The fc-rDG/DDD method was successfully applied in 2007 to the palau'amine congener tetrabromostyloguanidine (**1**),[5] leading to a revision of the relative configuration of palau'amine (**2**) and its congeners.[6] The re-assignment of the relative configuration was confirmed by the total synthesis of palau'amine (**2**) by Baran and co-workers in 2010.[7]

The originally reported structure of palau'amine (with the wrong relative configuration) served as a synthetic target for several groups worldwide for well over a decade.[8]

In order to learn more about the scope and the broader application of the fc-rDG/DDD method, the studies were extended to two additional dimeric pyrrole−imidazole alkaloids, axinellamine A (**3**)[9] and 3,7-*epi*-massadine chloride (**4**).[10] Axinellamine A (**3**) was first reported in 1999 with a relative configuration established by NMR spectroscopy,[9a] and its total synthesis was achieved in 2008.[9b,c] 3,7-*epi*-Massadine chloride (**4**) is a side product in the synthesis of massadine[10c] and has not yet been found in natural isolates, but it is expected to be a natural product. Axinellamines A and B are two natural products that are epimeric at C-5 and C-9 (**3**, Scheme 1), the positions that correspond to C-3 and C-7 in the massadines (**4**, Scheme 1). However, 3,7-*epi*-massadine analogues, which were synthesized by biomimetic cyclization in the Baran laboratory,[10c] have not been identified in sponge extracts. We believe it is likely that the *epi*-massadine series is actually present in Nature but has yet to be extracted and characterized.

Unlike the palau'amine congeners, axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**) possess tetracyclic cores, but still contain eight contiguous stereogenic centers, resulting in 128 possible relative configurations (diastereomers) when one center is fixed (which is relevant for the applied NMR studies). In this article we report the application of the fc-rDG/DDD method to assign the relative configuration of all eight stereogenic centers simultaneously. There are several arguments why

Scheme 1. Structural Formulas of Tetrabromostyloguanidine (1), Palau'amine (2), Axinellamine A (3), and 3,7-*epi*-Massadine Chloride (4)[a]

Tetrabromostyloguanidine (1)

Palau'amine (2)

Axinellamine A (3)

3,7-*epi*-Massadine chloride (4)
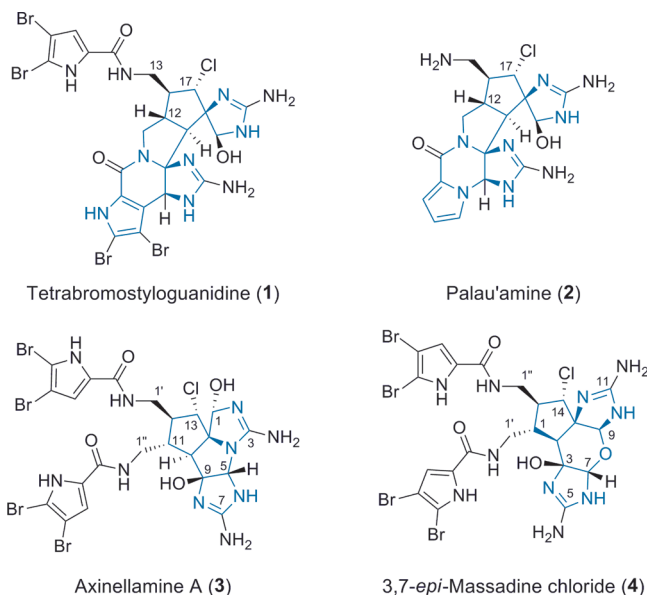
[a]The hexacyclic cores of **1** and **2** and the tetracyclic cores of **3** and **4** are indicated in blue. The configurations of C-12 and C-17 were revised for the palau'amine congeners (here **1** and **2**) in 2007. All structures herein are shown in their neutral forms.

the dimeric pyrrole—imidazole alkaloids **1** to **4** are ideal for a conformational and configurational analysis by NMR spectroscopy: (a) they are conformationally restricted molecules, (b) many of the protons are methine protons, which do not need any correction for pseudoatoms,[11] such as methyl groups or diastereotopically nonassigned methylene groups, and (c) the stereogenic centers are all in close proximity to each other.

