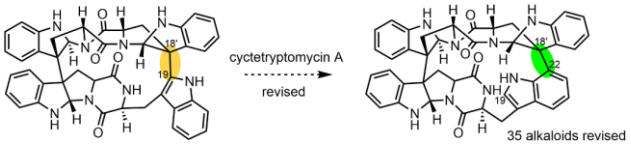
DU8ML: Machine Learning-Augmented DFT NMR Computations for High-Throughput *in Silico* Solution Structure Validation and Revision of Complex Alkaloids.

Ivan M. Novitskiy and Andrei G. Kutateladze*

Department of Chemistry and Biochemistry, University of Denver, Denver, CO 80208



ABSTRACT: Machine Learning (ML) profoundly improves accuracy of the fast DU8+ hybrid DFT/parametric computations of NMR spectra, allowing for high throughput *in silico* validation and revision of complex alkaloids and other natural products. Of nearly 170 alkaloids surveyed, 35 structures are revised with the next generation ML-augmented DU8 method, termed DU8ML.

INTRODUCTION

Computational NMR is in a twilight zone today. Quantum chemistry methods exist to compute chemical shifts and nuclear spin coupling constants with high accuracy, but as a rule these methods are not yet practical for large natural products (NPs), where such computations may take days, if not weeks of cpu time.[1] On the other hand, existing Computer Assisted Structure Elucidators (CASE)[2] – usually based on various neural networks (NN) algorithms – generate and evaluate massive numbers of structural candidates in spectacularly short computational times. Alas many of them severely lack in the accuracy department, unless augmented by DFT.

The niche between these two extremes must be filled with both fast and accurate computational approaches. One of these is to utilize calculations at the lower levels of DFT theory as a "zeroth order" baseline, with subsequent parametric corrections for the inevitable systematic errors. Utilization of such remedial correction factors is broadly precedented in quantum chemistry. One needs to look no further than the venerable 0.89 to 0.96 scaling factors for vibrational frequencies and zero-point energies calculated at lower levels of *ab initio* theory.[3]

Today the introduction and optimization of massive sets of such parametric corrections is termed *machine learning* (ML), but it does not change the fact that one achieves *both* a considerable acceleration of the computations *and* the improved accuracy at the same time, provided that the output of a minimalist level DFT calculation is tweaked with a set of ML-derived empirical corrections. This has been our approach, which resulted in DU8+ hybrid DFT-parametric method. In the lingo of the simplest (human-mediated) machine learning, we "labeled" the substructure fragments – which accounted for major deviations in the DFT calculated values – with the appropriate SMARTS strings[4] and trained the machine on a large set of experimental nuclear spin coupling constants or chemical shifts to detect the discrepancies and correct for them. In 2014, we expanded on Bally and Rablen's idea[5] of scaling Fermi contacts and developed a fast and accurate method for computing nuclear spin-spin coupling constants based on a substructure-aware scaling.[6] Later we applied a similar methodology for calculations of chemical shifts.[7] Most notably, for molecules containing heavy atoms, e.g. halogens, this pragmatic approach circumvented the need to deploy arduous spin-orbit coupling, SOC, calculations. Uncorrected, the computed C-Br chemical shifts deviate by more than 10ppm, and the C-I chemical shifts – by more than 30ppm. In fact, more than a decade ago Braddock and Rzepa in their reassignment of obtusallenes[8] suggested that a fixed value correction for non-relativistic computations of C-Br/I chemical shifts could be a practical substitute for the difficult SOC calculations. One imagines that Rzepa's approach was an early instance of human-mediated machine learning in computational NMR.

Shortly after our 2017 paper,[7] in which we presented binomial correction functions for chemical shifts of carbons bearing heavy atoms, Fandrick, Gonnella and co-workers published on what they call empirically derived systematic error correction terms, which has a similar philosophy, albeit the corrections were single fixed value offsets.[9] As the

terminology evolved, Zhang and co-workers call this approach *Machine Learning Augmented DFT*,[10] which probably most accurately reflects the nature of our efforts as well. Similar approaches are beginning to emerge in the field of computational NMR.[11] Whatever the term is, empirical corrections of NMR parameters calculated at the light DFT levels could significantly improve the cost-benefit outcomes for the current state of computational hardware.

One practical inference from our prior observations is that rigorous (and accurate) DFT or *ab initio* methods are too slow and simply could not keep up with the flood of published misassigned structures of natural products. As we have shown in the past, depending on the nature of "difficult" structural elements, the misassignment rate for halogenated marine natural products[7] or NPs containing oxirane,[12] oxetane,[13] triquinane,[14] or carboxylic anhydride[15] and other[16] moieties could be as high as 12-20%. This translates into a rather large number of incorrect structure assignments in need of revision, which accumulate with an alarming rate. Yet, on the other end of this spectrum, oversimplified zeroth order DFT methods augmented with some version of ML at this time produce insufficient accuracy of calculated ^{13}C chemical shifts, with rmsd's exceeding 4-6ppm.[11b,17] It may take time until the accuracy of these methods is (systematically) improved to give a practical reassignment tool.

Our pragmatic objective was to develop a method capable of accurate computations of NMR chemical shifts and spin-spin coupling constants (SSCC) of large NP's, such as C20-diterpenes, in 15-20 min of computational time on a node of a garden variety Linux cluster accessible to most practitioners in the field. With the most recent implementation of ML-augmented DU8+ method, *that we now term DU8ML*, we achieved these short computational times without sacrificing the accuracy of computations, which currently stands at rmsd=0.95ppm for ^{13}C chemical shifts (over the training set of 11K experimental chemical shifts) and 0.28Hz for proton spin-spin coupling constants (training set of ~4K experimental J's).

With the new DU8ML method, the rms deviations for correct structures of NPs normally range from under 1.0 to 1.5 ppm. We are currently developing a DU8 implementation of Goodman's absolute DP5 scheme[18] to attach probabilistic values to potential candidate structures. However, in this paper we will be using the rmsd values for ^{13}C chemical shifts as the primary criterion for our revisions or validations of the analyzed structures.

At this point our method is trained on the experimental data obtained in chloroform solutions. For spectra in chloroform we will report "rmsd" values. Other solvents require an additional linear correction, and the results will be reported as "crmsd," i.e. corrected-rmsd (see Supporting Information for more detail).

Figure 1 shows the flowchart for DU8ML. The orange blocks outline the order of the underlying Gaussian computations, while the purple ones illustrate the ML-derived refinement of the DFT-computed values.

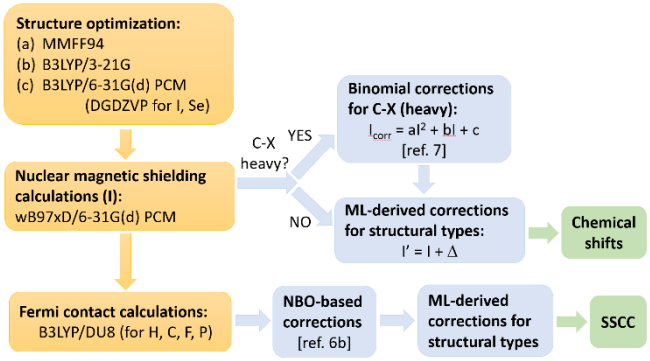


Figure 1. DU8ML flowchart.

Alkaloids are one of the most important classes of natural products. Their biological activity ranges from anti-inflammatory[19a,c] to analgesic,[19c] anti-cancer,[19a,b] antioxidant,[19c] antimicrobial activity[19c] and a whole array of other useful properties. The high value of alkaloids and their synthetic derivatives for drug design and discovery imparts additional burden on the accuracy of their solution structure elucidation in cases when x-ray analysis is not an option. Total synthesis of the putative structures remains a powerful tool for validation or revision of alkaloids,[20] although it is far from "high throughput," and also is not fully guaranteed from errors.[16b] DU8ML offers a practical and robust tool for structure elucidation workflow.

RESULTS AND DISCUSSION

DU8ML demonstrated excellent performance when tested against a probe set of correct structures containing synthetic nitrogen heterocycles, Table S1, and complex alkaloids, Table S2. The rms deviations of the ^{13}C chemical shift calculations for individual correctly assigned molecules in this rather extensive probe set ranged from 0.79 to 1.36ppm. Cumulatively we examined nearly 170 reported alkaloids and discovered 35 misassigned structures, for which we proposed revisions matching the experimental NMR data. While these examples were selected arbitrarily for this paper, the misassignment rate of approximately 20% is in keeping with our previous observations. This fast *in silico* structure revision work was only feasible because of the short computational times afforded by DU8ML. A typical distribution of the "wall clock" times (i.e. elapsed, not cpu time) for the underlying Gaussian16 DFT computations is presented in Figure 2 as a function of molecular weight. Most calculations for diterpenoid- and similar size alkaloids, including the structure optimization step, were completed under 20 min, allowing for expeditious and confident validation or revision of the structures. Including synthetic nitrogen heterocycles, nearly 150 validation cases are presented in the Supporting Information section and attest to the high fidelity of the method. Figure 3 illustrates a subset of validated structures to showcase the complexity of alkaloids in the probe set and the accuracy of DU8ML.

