Computational Chemistry Driven Solution to Rubriflordilactone B

Nicolás Grimblat,[ID] Teodoro S. Kaufman,* and Ariel M. Sarotti*[ID]

Instituto de Química Rosario (IQUIR, CONICET-UNR) and Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531 (S2002LRK) Rosario S2002LRK, Argentina

S *Supporting Information*

ABSTRACT: The structure of rubriflordilactone B (**2**) was determined by X-ray crystallography. However, the NMR data of the synthetic sample did not match those reported for **2**. It was then suggested that the original sample contained an additional isomer of different solubility, pseudorubriflordilactone B (**3**), whose structure remained unknown. From theoretical calculations, reexamination of the NMR data, and biogenetic considerations, it is proposed that **3** should be the $16S,17R$ isomer of **2**.

Rubriflordilactone B (**2**)   Pseudorubriflordilactone B (**3**)
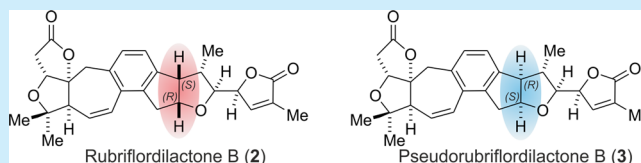
The Schisandraceae, comprising about 50 species of medicinally and economically relevant climbing plants belonging to the genera Katsura and Schisandra, are chiefly distributed in North America and the Southeast of Asia. Their phytochemistry has been recently reviewed, covering around 400 triterpenoids identified up to 2014.[1]

In 2006, Sun and co-workers reported the isolation of two highly unsaturated rearranged bisnortriterpenoids from *Schisandra rubriflora*, designated as rubriflordilactones A (**1**) and B (**2**), Figure 1. These structures, featuring unique polysub-
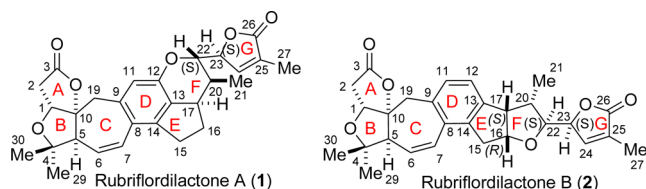

Figure 1. Proposed structures of rubriflordilactones A (**1**) and B (**2**).

stituted central aromatic rings, were respectively assigned to the novel compounds after detailed NMR spectroscopic analyses. Further, both proposed structures were confirmed by X-ray crystallography.[2]

Both natural products exhibited anti-HIV activity; the EC_{50} of **2** against viral replication in HIV-1$_{IIIB}$-infected C8166 cells was 9.75 μg/mL. The latter also displayed low cytotoxicity against K562 cells.

Two total syntheses of compound **1** have been reported so far,[3] whereas synthetic studies of the CDE, 5-*epi*-ABCDE, and DEFG ring fragments of **2**[4] were crowned by its recent total synthesis, disclosed by Li and co-workers.[5] Surprisingly, however, although the X-ray analysis of the synthesized structure **2** clearly confirmed its identity, its ¹H and ¹³C NMR spectra were unlike those of the authentic product.

Since the differences could not be attributed to solvent, temperature, concentration, or other usual factors, the authors hypothesized that the original sample of "rubriflordilactone B"

may have contained two compounds with different solubility. The minor one crystallized, affording the reported X-ray structure which agrees with **2**, whereas the major component, probably an isomer of **2**, which was termed "*pseudo*rubriflordilactone B", could be responsible for the observed bioactivities and published specific rotation and NMR spectroscopic data.

Taking into account our interest in natural product characterization[6] using quantum chemical calculations of NMR shifts[7] and considering the remote possibility of unveiling the structure of pseudorubriflordilactone B (**3**) by reisolation of the natural product, we decided to examine the problem from a theoretical viewpoint. This approach has been employed in recent and resonant cases of structural misassignments,[8a–d] providing useful insight in subtle issues beyond the limits of X-ray methods as well.[8e]

As most of the ¹³C chemical shifts of isolated **3** were in close agreement with those found for synthetic compound **2** (except for δC15, vide infra),[9] it seemed sound to postulate that the misassignment problem was stereochemical in nature, as suggested by Li.[5]

A wide variety of strategies have been employed recently to settle stereochemical issues of complex organic molecules from quantum chemically computed NMR shifts.[10] Among them, the DP4 probability has emerged as one of the most popular methods of choice when only one set of experimental data is available, as in the present case.[11] We have recently developed an updated version, termed DP4+, proving that the use of both scaled and unscaled shifts, computed at higher levels of theory, remarkably improves the predictive performance.[7a] Hence, DP4+ became the method of choice to unveil the most likely structure of pseudorubriflordilactone B.