■ RESULTS AND DISCUSSION

The fc-rDG/DDD method was applied to axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**). For both compounds more than 30 ROE restraints were used. All interproton distances can be used without correction factors since almost exclusively methine protons are present in these structures. Diastereotopically nonassigned methylene groups, methyl groups, nonassigned dimethyl groups, and aromatic *ortho/meta* protons have no defined proton positions and therefore require pseudoatom correction, leading to larger ranges of the distance restraints. Here we used ±10% for the conversion of the ROE-derived distances into distance restraints (upper/lower bounds) in order to account for experimental errors.

For axinellamine A (**3**), 35 interproton distances were obtained from ROESY spectra. These were used as experimental distance restraints for the fc-rDG/DDD calculations. In principle the DG/DDD approach can generate all 256 possible stereoisomers for **3**. Since NMR spectroscopy in achiral or racemic media cannot distinguish between enantiomers, one stereogenic center was set as a reference. This entails the possibility to generate 128 diastereomers. Results for the 99 generated possible structures of axinellamine A (**3**) are shown in Figure 1 as a graphical representation of the total error (dimensionless) for each structure, ordered according to ascending total errors.

Altogether, only 4 out of 128 possible diastereomers of **3** resulted from several runs with the fc-DG/DDD method using
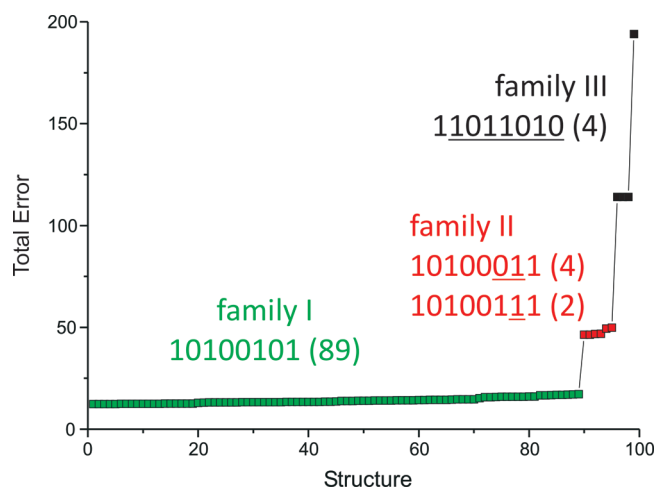

Figure 1. Results of the fc-rDG/DDD calculations of axinellamine A (**3**). Only 4 out of 128 possible relative configurations (diastereomers) for **3** were generated (the frequency is given in parentheses); 89 out of 99 generated structures have the same relative configuration (in green), which is identical to the one published in 1999. Underlined numbers within the binary code indicate a sign inversion of the chiral volume of the stereogenic center.

the experimental ROE restraints. The remaining 124 diastereomers fulfilled the restraints so poorly that they were not generated in any of the calculations. The 89 structures with the lowest total errors had the same relative configuration. As demonstrated recently for tetrabromostyloguanidine (**1**), the structures were clustered into different families that comprise structures with similar errors, but potentially different conformation and configuration.[5] Here, the families consisted of structures with different relative configurations (compare Figures 1 and 4). For axinellamine A (**3**), the 99 structures were divided into three structural families (I, II, and III; see Figure 1) according to their total errors. A binary code was used for the description of the configuration in order to simplify the distinction between the different diastereomers. The order within the binary code is identical to the sequence of the stereogenic centers (atom numbering). Family I (structures 1 to 89) exhibited the lowest total errors, with values between 15 and 20. An analysis of all 89 structures of family I indicated the same relative configuration as published for **3**.[9a] Family II consists of six structures (two different relative configurations) with total errors around 50. Family III consists of the final four structures with total errors between 100 and 200. The 89 superimposed structures of family I (the structures with the lowest total errors) are shown in Figure 2. These results supported the originally proposed structure[9a] that was confirmed in 2008 by the total synthesis[9b,c] of axinellamine A (**3**).