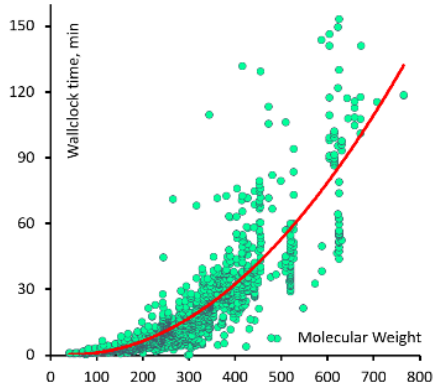


Figure 2. Computational times as a function of molecular weight.

The full set of validated structures, including natural alkaloids and synthetic nitrogen heterocycles, is presented in the Supporting Information section. Below we discuss our alkaloid structure revision results, beginning with less complicated cases of smaller alkaloids and progressing to more complex ones.

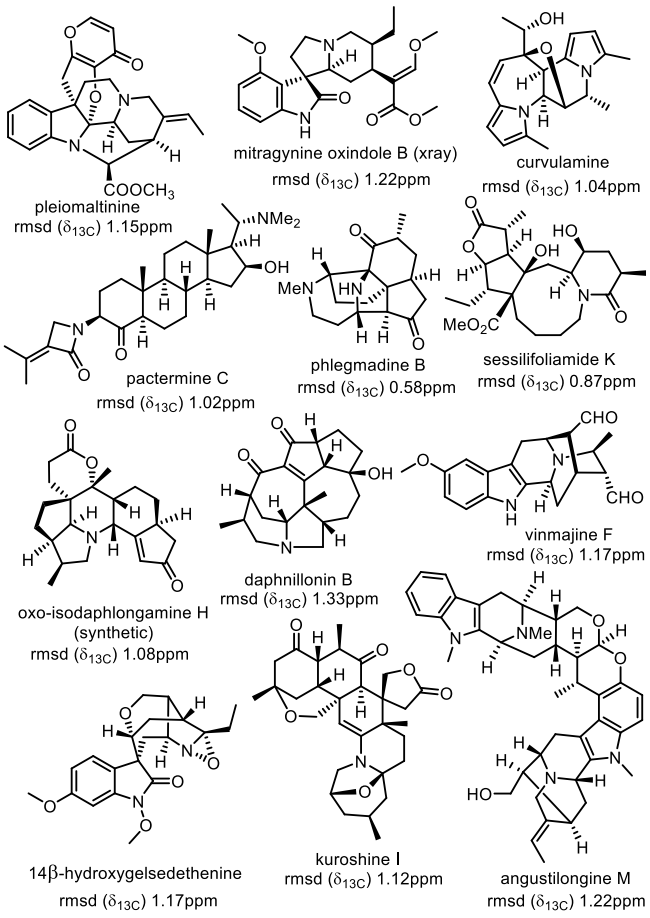


Figure 3. A small subset of validated alkaloids (see Supporting Information for the complete list).

Proton-deficient substituted alkaloids containing fused fully- or partially conjugated rings present added challenge for structure elucidation, complicating COSY- or HMBC-based assignment of connectivity between the fused heterocyclic moieties. Acanthiline A,[21] isolated from the Chinese mangrove *Acanthus ilicifolius* Linn is an instructive example, where the sole proton in the azaquinone ring is seven bonds away from the closest proton in the indole ring, Figure 4. DU8ML gave poor match for the originally proposed structure, rmsd(δ_{13C})=9.48ppm, indicating misassignment. Further analysis of discrepancies suggested that the proposed indole moiety is unlikely. As the DU8ML computational time for each of the $C_{14}H_{11}NO_4$ candidate structures was under 10 min, we were able to generate 21 revision candidates and promptly check them against the experimental NMR data, converging on the shown revised structure, rmsd(δ_{13C})=1.38ppm. Literature search has revealed that the revised structure belongs to a known compound, baphicacanthin A, isolated from a related plant, *Baphicacanthus* (Acanthaceae).[22]

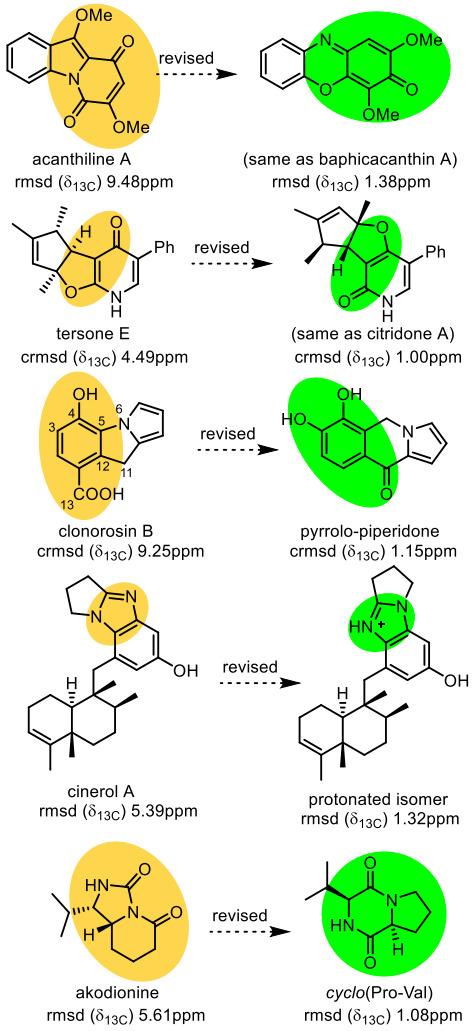


Figure 4. Revision of acanthiline A, tersone E, clonorosin B, cinerol A and akodionine.

The next two alkaloids in Figure 4 were reported in methanol-d4, necessitating additional linear correction of the calculated chemical shifts. For this reason, we report crmsd, i.e. corrected rmsd values.A phenylfuropyridone, tersone E, isolated from the deep-sea fungus *Phomopsis tersa*,[23] was assigned a 4-pyridone structure. This putative

structure was rejected by DU8ML with poor crmsd(δ_{13C})=4.49ppm. "Flipping" the fused cyclopentenofuran produced the shown revised structure, crmsd(δ_{13C})=1.00ppm. Literature search uncovered that this alkaloid was previously isolated from *Penicillium* sp. FKI-1938 and named citridone A.[24] Curiously, a recent revision of drazepinone,[25] isolated from another marine-derived fungus *Penicillium sumatrense*, also resulted in the same citridone A structure, attesting to the importance of dereplication efforts.

Clonorosin B, the third example in Figure 4, an indole alkaloid isolated from a soil-derived *Clonostachys rosea*,[26] also showed a very poor match with the DU8ML-computed values, crmsd(δ_{13C})=9.25ppm. The most prominent discrepancies were associated with the carboxylate and two ortho-carbons, C1 and C12. The peak for methylene C11 at 44.4ppm indicated C-N connection. These observations narrowed the number of candidate structures to twelve, *for which DU8ML calculations took 2-5min per candidate*, resulting in the shown pyrrolo-piperidone as the final revision, crmsd(δ_{13C})=1.15ppm.

Nitrogenous meroterpenoid cinerol A,[27] isolated from the marine sponge *Dysidea cinerea*, also required a "flip" of the amidine moiety fused to the phenol pendant. Additionally, cinerol A appears to be protonated, as confirmed by the matching computed chemical shifts for the shown protonated species, rmsd(δ_{13C})=1.32ppm. (See Supporting Information for additional examples of complications due to protonation in NMR solvents ostensibly containing acidic impurities).

Finally, an unusual structure was proposed for alkaloid akodionine, a secondary metabolite from the endophytic fungus *Xylaria cubensis*,[28] Figure 4. DU8ML showed significant discrepancies for the NMR calculated for its putative structure, rmsd=5.61ppm. The spectral data for akodionine was reminiscent of proline-based diketopiperazines, and the calculated values for *cyclo*(Pro-Val) provided an excellent match with the experimental data, rmsd(δ_{13C})=1.08ppm, settling the revision.

Several examples of structure revisions not matching the original mass-spectrometry (MS) data are shown in Figure 5. The proposed structure of diketopiperazine matsudipeptide A, isolated recently from the fruiting bodies of the edible mushroom *Tricholoma matsutake*,[29] contains an endoperoxide moiety, which was inconsistent with the calculated ^{13}C chemical shifts, rmsd=2.85ppm. As the major deviations were co-located in the 1,3,4-azadioxane ring, we explored different isomers of oxazolidino-diketopiperazine, and revised the structure of matsudipeptide A to the *cis*-dimethyloxazolidine shown in Figure 5, which matched the experimental NMR data with rmsd=0.71ppm.