Given that the target molecule (**3**) contains eight stereo-centers, a total number of 128 different diastereoisomers are plausible (**PRB1−PRB128**).[9] In order to narrow down the

number of isomers for preliminary calculations, we assumed that the ABC ring system of **2** and **3** might display the same relative configuration, as minor differences in the ^{13}C NMR spectra of both compounds were noted for C1−C10.[9] Moreover, we compared the ^{13}C NMR shifts of other 12 structurally related natural products with the same configurations at rings ABC, also finding a close similarity in the chemical shifts assigned to carbons C1−C5 and C10 with those reported for **2** and **3**.[9]

Thus, we initially studied the 32 different isomers bearing the 1*R*,5*S*,10*R* configuration, identified as **PRB33−PRB64** according to our arbitrary nomenclature system. Although these molecules are fairly rigid, unique conformations (up to 33) were located for each isomer after exhaustive exploration of the conformational space using the MMFF force field. All rotamers were optimized at the B3LYP/6-31G* level of theory for further calculations of the NMR chemical shifts at the PCM/mPW1PW91/6-31+G** level (the optimal for DP4+ calculations)[7a] using the polarizable continuum model PCM (with pyridine as the solvent) as implemented in Gaussian 09.[12]

To validate our methodology, we first correlated the NMR data computed for isomers **PRB33−PRB64** with the experimental NMR values reported for synthetic **2**, whose structure was unambiguously determined by X-ray studies.[5] To our delight, we found that our DP4+ probability strongly suggested (92.6%) isomer **PRB48** (with a 1*R*,5*S*,10*R*,16*R*,17*S*,20*S*,22*S*,23*S* configuration) as the most likely candidate, which was indeed the correct assignment (compound **2**, Figure 1). The ^{13}C NMR data computed for **PRB48** showed excellent agreement with the reported values, with a CMAE (corrected mean average error, defined as $\sum_n |\delta_{scaled} - \delta_{exp}|/n$) and CMaxErr (defined as $\max |\delta_{scaled} - \delta_{exp}|$) values of 1.5 and 5.5 ppm, respectively. Interestingly, such parameters were the lowest among the computed for the other 31 isomers (CMAE range: 1.7−2.8 ppm, CMaxErr range: 5.5−11.6 ppm).[9] Not unexpectedly, the DP4+ probability computed using only ^{13}C NMR data (C-DP4+) was high (94.0%) for **PRB48**. The ^{1}H NMR shift prediction was also good (CMAE = 0.13 ppm, CMaxErr = 0.28 ppm), although not the best one. In particular, **PRB49** (with all the configurations at rings E, F, and G inverted relative to **PRB48**) displayed the lowest CMAE (0.12 ppm) and CMaxErr (0.27 ppm). As a result, the DP4+ probability computed using only ^{1}H NMR data (H-DP4+) for **PRB49** was higher than that for **PRB48** (58.8% vs 6.1%).[9] The fact that the overall DP4+ probability was high for **PRB48**, even with unfavorable H-DP4+ values, strengthens the need for using both types of data whenever possible.[7a,11]

With the validated methodology, the NMR data computed for **PRB33−PRB64** were next correlated with the experimental shifts reported for **3**.[2] In this case, the DP4+ probability showed extremely low confidence for the originally proposed structure **PRB48** (<0.01%). On the other hand, isomer **PRB60** (with a 1*R*,5*S*,10*R*,16*S*,17*R*,20*S*,22*S*,23*S* configuration) was identified as the most likely candidate (DP4+ = 99.5%, Figure 2A). Both isomers differ only on the configurations at C16 and C17, which describe the E/F ring fusion.