In order to understand the influence of different upper and lower boundaries and different mixing times, several fc-rDG/DDD calculations were carried out for **3**. The use of three different upper/lower boundaries (±10%, ± 20%, and ±30%) for the distance restraints of each mixing time of **3** resulted in a total of nine calculations. The results are summarized in Figure 3. For axinellamine A (**3**), at least 90% of the structures generated with upper/lower boundaries of ±10% or ±20% (Figure 3, bars 1, 2, 4, 5, 7, and 8) and even 70—85% (Figure 3, bars 3, 6, and 9) of the structures generated with 30% upper/lower boundary limits yielded the correct relative configuration. Since 30% boundary limits are very loose restrictions, these results indicate unambiguously the

Figure 2. Superimposed structures of family I (left, see Figure 1) of axinellamine A (**3**) and 89 substructures (14 atoms of the tetracyclic ring system) of axinellamine A (**3**) (right). Carbons are represented in gray, hydrogens in white, nitrogens in blue, oxygens in red, chlorine in green, and bromines in purple.
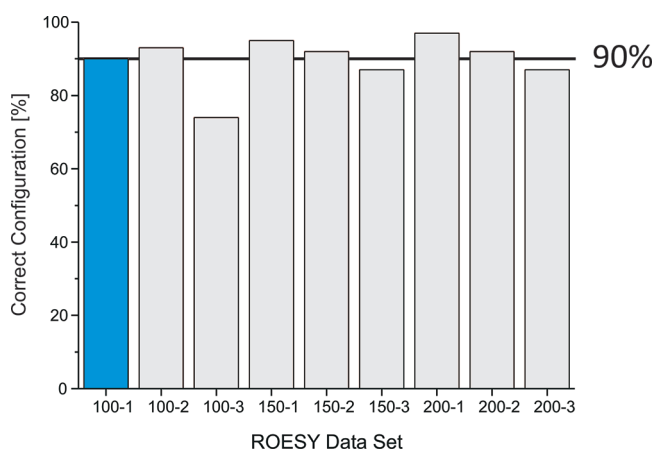

Figure 3. Summary of the results of the nine fc-rDG/DDD calculations of axinellamine A (**3**). The percentage of the correct relative configuration for all nine data sets is indicated by the bars. The abbreviation "100-1" stands for a mixing time of 100 ms and upper/lower boundaries of ±10%, "100-2" for 100 ms and ±20%, and so on. The blue bar indicates the calculation corresponding to the results shown in Figures 1 and 2.

strength of the fc-rDG/DDD method for these examples of conformationally restricted molecules.

For 3,7-*epi*-massadine chloride (**4**), 36 interproton distances were obtained from a ROESY spectrum.[12] In this case only one ROESY spectrum with a mixing time of 300 ms was measured. The 36 interproton distances were used as experimental distance restraints for the fc-rDG/DDD calculations. The results for the 98 generated structures are shown in Figure 4. Out of the 128 possible relative configurations (diastereomers) again only four were generated; 74 out of 98 generated structures had the same relative configuration. In contrast to axinellamine A (**3**) the correct configuration of 3,7-*epi*-massadine chloride (**4**) can be found in families I and II. Out of the 19 structures in family II of 3,7-*epi*-massadine chloride (**4**), 15 had the same relative configuration (11111110) as family I and four differed only in the configuration of one stereogenic center (1111111<u>1</u>; see Figure 4). The fc-rDG/DDD calculations for 3,7-*epi*-massadine chloride (**4**) were run with upper/lower bounds for the distance restraints of ±10% (see Figure 4), ± 20%, and ±30%.