Another error was identified in the series of N-formylnortropane alkaloids, isolated from the leaves and bark of *Pellacalyx saccardianus*,[30] which we revised to their thioformyl derivatives. The listed ^{13}C values for the formyl groups were much too high (~180ppm) for a formamide derivative. Another compound in this series, 3α-cinnamoyl-N-formylnortropane (not shown), which was assigned correctly, has the formyl peak expectedly at a higher field, 157.4 ppm. Revision of these nortropanes is all the more important given the fact that the 3α-benzoyloxy derivative has demonstrated high relaxant activity in a carbachol-induced contraction assay, with EC_{50} = 3.6μM.

Magnificines A and B, alkaloids containing hydroperoxy-oxazolidinone moiety, were recently isolated from the Red Sea sponge *Negombata magnifica*,[31] Figure 5. However, the DU8ML calculated spectra showed irreconcilable differences with the reported NMR data. As these alkaloids demonstrated modest antimicrobial activity against three pathogens, we deemed it important to pursue the revision of their misassigned structures. Our efforts resulted in an unexpected finding: DU8ML computations of a series of candidate structures led to our revision of magnificines to the known loliolide (at times dubbed the most ubiquitous monoterpenoid lactone) and its 3-epimer. The experimental NMR data for loliolide and epi-loliolide also match that of magnificines perfectly. The difference in the exact masses between the putative magnificines and loliolides account for a molecule of nitric acid, HNO_3, but the by far largest peak in the LRESIMS spectrum found in the supporting information for the original paper is labeled 219.08, which is close to 219.099 [loliolide+Na$^+$].

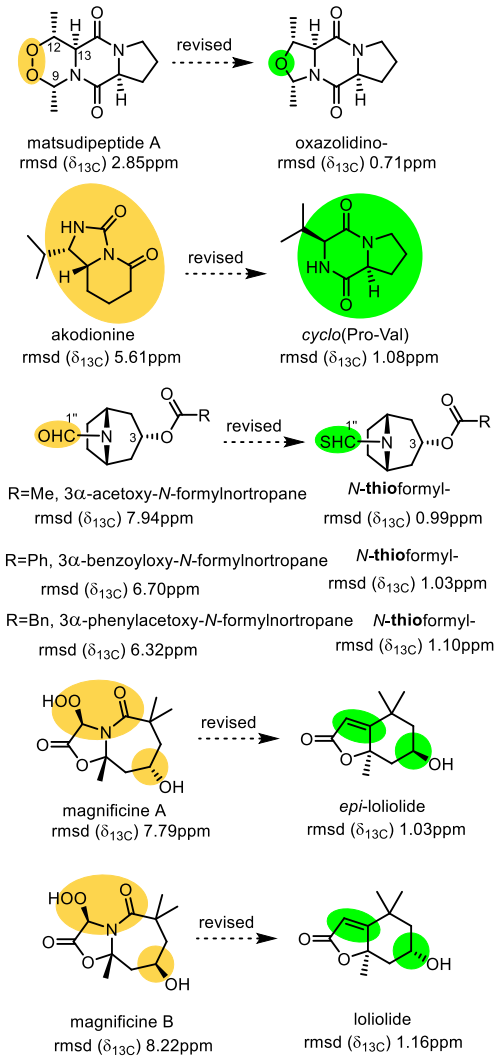


Figure 5. Revisions of matsudipeptide A, akodionine, nortropane alkaloids, magnificines A and B.

Human chemical intuition was used in this study to limit the range of candidate structures to a small subset of the most likely ones. Yet, the short computational times afforded by DU8ML could potentially allow for screening massive sets of computer-generated candidates in a fully automated manner in a practical time frame. Such fully automated structure generators based on chemical formulae are broadly available. Still, the challenges for automation are highlighted by revisions in Figure 5, where the incorrect chemical formulae were derived from the erroneous analysis of the MS data.

Figure 6 shows another common misassignment in natural products, including alkaloids, i.e. incorrect regioplacement of substituents. For example, a C20-diterpenoid alkaloid majusidine B, isolated from *Delphinium majus*,[32] was assigned a 3-oxo group in the A ring. While the rest of the molecule was matching the predicted values, our calculations were consistent with the carbonyl group in position 2, rmsd=1.17ppm (note that we truncated the acyl tail to acetate for computational simplicity).

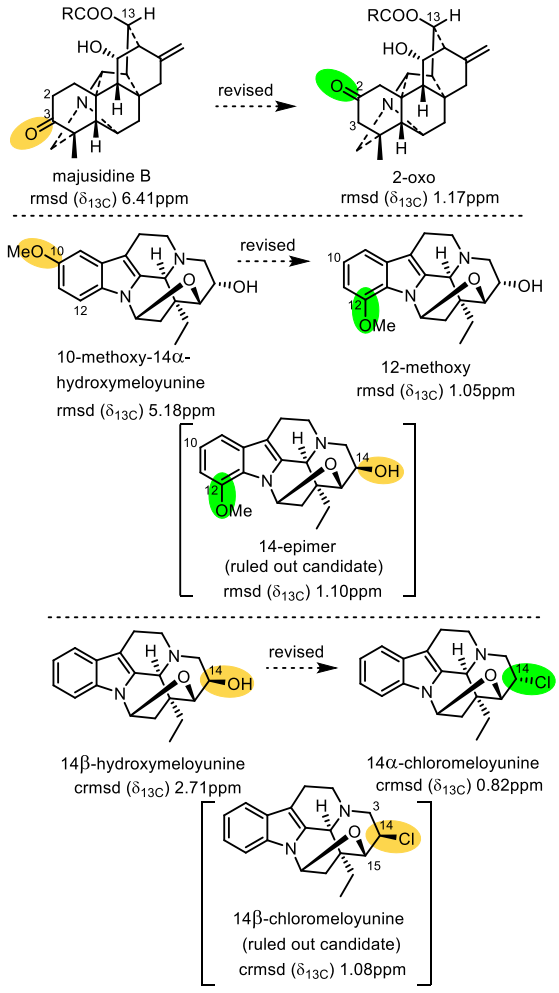
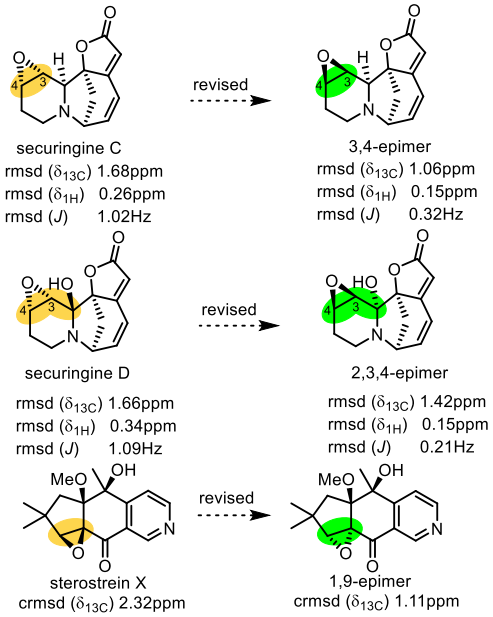


Figure 6. Revision of majusidine B and hydroxymeloyunines.

The structure of a vincan-type alkaloid 10-methoxy-14α-hydroxymeloyunine, Figure 6, recently isolated among the other thirty alkaloids from aerial parts of *Tabernaemontana bovina*,[33] was assigned based on a similar 14β-hydroxymeloyunine isolated earlier from leaves and twigs of *Melodinus yunnanensis*.[34] While DU8ML confirms the core structure for both, it is clear that the methoxy group in "10"-methoxy-14α-hydroxymeloyunine should be moved to position 12.

A related alkaloid 14β-hydroxymeloyunine exhibited additional deviations of the computed values in the vicinity of C14. Inversion of stereoconfiguration at C14-OH did not help to reconcile the calculated chemical shift for this carbon, which differed from its experimental value by 11ppm. At the same time, exclusion of the C14 values from the analysis produced an excellent crmsd(δ_{13C})=1.04ppm. This behavior, i.e. strictly local perturbation of a single chemical shift, was precedented in our prior work and pointed to the *chloro*-substitution. We therefore revised 14β-hydroxymeloyunine to the shown 14α-chloromeloyunine, which produced an excellent match between the calculated and experimental chemical shifts, crmsd(δ_{13C})=0.82ppm. It is not clear whether this nitrogen mustard is formed during an isolation step involving hydrochloric acid. The 14-epimer of the chloride was also considered but ruled out due to inferior matches for both chemical shifts and proton spin coupling constants.