In this case, the ^{1}H NMR data conclusively favored the most likely candidate (H-DP4+ = 99.9%). This resulted in an interesting observation considering that the overall agreement between experimental and predicted ^{1}H shifts (CMAE = 0.18−0.32 ppm) was not high as in the case of **2** (CMAE = 0.12−0.28 ppm). Since the shifts of the aliphatic protons were nicely
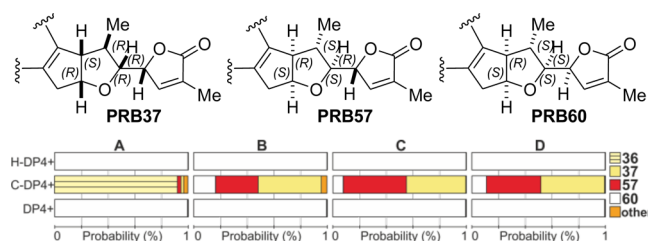

Figure 2. Structures of **PRB37**, **PRB57**, and **PRB60** (only the EFG ring systems are shown) and H-DP4+, C-DP4+, and DP4+ values using the experimental data of **3**: (A) as published; (B) suppressing $\delta(C15)$, (C) $\delta_{exp}(C15) = 35.9$ ppm, (D) $\delta_{exp}(C15) = 38.2$ ppm.

reproduced (the CMAE and CMaxErr computed for **PRB60** were only 0.11 and 0.30 ppm, respectively), this was attributed to unusually high errors in the aromatic region. In turn, this depended on the significant differences between the experimental values reported for those protons in compounds **2** and **3**. For instance, the H12 signal of **3** was considerably shifted downfield than that of **2** (7.89 vs 7.23 ppm, respectively). However, those outliers did not affect our conclusions, as both the H-DP4+ and DP4+ probabilities remained almost unchanged (99.8% and 99.2%, respectively, favoring **PRB60**) after removal of the data of the aromatic protons and recomputing. On the opposite side, the carbon data provided weak arguments toward **PRB60**, with a C-DP4+ value of only 0.5%, with **PRB36** (bearing a *trans*-E/F ring fusion) the most likely candidate (C-DP4+ = 92.0%). Although it is not uncommon that carbon and proton data point toward different directions, the large differences between H-DP4+ and C-DP4+ caught our attention. After closer inspection of the experimental and calculated ^{13}C NMR data for **PRB60**, we identified a high discrepancy in the shift assigned to the C15 methylene. Whereas the reported value was 30.7 ppm, our calculations placed that carbon in a more deshielded region (41.6 ppm, unscaled data). The magnitude of the error gained additional relevance considering that the other sp^3 carbons of the molecule were nicely reproduced by our calculations (CMAE = 1.1 ppm).

Alarmed by this finding, we thoroughly revised the Supporting Information of the isolation paper.[2] In the HSQC spectrum of **3** we found that the signal at 30.7 ppm (a methylene unit according to the DEPT spectrum) lacked cross-peaks with protons of the 3.2−2.8 region, as originally indicated; instead, it exhibited correlation with protons in the region of ∼1.2 ppm. Furthermore, no long-range correlations were observed for this resonance in the HMBC experiment. On the basis of this evidence, we considered that the original authors might have misinterpreted this signal, which probably originated from an impurity present in the sample. In fact, after the calculated and experimental chemical shifts for C15 were removed from all isomers, the C-DP4+ probability computed for **PRB36** and **PRB60** changed to 0.12 and 16.2%, respectively (Figure 2B), indicating that the high C-DP4+ value initially computed for the former was strongly related to the C15 resonance.

Nevertheless, if our hypothesis was right, we still had to perform the assignment of the missing C15 methylene. The DEPT spectra of **3** exhibited two additional CH$_2$ carbons within the 30−45 ppm region: C2 (35.9 ppm) and C19 (38.2 ppm).[2] Since, according to the HSQC spectrum, both carbons correlate with protons in the 3.2−2.8 ppm region, the same

zone suggested for the H15 resonances, we speculated that the signal of C15 could be overlapped with that of either C2 or C19. The C-DP4+ values calculated using these assignments were 7.6% (Figure 2C) and 10.8% (Figure 2D), respectively, suggesting the latter as the most likely scenario. Even though the overall DP4+ probability was always high for **PRB60** (>99%), we performed a detailed analysis of those isomers that afforded high C-DP4+ values. Assuming that the chemical shift of C15 was 38.2 ppm, those isomers were **PRB37**, **PRB57**, and **PRB60** with C-DP4+ values of 47.2%, 40.3%, and 10.8%, respectively (Figure 2D). Interestingly, **PRB37** is the pseudoenantiomer of **PRB60** at the right side of the molecule (that is, with all the stereocenters of this region are inverted), whereas **PRB57** is the C23-epimer of **PRB60**.