The results obtained with the fc-rDG/DDD method for axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**) demonstrated the power of this approach for the structural assignment of densely functionalized small molecules. The fc-rDG/DDD method

allowed the simultaneous assignment of the relative configuration of all eight stereogenic centers in axinellamine A (**3**) and 3,7-*epi*-massadine chloride (**4**). It was shown that even with very loose restraints (up to 30%), these molecules could be assigned a relative
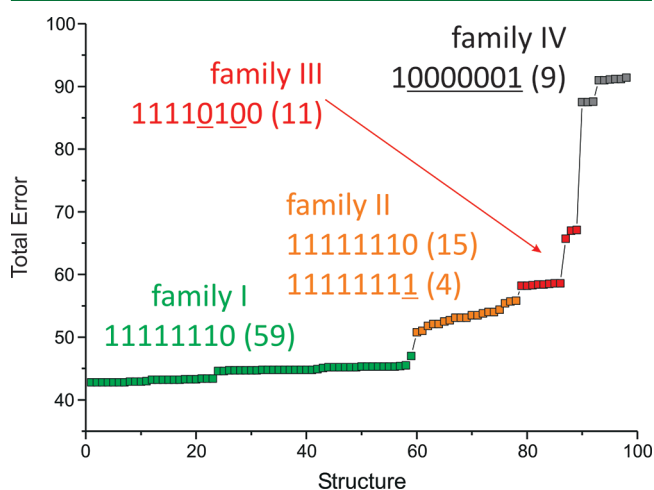

Figure 4. Results of the fc-rDG/DDD calculations of 3,7-*epi*-massadine chloride (**4**). Only 4 out of 128 possible relative configurations (diastereomers) for **4** are generated; 74 out of 98 generated structures have the same relative configuration (59 in green, 15 in orange), which is identical to the one published in 2008. Underlined numbers within the binary code indicate a sign inversion of the chiral volume of the stereogenic center.

configuration in a single calculation without necessitating prior knowledge of the relative configuration at any of the eight contiguous centers. This method represents a very useful tool to be combined with qualitative NMR analysis for isolation and synthesis chemists alike. It has already enabled the reassignment of very noteworthy natural products such as palau'amine (**2**) [via tetrabromostyloguanidine (**1**)]. Its application to the assignment of natural products could prevent future misassignments of natural products, including those with more flexible architectures.

■ EXPERIMENTAL SECTION

For the so-called NOE/ROE-restrained calculations the cross-peaks of NOESY or ROESY spectra are translated into interproton distances, which are used as restraints for conformational and configurational analysis. The usage of a NOESY or a ROESY spectrum is mainly dependent on the size of the molecule. NOESY spectra have the disadvantage that for medium-sized molecules the NOE becomes 0 if $(\omega_0 \tau_c) = 1.12$. This is approximately the case for a correlation time τ_c of 0.3 ns at 400 MHz. ROESY spectra do not have this disadvantage, but because of the spinlock, other effects (e.g., offset dependence) have to be considered. A quantitative ROESY analysis was performed for compounds **3** and **4**. Interproton distances from ROESY spectra with different mixing times (100, 150, and 200 ms) of axinellamine A (**3**) were used for the configurational assignment. For 3,7-*epi*-massadine chloride (**4**) only one ROESY spectrum with a mixing time of 300 ms was used. The ROESY spectra were calibrated and baseline corrected with the Bruker Topspin 2.0 software and integrated with the Sparky 3.1 software. In Sparky, the contour levels were adjusted to display peaks just below the baseline but exclude the baseline. Integration was performed according to the scheme (peak picking−peak assignment−peak locking−peak integration). Multiplet peaks were grouped into one assigned peak. Where possible, the Gaussian fit was used for integration and the line widths were adjusted to fit the displayed peak contour. Peaks that overlaid the diagonal were integrated by the sum-overbox or sum-overellipse fit and labeled to be used carefully during distance calculations. Subsequently, the integrals of

corresponding cross-peaks were checked manually for compliance and rated accordingly for the distance calculation program.