With modern 2D NMR experiments available for thorough solution structure elucidation of natural products – especially in the context of atomic and fragment connectivity – misassignments of individual stereogenic centers remain the most common error. As we reported in the past, the oxirane moiety in complex natural products presents difficulty for the assignment of its stereoconfiguration.[12] Alkaloids are no exception: secondary metabolites isolated from the twigs of *Securinega suffruticosa*, securingines C and D possess the C2-C3 epoxide moiety, Figure 7. DU8ML analysis of their ^{13}C chemical shifts gave marginally acceptable rmsd's under 1.7ppm. However, the 3,4-epimer for securingine C and the 2,3,4-epimer for securingine D match the experimental data with superior rmsd's not only for ^{13}C chemical shifts, but also for SSCCs and ^{1}H chemical shifts.



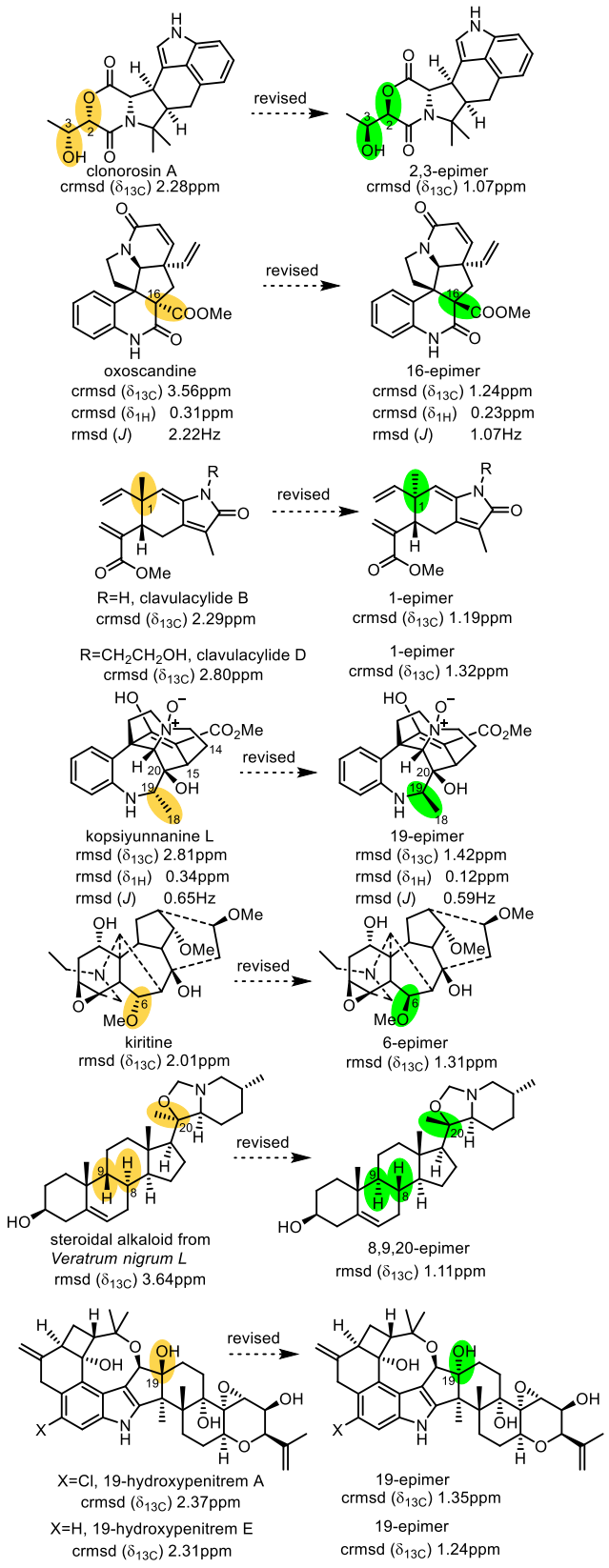


Figure 7. Revision of misassignments in stereoconfiguration.

We therefore revised securingines C and D to their diastereomers shown in Figure 7. Similar revision of the epoxide configuration was required for sterostrein X.[35] Indole alkaloid clonorosin A, isolated from a soil-derived fungus *Clonostachys rosea*[26] (same as clonorosin B above), required correction of stereoconfiguration at C2-C3, as shown, which brought the crmsd(δ_{13C}) from 2.28ppm to 1.07ppm.

Quinolinic scandine-type monoterpenoid alkaloid oxoscandine, isolated from a woody liana *Melodinus henryi* Craib,[36] was assigned an unusual *trans*-fusion of a cyclopentane ring to the quinolinone moiety which, according to DU8ML, needed correction. The structure of oxoscandine is now revised to the shown 16-epimer.Stereoconfiguration at C1 for two elemane alkaloids, clavulacylides B and D,[37] isolated from the soft coral *Clavularia inflata*, needed correction to 1-epimer, matching the C1-stereoconfiguration of their saturated congeners clavulacylides A and C (not shown) isolated in the same campaign.

Revision of kopsiyunnanine L,[38] a terpenoid indole alkaloid[39] isolated from *Kopsia arborea*, to its 18-epimer illustrates a challenging circumstance when the stereochemical assignment could not reliably be made with the analysis of proton coupling constants. Both the original and the revised structures give very similar rmsd(J), 0.65Hz and 0.56Hz. However, rmsd(δ_{13C}) differ by a factor of two, i.e. 2.81 and 1.42ppm, even though the revision involves only one stereogenic center. A similar single C6-stereocenter revision was needed for kiritine, a diterpenoid alkaloid from the roots of *Aconitum kirinense* Nakai.[40]

Three stereogenic centers in the steroidal alkaloid from *Varatrum nigrum* L[41] needed correction, Figure 7. Despite the formidable size of this 28-carbon structure, DU8ML computations for individual conformers required relatively short time in the range of 20-35 min, including the Fermi contact calculations. Finally, the last examples in Figure 7 show our revision of stereoconfiguration in indole-diterpenoids, 19-hydroxypenitrems A and B. These two 37-carbon structures, approaching MW of 650, also required less than an hour of wall clock time per conformer, which we deem very much practical.

Figure 8 illustrates that the conformationally flexible perhydro pyrrolo[2,1,5-de]quinolizine core of crepidine and its congeners seem to present significant challenges for stereochemical assignment in this family. Crepidine itself was isolated in 1970 and characterized at least twice by x-ray crystallography, [42] last time in 2019 but, to the best of our knowledge, its ^{13}C NMR was not reported. However, its diastereomers and other alkaloids possessing the pyrrolo[2,1,5-de]quinolizine core have subsequently been isolated and characterized by NMR. *Yet most required revision*.

Crepidatumines C and D were recently isolated together with crepidamine from an orchid family plant *Dendrobium crepidatum* Lindl. ex. Paxt. As shown in Figure 8, crepidatumine D needed a revision to 1,5-epimer, which brought its stereochemistry in agreement with crepidine. Crepidatumine C showed significant mismatch between the computed and experimental NMR data, and is discussed below, in Figure 9.

Figure 8 also shows dendrocrepidines[43] which were recently isolated from another species of orchid, *Dendrobium crepidatum*. Stereoconfiguration at carbons C1, C5, and C9 illustrate a diversity of ring fusions in the crepidine family. Dendrocrepidine A required correction of stereoconfiguration at C6, making it a 3-epimer of crepidine.

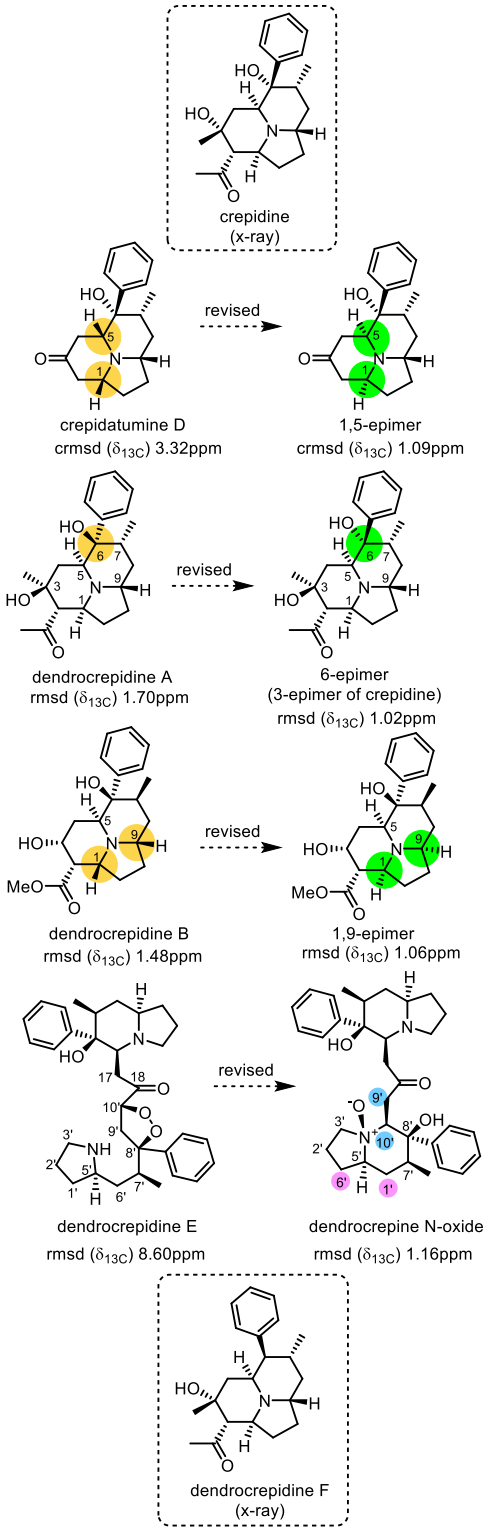


Figure 8. Revision of crepidine congeners.