The parity among the C-DP4+ values paralleled the degree of congruence between the experimental and calculated ^{13}C NMR data exhibited by all three isomers; for instance, **PRB37**, **PRB57**, and **PRB60** displayed the same CMAE level (1.7 ppm). Thus, it seemed judicious to conjecture that the ^{1}H NMR data could harbor the discriminating information. Accordingly, we found significantly higher errors in the H22 and H23 signals of the former two (0.15 and 0.20 ppm, respectively, for **PRB37**, and 0.37 and 0.47 ppm, respectively, for **PRB57**) than those observed for **PRB60** (0.06 and 0.05 ppm, respectively). Such differences accounted for the high H-DP4+ probability computed for **PRB60**.

An additional argument favoring **PRB60** resulted from the detailed examination of the NOESY spectrum of **3**, where a correlation between H1 (4.42 ppm) and H27 (1.76 ppm) was noticed.[2] Although the signals seem slightly offset, there are no other NMR resonances overlapping those corresponding to H1 and H27. Curiously, this apparently key correlation that would completely rule out the stereochemistry of **2**, proposed for **3**, was missed by the authors. Therefore, we evaluated whether this finding suited our DP4+ analysis presented above. Hence, all of the relevant conformations of the 32 isomers were again scrutinized to understand the geometrical factors that would be positioned in close proximity these protons, located in the antipodes of the molecule. From this analysis, the need for a *cis*-E/F ring-junction with a 16R,17S configuration became clear. This would allow the right side of the molecule to bend inward along its β-face, to which H1 is directed, nearing the latter. Moreover, the configuration at C22 must be *S* to enable the pendant lactone unit to approach the AB ring system. Only four isomers met these requirements: **PRB51**, **PRB54**, **PRB57**, and **PRB60**, and in each of these cases at least one conformation was found that showed an H1/H27 distance lower than 5 Å (Figure 3). However, only in the case of **PRB60** were such conformers among the most stable conformations. In the remaining molecules, the conformations that would account for this NOESY interaction are clearly disfavored energetically (above 2.5 kcal/mol from the global minima).

Finally, to rule out the possibility that any another stereoisomer besides the 32 considered thus far could be the true structure of **3**, we calculated the chemical shifts of the remaining 96 isomers at the PCM/mPW1PW91/6-31+G**// B3LYP/6-31G* level of theory. Then, the DP4+ was recomputed considering the whole set of 128 candidates, and we were glad to observe that the values shown in Figure 3 exhibited only minor changes. Once again, isomer **PRB60** was identified as the most likely candidate with a high level of confidence (H-DP4+ = 86.8%, C-DP4+ = 10.8%, and DP4+ = 99.5%).[9]
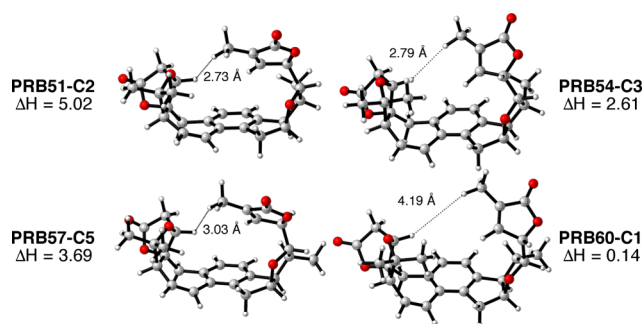

Figure 3. B3LYP/6-31G*-optimized geometries of the conformers of **PRB51**, **PRB54**, **PRB57**, and **PRB60** that showed proximity between H1/H27. The enthalpies (relative to the global minimal conformation found in each case) are given in kcal/mol.