After volume integration of all cross-peaks in the ROESY spectra, the intensity data were calibrated setting the distance between the geminal protons at C-1′ (for **3**) and C-1″ (for **4**) to 178 pm. These positions correspond to the same methylene unit used to calibrate **1** (C-13). For axinellamine A (**3**) each ROESY spectrum was analyzed separately under the assumption of the linear approximation approach. This assumption is valid because the distances for the different mixing times were identical. Since all molecules studied here (including **1**) have more or less the same molecular weight, it is sufficient to use only one ROESY spectrum with the previously mentioned mixing time of 100 to 200 ms (max. 300 ms). The latter one was used for 3,7-*epi*-massadine chloride (**4**).

The distance geometry calculations were carried out using a modified version of the DISGEO program kindly provided by Prof. Ruud Scheek (University of Groningen, The Netherlands). The list of distance bounds for the DG calculation was constructed from the covalent primary structure using bond lengths, bond angles, chirality, and planarity. The upper and lower bounds were taken as ±2%, respectively, of these calculated distances. For atoms that have a rotatable torsion between them, the lower and upper bound are calculated from the minimum and maximum possible distance, respectively, with a complete rotation of the dihedral. In addition to these geometric distance bounds, the experimental interproton distances, derived from 35 ROEs for **3** (100 ms) and 36 ROEs for **4**, were utilized. The upper and lower distance restraints were calculated with ±10%, respectively, of the actual distances obtained from ROESY spectra.

Distances within these constraints that also satisfy the triangular inequalities were chosen randomly and embedded with the metric matrix method (random metrization). Around 100 structures were successfully embedded (usually 98 to 100) and were partially minimized with a conjugate gradient minimizer (Optimize). This was followed by a distance-bounds-driven dynamics (DDD) calculation beginning at 500 K for 5 ps (2500 steps, step size 20 fs) and then with a gradual reduction in temperature over the next 2 ps. Both Optimize and DDD utilized only the holonomic constraints (i.e., geometric and experimental distance constraints) as the potential and for generation of forces. The structures are ranked by their total errors (pseudoenergies). The pseudoenergy is not a real energy; it is a quality factor that describes the degree of satisfaction of the given distances and the given chiral volumes with the actual values. It is without any dimension. The distance term consists of experimental and holonomic restraints. The latter ones are defined by the constitution of the molecule. The calculation and optimization of 100 structures required less than 5 min of CPU time using a single processor PC.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +49-471-4831-1497. Fax: +49-471-4831-1425. E-mail: mkoeck@awi.de.

■ REFERENCES

(1) (a) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *Bull. Math. Biol.* **1983**, *45*, 665−720. (b) Havel, T. F. *DISGEO*, Quantum Chemistry Exchange Program Exchange No. 507, 1986, Indiana University. (c) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press LTD.: Somerset, England, 1988. (d) Kuntz, I. D.; Thomason, J. F.; Oshiro, C. M. *Methods Enzymol.* **1989**, *177*, 159−204. (e) Havel, T. F. *Prog. Biophys. Mol. Biol.* **1991**, *56*, 43−78.

(2) (a) Kaptein, R.; Boelens, R.; Scheek, R. M.; van Gunsteren, W. F. *Biochemistry* **1988**, *27*, 5389−5395. (b) Scheek, R. M.; van Gunsteren, W. F.; Kaptein, R. *Methods Enzymol.* **1989**, *177*, 204−218.

(3) (a) Weber, P. L.; Morrison, R.; Hare, D. *J. Mol. Biol.* **1988**, *204*, 483−487. (b) Holak, T. A.; Gondol, D.; Otlewski, J.; Wilusz, T. *J. Mol. Biol.* **1989**, *210*, 635−648.