DU8ML calculations also disagree with the original structure proposed for dendrocrepidine B, which we revised to the shown 1,9-epimer (here, and throughout the paper, relative stereochemistry is implied). The last revised alkaloid in Figure 8, dimeric dendrocrepidine E, was assigned an unusual 1,2-dioxolane structure, which also showed irreconcilable differences with our calculations. The spectra of dendrocrepidine E were reminiscent of dendrocrepine,[42b,44] except for the low field peaks suggesting that one of the monomeric units is an N-oxide. Further calculations supported this hypothesis. We revised the structure of dendrocrepidine E to the shown mono N-oxide of dendrocrepine, rmsd(δ_{13C})=1.16ppm.

In the dendrocrepidine F case, bottom of Figure 8, x-ray structures for both the (-)-enantiomer and the racemate were obtained, so the structure of dendrocrepidine F is unambiguously established. However, the published tabulated ^{13}C NMR chemical shift values are not in agreement with the x-ray structure, rmsd(δ_{13C})=3.84ppm. This illustrates another pitfall of the natural product research. The ^{13}C spectrum in the SI section of the original paper contains seven extra peaks, and it is not entirely clear how the reported values were selected for tabulation in the main manuscript.

Figure 9 addresses the special case of crepidatumine C. The most offending deviation was for C-1 listed at 74.8ppm, which hinted at a C-O, not C-N bond. Insertion of oxygen atom into the N-C1 bond gave the shown revised structure, rmsd(δ_{13C})=1.06ppm. Such perhydro-pyridooxazine cores are rare, but precedented in phyllanthidines[45] (see Supporting Information) and securingine B.[46] DU8ML gave an excellent match for both, Figure 9. However, another purported occurrence of the pyridooxazine core in asiaticumine B[47] did not withstand the scrutiny of DU8ML, which revealed misassignment, crmsd(δ_{13C})=4.77ppm

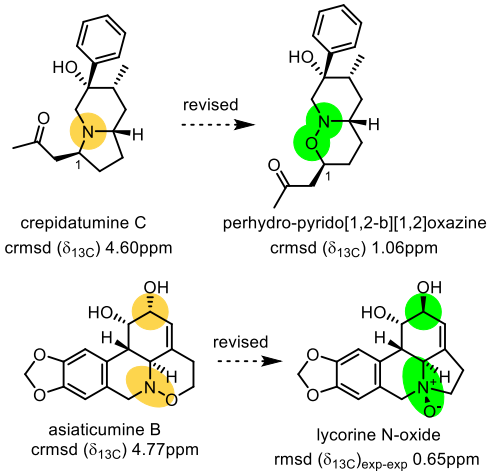


Figure 9. Revision of crepidatumine C and asiaticumine B.

Our calculations confirmed that asiaticumine B is in fact lycorine-N-oxide, a known alkaloid.[48] The experimental ^{13}C chemical shifts reported for lycorine-N-oxide in methanol matched the experimental data for asiaticumine B with rmsd$_{exp-exp}$=0.65ppm. It appears that the revised structure of crepidatumine C is the first example of the

perhydro-pyridooxazine core found among the crepidine congeners, and only a second alkaloid type which contains such core, in addition to securingins. Given a large number of alkaloids possessing an N-oxide moiety, it is only a matter of time before other families of pyrido-oxazine containing NPs are discovered.

Another particularly challenging NMR problem is the assignment of the N-oxide stereochemistry in complex alkaloids. As an example, for the vincanmine derivatives shown in Figure 10 – which were isolated from *Tabernaemontana bovina* in the same campaign as was the revised 10-methoxy-14α-hydroxymeloyunine above[33] – the authors did not attempt to assign stereoconfiguration of the N4-oxide. DU8ML allowed for differentiation between the *4α* and *4β* N-oxides with high confidence, consistently pointing to the *4α* stereoconfiguration at N4.

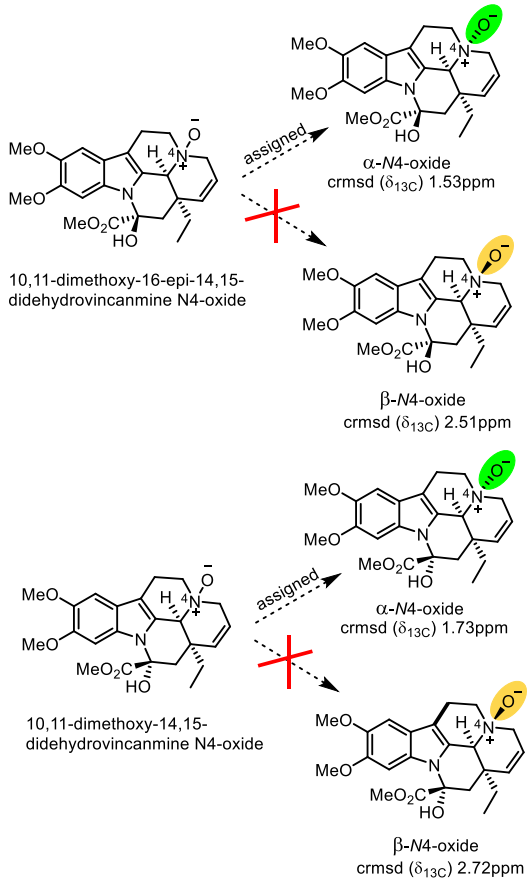


Figure 10. Stereoconfiguration assignment of N-oxides.

Complexity of alkaloids seemingly does not know limits. For example, rupestrisines A and B which were recently isolated from plant *Alstonia rupestris* and characterized as unprecedented yohimbine-kopsinine type heterodimeric indole alkaloids,[49] possess 42 carbon atoms with MW approaching 700 Daltons. Their solution structures were elucidated with the full assortment of modern NMR experiments augmented with computations of ECD spectra. The authors correctly deciphered all the atomic connectivities and assigned correct stereoconfiguration to most stereogenic centers. DU8ML computations corrected a small number of stereogenic centers as shown in Figure 11; i.e. rupestrisine A is revised to its 3'-epimer, and rupestrisine B is revised to its 3',15,16-epimer. Importantly, even for these large molecules, the computational (elapsed) time was 1.2-1.5 hours for each of the four conformers on a single node of a Linux cluster.

An unusual representation of stereochemistry around C3'-C14'-C15' bridge necessitated an additional check with a 3D drawing of rupestrisines in the original publication to orient the methylene C14' as the authors intended. Our drawing of this fragment in the revised structures helps to avoid confusion, provided one adheres to the prevalent mode of representation for the bicyclic moieties. Drawing uncertainties like this one are ubiquitous in the natural products literature. We refer everyone interested in honing their structure drawing skills to an excellent piece by Tantillo, appropriately titled Drawing Polycyclic Molecules[50] (see Supporting Information for more examples).

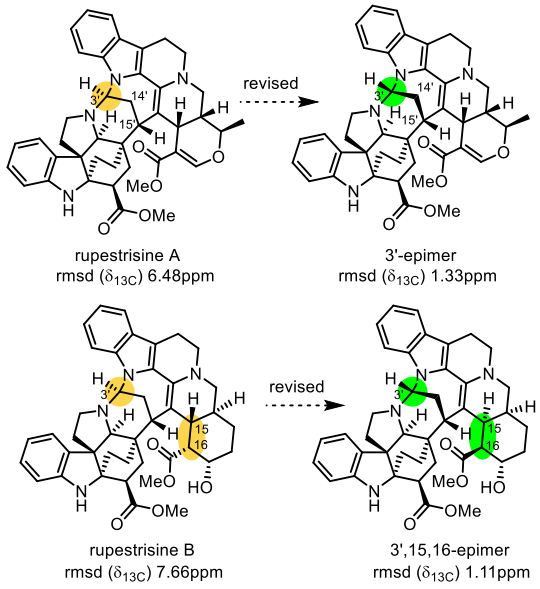


Figure 11. Stereogenic centers revision in rupestrisines A and B.