Biogenetic postulates may also account for the proposed structures for **1**−**3** (Scheme 1). Hypothetically, these

Scheme 1. Proposed Common Biogenetic Origins of 1−3
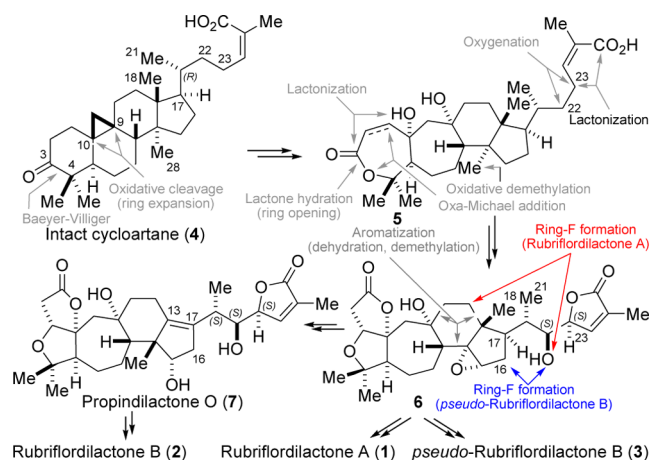
bisnortriterpenoids arise from an intact cycloartane (**4**) through the intermediacy of **5** and a key schiartane like micrandilactone B (**6**), formed as proposed.[1b] In turn, the biosynthesis of **1** would entail aromatization of the D-ring through dehydration and oxidative demethylation reactions (including decarboxylation) and formation of the F-ring by etherification of C12 with the C22 alcohol moiety. In this process, all the intermediates would conserve the 17R-configuration of the starting cycloartane (**4**).

Analogously, the biosynthesis of **3** would involve D-ring aromatization, whereas generation of the F-ring would take place by etherification of C16 with the alcohol pending from C22. As before, the stereochemical integrity of C-17 along the biosynthetic process would result in the opposite configurations at C-16 (*S*) and C17 (*R*) with regard to **2** (16R,17S) as a result of the preferred E/F-ring *cis*-fusion. On the other hand, the biogenesis of **2**, isolated concomitantly and structurally assessed in an unambiguous form, could similarly be proposed as occurring through 18(13 → 14)-*abeo*schiartane intermediates, like kadcoccilactone F or propindilactone O (**7**), which have lost their original configuration at C-17 as a result of a C18 rearrangement.[13]

In summary, relying on the assumption that the true structure of **3** is indeed a diastereoisomer of the originally proposed polycycle **2**, careful analysis of the published NMR

spectral data of "authentic rubriflordilactone B", coupled with exhaustive theoretical calculations, led us to suggest the 16S,17R isomer of Li's synthetic rubriflordilactone B (**2**) as the most likely structure of pseudorubriflordilactone B (**3**). Coherent biogenetic paths were also postulated for both diastereomeric heterocycles and for the related compound **1**, from a common intermediate, reinforcing the structural proposition. In this regard, the strict requirements imposed by the structural problem in terms of precision of signal positioning emphasize the importance of disseminating NMR data through their original FIDs to enable more detailed subsequent analyses.[8a,14] We hope that this study, which underscores the potential of computational methods as a key tool for assessing the structure of complex organic molecules, will provide helpful guidance to further synthetic work needed to ultimately unravel the mystery that surrounds rubriflordilactone B.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.6b03318.

> Computational methods; NMR resonances of synthetic and isolated rubriflordilactone and 12 other structurally related natural products; Boltzmann-averaged isotropic shielding tensors, scaled and unscaled chemical shifts, DP4+ results, and Cartesian coordinates for all compounds under study (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: kaufman@iquir-conicet.gov.ar.
*E-mail: sarotti@iquir-conicet.gov.ar.

ORCID Ⓘ

Nicolás Grimblat: 0000-0003-3285-8485
Ariel M. Sarotti: 0000-0002-8151-0306

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) (a) Xiao, W.-L.; Li, R.-T.; Huang, S.-X.; Pu, J.-X.; Sun, H.-D. *Nat. Prod. Rep.* **2008**, *25*, 871−891. (b) Shi, Y.-M.; Xiao, W.-L.; Pu, J.-X.; Sun, H.-D. *Nat. Prod. Rep.* **2015**, *32*, 367. (c) Xia, Y.-G.; Yang, B.-Y.; Kuang, H.-X. *Phytochem. Rev.* **2015**, *14*, 155.

(2) Xiao, W.-L.; Yang, L.-M.; Gong, N.-B.; Wu, L.; Wang, R.-R.; Pu, J.-X.; Li, X.-L.; Huang, S.-X.; Zheng, Y.-T.; Li, R.-T.; Lu, Y.; Zheng, Q.-T.; Sun, H.-D. *Org. Lett.* **2006**, *8*, 991.