(4) (a) Mierke, D. F.; Reggelin, M. *J. Org. Chem.* **1992**, *57*, 6365−6367. (b) Reggelin, M.; Köck, M.; Conde-Frieboes, K.; Mierke, D. F. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 753−755. (c) Köck, M.; Junker, J. *J. Mol. Model.* **1997**, *3*, 403−407. (d) Köck, M.; Junker, J. *J. Org. Chem.* **1997**, *62*, 8614−8615; (*Erratum*) **1998**, *63*, 2409. (e) Junker, J.; Reif, B.; Steinhagen, H.; Junker, B.; Felli, I. C.; Reggelin, M.; Griesinger, C. *Chem.—Eur. J.* **2000**, *6*, 3281−3286.

(5) Grube, A.; Köck, M. *Angew. Chem., Int. Ed.* **2007**, *46*, 2320−2324.

(6) (a) Kinnel, R. B.; Gehrken, H. P.; Scheuer, P. J. *J. Am. Chem. Soc.* **1993**, *115*, 3376−3377. (b) Kinnel, R. B.; Swali, R.; Skoropowski, G.; Gehrken, H.-P.; Scheuer, P. J. *J. Org. Chem.* **1998**, *63*, 3281−3286.

(7) Seiple, I. B.; Su, S.; Young, I. S.; Lewis, C. A.; Yamaguchi, J.; Baran, P. S. *Angew. Chem., Int. Ed.* **2010**, *49*, 1095−1098.

(8) For a review, see: Köck, M.; Grube, A.; Seiple, I. B.; Baran, P. S. *Angew. Chem., Int. Ed.* **2007**, *46*, 6586−6594.

(9) (a) Urban, S.; de Almeida Leone, P.; Carroll, A. R.; Fechner, G. A.; Smith, J.; Hooper, J. N. A.; Quinn, R. J. *J. Org. Chem.* **1999**, *64*, 731−735. (b) Yamaguchi, J.; Seiple, I. B.; Young, I. S.; O'Malley, D. P.; Maue, M.; Baran, P. S. *Angew. Chem., Int. Ed.* **2008**, *47*, 3578−3580. (c) O'Malley, D. P.; Yamaguchi, J.; Young, I. S.; Seiple, I. B.; Baran, P. S. *Angew. Chem., Int. Ed.* **2008**, *47*, 3581−3583. (d) Seiple, I. B.; Su, S.; Young, I. S.; Nakamura, A.; Yamaguchi, J.; Jørgensen, L.; Rodriguez, R. A.; O'Malley, D. P.; Gaich, T.; Köck, M.; Baran, P. S. *J. Am. Chem. Soc.* **2011**, *133*, 14710−14726.

(10) (a) Nishimura, S.; Matsunaga, S.; Shibazaki, M.; Suzuki, K.; Furihata, K.; van Soest, R. W. M.; Fusetani, N. *Org. Lett.* **2003**, *5*, 2255−2257. (b) Grube, A.; Immel, S.; Baran, P. S.; Köck, M. *Angew. Chem., Int. Ed.* **2007**, *46*, 6721−6724. (c) Su, S.; Seiple, I. B.; Young, I. S.; Baran, P. S. *J. Am. Chem. Soc.* **2008**, *130*, 16490−16491.

(11) A pseudoatom is the geometrical averaged position of *n* protons. For nonstereospecifically assigned methylene groups its position is between the two real protons. For methyl groups it is the averaged postions of the three real protons. When pseudoatoms are used within simulations, the corresponding values for the upper bounds have to be corrected.

(12) For the ROESY analysis (and fc-rDG/DDD calculations) of 3,7-*epi*-massadine chloride (**4**) the diazide derivative was used (the azides are attached to C1′ and C1″).