DU8ML predicts NMR properties of individual conformers exceptionally well. However, the structure optimizations are carried out at a *light* B3LYP/6-31G(d) level of DFT theory. This implies that one typical source of uncertainty in large, conformationally-flexible systems is the imperfect relative energies of conformers used to calculate their Boltzmann populations. For molecules possessing shallow minima on the potential energy surface (PES), this difficulty is particularly pronounced. Often, spin-spin coupling constants calculated for individual conformers differentiate significantly, providing an adequate dynamic range to refine the position of conformational equilibrium via J-fitting, i.e. fine-tuning of the conformer ratio to better match the calculated constants to the experimental ones.[6b] However, in many cases of proton-deficient NPs this is not an option. We suggest that as the accuracy and reliability of the ^{13}C calculations improves, one could use a similar δ-fitting procedure to refine the conformer ratios, as long as

such tweaking of the Boltzmann populations does not outstrip the energy calculation errors, expected for a given structure optimization method.[51] An instructive example of this is our revision of the structure of neuroprotective cyclopeptide cyctetryptomycin A isolated with its isomer, cyctetryptomycin B, from *Saccharopolyspora hirsuta* DSM 44795,[52] Figure 12. With crmsd(δ_{13C})=1.25ppm, DU8ML confirmed the structure of its congener, cyctetryptomycin B, where the "*south-east*" indole moiety was connected via N20 to carbon 18' of the "*north-east*" indoline. However, the calculated chemical shifts for the structure of cyctetryptomycin A were in a bad disagreement with the experimental data, crmsd(δ_{13C})=3.70ppm. The most obvious mismatch was for C19, where the experimental and computational data were more than 14 ppm apart. Another offending carbon was C22, $\Delta\delta_{exp-calc} > 9$. The HSQC data seem to relate C23 to a very narrow multiplet of H23, not consistent with the expected aromatic triplet and its calculated proton spin coupling constants J_{22-23}=8.2Hz and J_{23-24}=7.1Hz. We reassigned this narrow multiplet to H19, assuming a small splitting on the neighboring NH. The calculated ^{13}C chemical shift for C19 also supported this assignment. Cumulatively, these observations suggested that the connecting bond C19-C18' ought to be revised to C22-C18'.

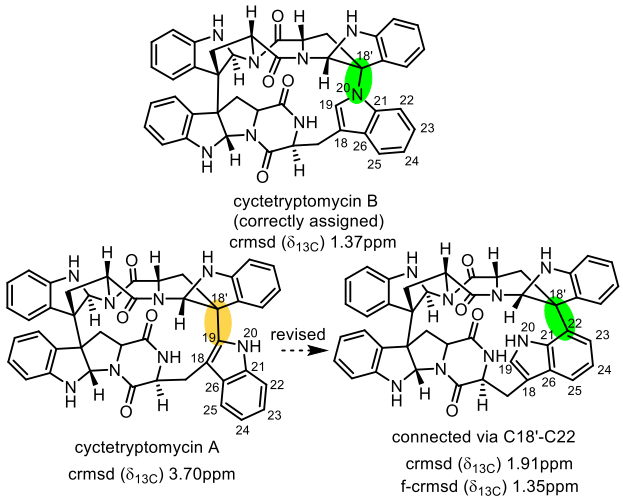


Figure 12. Structure revision of cyctetryptomycin A.

The revised structure gave much improved crmsd(δ_{13C})= 1.91ppm. Yet, additional fitting of the Boltzmann populations (the conformer ratio) of the two main conformers to better match the experimental data, gave f-crmsd(δ_{13C})=1.35ppm. We are now confident that the structure of cyctetryptomycin A should be revised to reflect the C22-C18' connection as shown in Figure 12.

Cyctetryptomycins A and B both have chemical formula $C_{44}H_{36}N_8O_4$, i.e. these are large molecules with MW=740 and yet, computational time was under 1.5 hours for the structure optimization and calculations of chemical shifts. Computations of Fermi contacts for accurate ^{1}H-^{1}H and ^{1}H-^{13}C nuclear spin-spin coupling constants added additional 30-35min per conformer. These practical times allowed for fast and confident validation of the structure of cyctetryptomycin B and revision of the structure cyctetryptomycin A.

CONCLUSIONS

DU8ML offers an exceptional cost-benefit balance, providing a quick and practical tool for solution structure elucidation. In this paper we demonstrated its effectiveness in structure revisions of 35 misassigned alkaloids and validation of many other correctly assigned structures. Our approach is not quite ready for human-free workflow of structure elucidation. The easy task, i.e. the detection and recognition of misassignments, could plausibly be human-free soon, although the lack of reliable fully automated programs for generation of conformers *in complex cyclic systems* – i.e. not for the freely rotatable bonds – remains challenging. Much more difficult task is finding the correct revision (or assignment of the correct structure in the first place). Here, CASE programs could offer help, by generating a subset of structures adhering to the constraints derived from the MS data and advanced NMR experiments, and further ranking them based on fast neural network algorithms designed to roughly evaluate chemical shifts.

"The last mile" in this journey currently belongs to DFT methods for the final judgment about differentiating and ranking the candidates, especially in the difficult situations of nearly-indistinguishable diastereomers. That is where the accuracy of calculations is of critical importance. ML-augmented DFT approaches are seemingly gaining momentum as a pragmatic compromise between, on the one hand, rigorous-yet-slow quantum chemistry methods and, on the other hand – the black box of NN/ML, most of the time correct, but no one could tell when (and why) erroneous. Obviously, DU8ML is our current goldilocks zone, and a large body of evidence presented in this paper hopefully supports our optimism.

ASSOCIATED CONTENT

Supporting Information. Computational details, cartesian coordinates for optimized structures, DFT energies, and the computed chemical shifts and SSCCs. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: akutatel@du.edu
ORCID: 0000-0003-3066-517X

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This research is supported by the NSF, CHE-1955892

Notes and references

(1) (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational Prediction of ^{1}H and ^{13}C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112* (3), 1839–1862. https://doi.org/10.1021/cr200106v. (b) see also Tantillo's excellent compilation of various NMR computational approaches at http://cheshirenmr.info (accessed 1/3/2022).

(2) (a) Elyashberg, M.; Argyropoulos, D. Computer Assisted Structure Elucidation (CASE): Current and Future Perspectives. *Magn. Reson. Chem.* **2021**, *59* (7), 669–690. https://doi.org/10.1002/mrc.5115. (b) Costa, F. L. P.; Albuquerque, A. C. F. de; Fiorot, R. G.; Lião, L. M.; Martorano, L. H.; Mota, G. V. S.; Valverde, A. L.; Carneiro, J. W. M.; Junior, F. M. dos S. Structural Characterisation of Natural Products by Means of Quantum Chemical Calculations of NMR Parameters: New Insights. *Org. Chem. Front.* **2021**, *8* (9), 2019–2058. https://doi.org/10.1039/D1QO00034A. (c) Guan, Y.; Sowndarya, S. V. S.; Gallegos, L. C.; St. John, P. C.; Paton, R. S. Real-time prediction of ^{1}H and ^{13}C chemical shifts with DFT accuracy using a 3D graph neural network. *Chem. Sci.* **2021**, *12*, 12012–12026. https://doi.org/10.1039/D1SC03343C (d) Buevich, A. V.; Elyashberg, M. E. Enhancing Computer-Assisted Structure Elucidation with DFT Analysis of J-Couplings. *Magn. Reson. Chem.* **2020**, *58* (6), 594–606. https://doi.org/10.1002/mrc.4996. (d) Koos, M. R. M.; Navarro-Vázquez, A.; Anklin, C.; Gil, R. R. Computer-Assisted 3D Structure Elucidation (CASE-3D): The Structural Value of 2JCH in Addition to 3JCH Coupling Constants. *Angew. Chem. Int. Ed.* **2020**, *59* (10), 3938–3941. https://doi.org/10.1002/anie.201915103. (e) Elyashberg, M. E.; Williams, A. J.; Martin, G. E. Computer-Assisted Structure Verification and Elucidation Tools in NMR-Based Structure Elucidation. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53* (1), 1–104. https://doi.org/10.1016/j.pnmrs.2007.04.003. (f) Elyashberg, M.; Williams, A. J.; Blinov, K. Structural Revisions of Natural Products by Computer-Assisted Structure Elucidation (CASE) Systems. *Nat. Prod. Rep.* **2010**, *27* (9), 1296–1328. https://doi.org/10.1039/C002332A. (g) Reher, R.; Kim, H. W.; Zhang, C.; Mao, H. H.; Wang, M.; Nothias, L.-F.; Caraballo-Rodriguez, A. M.; Glukhov, E.; Teke, B.; Leao, T.; Alexander, K. L.; Duggan, B. M.; Van Everbroeck, E. L.; Dorrestein, P. C.; Cottrell, G. W.; Gerwick, W. H. A Convolutional Neural Network-Based Approach for the Rapid Annotation of Molecularly Diverse Natural Products. *J. Am. Chem. Soc.* **2020**, *142* (9), 4114–4120. https://doi.org/10.1021/jacs.9b13786.