(3) (a) Goh, S. S.; Chaubet, G.; Gockel, B.; Cordonnier, M.-C. A.; Baars, H.; Phillips, A. W.; Anderson, E. A. *Angew. Chem., Int. Ed.* **2015**, *54*, 12618. (b) Li, J.; Yang, P.; Yao, M.; Deng, J.; Li, A. *J. Am. Chem. Soc.* **2014**, *136*, 16477.

(4) (a) Peng, Y.; Duan, S.-M.; Wang, Y.-W. *Tetrahedron Lett.* **2015**, *56*, 4509. (b) Wang, Y.; Li, Z.; Lv, L.; Xie, Z. *Org. Lett.* **2016**, *18*, 792. (c) Goh, S. S.; Baars, H.; Gockel, B.; Anderson, E. A. *Org. Lett.* **2012**, *14*, 6278.

(5) Yang, P.; Yao, M.; Li, J.; Li, Y.; Li, A. *Angew. Chem., Int. Ed.* **2016**, *55*, 6964.

(6) Simonetti, S. O.; Larghi, E. L.; Bracca, A. B. J.; Kaufman, T. S. *Org. Biomol. Chem.* **2012**, *10*, 4124.

(7) (a) Grimblat, N.; Zanardi, M. M.; Sarotti, A. M. *J. Org. Chem.* **2015**, *80*, 12526. (b) Zanardi, M. M.; Sarotti, A. M. *J. Org. Chem.* **2015**, *80*, 9371. (c) Novaes, L. F. T.; Sarotti, A. M.; Pilli, R. A. *J. Org. Chem.* **2015**, *80*, 12027. (d) Sarotti, A. M.; Suárez, A. G.; Spanevello, R. A. *Tetrahedron Lett.* **2011**, *52*, 3116.

(8) (a) Reddy, D. S.; Kutateladze, A. G. *Org. Lett.* **2016**, *18*, 4860. (b) Rychnovsky, S. D. *Org. Lett.* **2006**, *8*, 2895. (c) Lodewyk, M. W.; Tantillo, D. J. *J. Nat. Prod.* **2011**, *74*, 1339. (d) Lodewyk, M. W.; Soldi, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012**, *134*, 18550. (e) Siskos, M. G.; Choudhary, M. I.; Tzakos, A. G.; Gerothanassis, I. P. *Tetrahedron* **2016**, *72*, 8287.

(9) For further details on this issue, see the SI.

(10) (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839. (b) Grimblat, N.; Sarotti, A. M. *Chem. - Eur. J.* **2016**, *22*, 12246.

(11) Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946.

(12) Frisch, et al. *Gaussian 09*, C.01 ed.; Gaussian, Inc., Wallingford, CT, 2009. See the Supporting Information for full reference.

(13) (a) Gao, X.-M.; Pu, J.-X.; Huang, S.-X.; Lu, Y.; Lou, L.-G.; Li, R.-T.; Xiao, W.-L.; Chang, Y.; Sun, H.-D. *J. Nat. Prod.* **2008**, *71*, 1182. (b) Huang, S. X.; Yang, L. B.; Xiao, W. L.; Lei, C.; Liu, J. P.; Lu, Y.; Weng, Z. Y.; Li, L. M.; Li, R. T.; Yu, J. L.; Zheng, Q. T.; Sun, H. D. *Chem. - Eur. J.* **2007**, *13*, 4816. (c) Yang, J. H.; Wen, J.; Du, X.; Li, X. N.; Wang, Y. Y.; Li, Y.; Xiao, W. L.; Pu, J. X.; Sun, H. D. *Tetrahedron* **2010**, *66*, 8880. (d) Lei, C.; Pu, J. X.; Huang, S. X.; Chen, J. J.; Liu, J. P.; Yang, L. B.; Ma, Y. B.; Xiao, W. L.; Li, X. N.; Sun, H. D. *Tetrahedron* **2009**, *65*, 164.

(14) Pauli, G. F.; Niemitz, M.; Bisson, J.; Lodewyk, M. W.; Soldi, C.; Shaw, J. T.; Tantillo, D. J.; Saya, J. M.; Vos, K.; Kleinnijenhuis, R. A.; Hiemstra, H.; Chen, S.-N.; McAlpine, J. B.; Lankin, D. C.; Friesen, J. B. *J. Org. Chem.* **2016**, *81*, 878.