(3) Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L. Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6–31G* and MP2/6–31G* Harmonic Frequencies. *Isr. J. Chem.* **1993**, *33* (3), 345–350. https://doi.org/10.1002/ijch.199300041.

(4) https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed 1/3/2022).

(5) Bally, T.; Rablen, P. R. Quantum-Chemical Simulation of ^{1}H NMR Spectra. 2. Comparison of DFT-Based Procedures for Computing Proton–Proton Coupling Constants in Organic Molecules. *J. Org. Chem.* **2011**, *76* (12), 4818–4830. https://doi.org/10.1021/jo200513q.

(6) (a) Kutateladze, A. G.; Mukhina, O. A. Relativistic Force Field: Parametric Computations of Proton–Proton Coupling Constants in ^{1}H NMR Spectra. *J. Org. Chem.* **2014**, *79* (17), 8397–8406. https://doi.org/10.1021/jo501781b. (b) Kutateladze, A. G.; Mukhina, O. A. Relativistic Force Field: Parametrization of ^{13}C–^{1}H Nuclear Spin–Spin Coupling Constants. *J. Org. Chem.* **2015**, *80* (21), 10838–10848. https://doi.org/10.1021/acs.joc.5b02001.

(7) Kutateladze, A. G.; Reddy, D. S. High-Throughput in Silico Structure Validation and Revision of Halogenated Natural Products Is Enabled by Parametric Corrections to DFT-Computed ^{13}C NMR Chemical Shifts and Spin–Spin Coupling Constants. *J. Org. Chem.* **2017**, *82* (7), 3368–3381. https://doi.org/10.1021/acs.joc.7b00188.

(8) Braddock, D. C.; Rzepa, H. S. Structural Reassignment of Obtusallenes V, VI, and VII by GIAO-Based Density Functional Prediction. *J. Nat. Prod.* **2008**, *71* (4), 728–730. https://doi.org/10.1021/np0705918.

(9) Xin, D.; Sader, C. A.; Chaudhary, O.; Jones, P.-J.; Wagner, K.; Tautermann, C. S.; Yang, Z.; Busacca, C. A.; Saraceno, R. A.; Fandrick, K. R.; Gonnella, N. C.; Horspool, K.; Hansen, G.; Senanayake, C. H. Development of a ^{13}C NMR Chemical Shift Prediction Procedure Using B3LYP/Cc-PVDZ and Empirically Derived Systematic Error Correction Terms: A Computational Small Molecule Structure Elucidation Method. *J. Org. Chem.* **2017**, *82* (10), 5135–5145. https://doi.org/10.1021/acs.joc.7b00321.

(10) (a) Gao, P.; Zhang, J.; Peng, Q.; Zhang, J.; Glezakou, V.-A. General Protocol for the Accurate Prediction of Molecular 13C/1H NMR Chemical Shifts via Machine Learning Augmented DFT. *J. Chem. Inf. Model.* **2020**, *60* (8), 3746–3754. https://doi.org/10.1021/acs.jcim.0c00388. (b) notice that Zhang and co-workers use QSAR-type descriptors, to ascertain the chemical environment, by parsing SMILES codes using RDKIT. The descriptors include atomic number, Gasteiger charge, total valence, size of the minimal ring, logP values, molar refractivity etc. Our approach is based solely on truncated substructures, which are shown empirically to cause errors in computed NMR values. For example, for ^{13}C chemical shifts, as the training set grows, these corrections could increase or decrease in magnitude, with the latest count of approximately 160 SMARTS strings contributing 1 ppm or greater as an ML-driven correction to the DFT computed ^{13}C shifts. For more details – see Supporting Information).

(11) (a) Unzueta, P. A.; Greenwell, C. S.; Beran, G. J. O. Predicting Density Functional Theory-Quality Nuclear Magnetic Resonance Chemical Shifts via Δ-Machine Learning. *J. Chem. Theory Comput.* **2021**, *17* (2), 826–840. https://doi.org/10.1021/acs.jctc.0c00979. (b) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015**, *6* (16), 3309–3313. https://doi.org/10.1021/acs.jpclett.5b01456.

(12) Kutateladze, A. G.; Kuznetsov, D. M.; Beloglazkina, A. A.; Holt, T. Addressing the Challenges of Structure Elucidation in Natural Products Possessing the Oxirane Moiety. *J. Org. Chem.* **2018**, *83* (15), 8341–8352. https://doi.org/10.1021/acs.joc.8b01027.

(13) Kutateladze, A. G.; Holt, T.; Reddy, D. S. Natural Products Containing the Oxetane and Related Moieties Present Additional Challenges for Structure Elucidation: A DU8+ Computational Case Study. *J. Org. Chem.* **2019**, *84* (12), 7575–7586. https://doi.org/10.1021/acs.joc.9b01005.

(14) Kutateladze, A. G.; Kuznetsov, D. M. Triquinanes and Related Sesquiterpenes Revisited Computationally: Structure Corrections of Hirsutanols B and D, Hirsutenol E, Cucumin B, Antrodins C–E, Chondroterpenes A and H, Chondrosterins C and E, Dichrocephone A, and Pethybrene. *J. Org. Chem.* **2017**, *82* (20), 10795–10802. https://doi.org/10.1021/acs.joc.7b02018.

(15) Novitskiy, I. M.; Kutateladze, A. G. DU8+ Computations Reveal a Common Challenge in the Structure Assignment of Natural Products Containing a Carboxylic Anhydride Moiety. *J. Org. Chem.* **2021**, *86* (23), 17511–17515. https://doi.org/10.1021/acs.joc.1c02291.

(16) (a) Kutateladze, A. G.; Krenske, E. H.; Williams, C. M. Reassignments and Corroborations of Oxo-Bridged Natural Products Directed by OSE and DU8+ NMR Computation. *Angew. Chem. Int. Ed.* **2019**, *58* (21), 7107–7112. https://doi.org/10.1002/anie.201902777. (b) Kutateladze, A. G.; Holt, T. Structure Validation of Complex Natural Products: Time to Change the Paradigm. What Did Synthesis of Alstofolinine A Prove? *J. Org. Chem.* **2019**, *84* (12), 8297–8299. https://doi.org/10.1021/acs.joc.9b00969.
(c) Holt, T. A.; Reddy, D. S.; Huple, D. B.; West, L. M.; Rodríguez, A. D.; Crimmins, M. T.; Kutateladze, A. G. The Discreet Structural Diversity of Briarellins: DU8+ Guided Multiple Structure Revisions Yielded Two Unknown Structural Types. *J. Org. Chem.* **2020**, *85* (9), 6201–6205. https://doi.org/10.1021/acs.joc.0c00555.

(17) Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L. Chemical Shifts in Molecular Solids by Machine Learning. *Nat Commun* **2018**, *9* (1), 4501. https://doi.org/10.1038/s41467-018-06972-x.

(18) Howarth, A.; Goodman, J. DFT, Molecular Mechanics and Raw NMR Data for: The DP5 Probability, Automatic Quantification and Visualization of Structural Uncertainty in Single Molecules. *Apollo*, 2021. https://doi.org/10.17863/CAM.65145.

(19) (a) Souza, C. R. M.; Bezerra, W. P.; Souto, J. T. Marine Alkaloids with Anti-Inflammatory Activity: Current Knowledge and Future Perspectives. *Mar. Drugs* **2020**, *18* (3), 147. https://doi.org/10.3390/md18030147. (b) Dyshlovoy, S. A. Recent Updates on Marine Cancer-Preventive Compounds. *Mar Drugs* **2021**, *19* (10), 558. https://doi.org/10.3390/md19100558. (c) Shen, Y.; Liang, W.-J.; Shi, Y.-N.; Kennelly, E. J.; Zhao, D.-K. Structural Diversity, Bioactivities, and Biosynthesis of Natural Diterpenoid Alkaloids. *Nat. Prod. Rep.* **2020**, *37* (6), 763–796. https://doi.org/10.1039/D0NP00002G.

(20) for a recent instructive example see Deng, Y.; Liang, X.; Wei, K.; Yang, Y.-R. Ir-Catalyzed Asymmetric Total Syntheses of Bisdehydrotuberostemonine D, Putative Bisdehydrotuberostemonine E and Structural Revision of the Latter. *J. Am. Chem. Soc.* **2021**, *143* (49), 20622–20627. https://doi.org/10.1021/jacs.1c11265.

(21) Cai, Y.-S.; Sun, J.-Z.; Tang, Q.-Q.; Fan, F.; Guo, Y.-W. Acanthiline A, a Pyrido[1,2-a]Indole Alkaloid from Chinese Mangrove Acanthus Ilicifolius. *Journal of Asian Natural Products Research* **2018**, *20* (11), 1088–1092. https://doi.org/10.1080/10286020.2018.1488834.

(22) Feng, Q.-T.; Zhu, G.-Y.; Gao, W.-N.; Yang, Z.; Zhong, N.; Wang, J.-R.; Jiang, Z.-H. Two New Alkaloids from the Roots of *Baphicacanthus Cusia*. *Chem. Pharm. Bull.* **2016**, *64* (10), 1505–1508. https://doi.org/10.1248/cpb.c16-00315.

(23) Chen, S.-C.; Liu, Z.-M.; Tan, H.-B.; Chen, Y.-C.; Li, S.-N.; Li, H.-H.; Guo, H.; Zhu, S.; Liu, H.-X.; Zhang, W.-M. Tersone A-G, New Pyridone Alkaloids from the Deep-Sea Fungus Phomopsis Tersa. *Mar. Drugs* **2019**, *17* (7), 394. https://doi.org/10.3390/md17070394.

(24) Fukuda, T.; Tomoda, H.; Ōmura, S. Citridones, New Potentiators of Antifungal Miconazole Activity, Produced by Penicillium Sp. FKI-1938. *J Antibiot* **2005**, *58* (5), 315–321. https://doi.org/10.1038/ja.2005.39.

(25) Cao, F.; Pan, L.; Gao, W.; Liu, Y.; Zheng, C.; Zhang, Y. Structure Revision and Protein Tyrosine Phosphatase Inhibitory Activity of Drazepinone. *Marine Drugs* **2021**, *19* (12), 714. https://doi.org/10.3390/md19120714.

(26) Jiang, C.-X.; Yu, B.; Miao, Y.-M.; Ren, H.; Xu, Q.; Zhao, C.; Tian, L.-L.; Yu, Z.-Q.; Zhou, P.-P.; Wang, X.; Fang, J.; Zhang, J.; Zhang, J. Z.; Wu, Q.-X. Indole Alkaloids from a Soil-Derived Clonostachys Rosea. *J. Nat. Prod.* **2021**, *84* (9), 2468–2474. https://doi.org/10.1021/acs.jnatprod.1c00457.

(27) Jiao, W.-H.; Li, J.; Wang, D.; Zhang, M.-M.; Liu, L.-Y.; Sun, F.; Li, J.-Y.; Capon, R. J.; Lin, H.-W. Cinerols, Nitrogenous Meroterpenoids from the Marine Sponge Dysidea Cinerea. *J. Nat. Prod.* **2019**, *82* (9), 2586–2593. https://doi.org/10.1021/acs.jnatprod.9b00471.

(28) Fan, N.-W.; Chang, H.-S.; Cheng, M.-J.; Hsieh, S.-Y.; Liu, T.-W.; Yuan, G.-F.; Chen, I.-S. Secondary Metabolites from the Endophytic Fungus Xylaria Cubensis. *Helvetica Chimica Acta* **2014**, *97* (12), 1689–1699. https://doi.org/10.1002/hlca.201400091.

(29) Zhao, Z.-Z.; Liang, X.-B.; Feng, W.-S.; Xue, G.-M.; Si, Y.-Y.; Chen, H.-P.; Liu, J.-K. Cyclic Dipeptides with Peroxy Groups from the Fruiting Bodies of the Edible Mushroom Tricholoma Matsutake. *Tetrahedron Lett.* **2020**, *61* (21), 151892. https://doi.org/10.1016/j.tetlet.2020.151892.

(30) Chan, Z.-Y.; Krishnan, P.; Modaresi, S. M.; Hii, L.-W.; Mai, C.-W.; Lim, W.-M.; Leong, C.-O.; Low, Y.-Y.; Wong, S.-K.; Yong, K.-T.; Leong, A. Z.-X.; Lee, M.-K.; Ting, K.-N.; Lim, K.-H. Monomeric, Dimeric, and Trimeric Tropane Alkaloids from Pellacalyx Saccardianus. *J. Nat. Prod.* **2021**, *84* (8), 2272–2281. https://doi.org/10.1021/acs.jnatprod.1c00374.

(31) Youssef, D. T. A.; Asfour, H. Z.; Genta-Jouve, G.; Shaala, L. A. Magnificines A and B, Antimicrobial Marine Alkaloids Featuring a Tetrahydrooxazolo[3,2-a]Azepine-2,5(3H,6H)-Dione Backbone from the Red Sea Sponge Negombata Magnifica. *Marine Drugs* **2021**, *19* (4), 214. https://doi.org/10.3390/md19040214.

(32) Chen, F.-Z.; Chen, D.-L.; Chen, Q.-H.; Wang, F.-P. Diterpenoid Alkaloids from Delphinium Majus. *J. Nat. Prod.* **2009**, *72* (1), 18–23. https://doi.org/10.1021/np800439a.

(33) Yu, Y.; Bao, M.-F.; Huang, S.-Z.; Wu, J.; Cai, X.-H. Vincan- and Eburnan-Type Alkaloids from Tabernaemontana Bovina and Their Hypoglycemic Activity. *Phytochem.* **2021**, *190*, 112859. https://doi.org/10.1016/j.phytochem.2021.112859.

(34) Cai, X.-H.; Li, Y.; Liu, Y.-P.; Li, X.-N.; Bao, M.-F.; Luo, X.-D. Alkaloids from Melodinus Yunnanensis. *Phytochem.* **2012**, *83*, 116–124. https://doi.org/10.1016/j.phytochem.2012.06.013.

(35) Pu, X.-J.; Hu, Q.-Y.; Li, S.-S.; Li, G.-H.; Zhao, P.-J. Sesquiterpenoids and Their Quaternary Ammonium Hybrids from the Mycelium of Mushroom Stereum Hirsutum by Medium Optimization. *Phytochem.* **2021**, *189*, 112852. https://doi.org/10.1016/j.phytochem.2021.112852.

(36) Yu, J.-Q.; Sun, X.-W.; Wang, Z.-W.; Fang, L.; Wang, X. Alkaloids from Melodinus Henryi with Anti-Inflammatory Activity. *Journal of Asian Natural Products Research* **2019**, *21* (8), 820–825. https://doi.org/10.1080/10286020.2018.1482878.

(37) Han, X.; Luo, X.; Xue, L.; van Ofwegen, L.; Zhang, W.; Liu, K.; Zhang, Y.; Tang, X.; Li, P.; Li, G. Dolabellane Diterpenes and Elemane Alkaloids from the Soft Coral Clavularia Inflata Collected in the South China Sea. *J. Nat. Prod.* **2022**, *85*, 276–283. https://doi.org/10.1021/acs.jnatprod.1c01103.

(38) Kitajima, M.; Nakazawa, M.; Wu, Y.; Kogure, N.; Zhang, R.-P.; Takayama, H. Kopsiyunnanines L and M, Strychnos-Related Monoterpenoid Indole Alkaloids from Yunnan Kopsia Arborea. *Tetrahedron* **2016**, *72* (42), 6692–6696. https://doi.org/10.1016/j.tet.2016.08.082.

(39) We are not entirely certain how the C20 indole alkaloids came to be broadly known as "MIAs"- monoterpenoid indole alkaloids.

(40) Jiang, G.-Y.; Qin, L.-L.; Gao, F.; Huang, S.; Zhou, X.-L. Fifteen New Diterpenoid Alkaloids from the Roots of Aconitum Kirinense Nakai. *Fitoterapia* **2020**, *141*, 104477. https://doi.org/10.1016/j.fitote.2020.104477.

(41) Li, Y.-L.; Zhang, Y.; Zhao, P.-Z.; Hu, Z.-X.; Gu, Y.-C.; Ye, J.; Hao, X.-J. Two New Steroidal Alkaloids from the Rhizomes of Veratrum Nigrum L. and Their Anti-TYLCV Activity. *Fitoterapia* **2020**, *147*, 104731. https://doi.org/10.1016/j.fitote.2020.104731.

(42) (a) Kierkegaard, P.; Pilotti, A.-M.; Leander, K. Studies on Orchidaceae Alkaloids. XX. The Constitution and Relative Configuration of Crepidine, an Alkaloid from Dendrobium Crepidatum Lindl. *Acta Chem. Scand.* **1970**, *24*, 3757–3759. https://doi.org/10.3891/acta.chem.scand.24-3757. (b) Elander, M.; Leander, K.; Rosenblom, J.; Ruusa, E. Studies on Orchidaceae Alkaloids. XXXII. Crepidine, Crepidamine and Dendrocrepine, Three Alkaloids from Dendrobium Crepidatum Lindl. *Acta Chem. Scand.* **1973**, *27*, 1907–1913. https://doi.org/10.3891/acta.chem.scand.27-1907. (c) Xu, X.; Li, Z.; Yang, R.; Zhou, H.; Bai, Y.; Yu, M.; Ding, G.; Li, B. Crepidatumines C and D, Two New Indolizidine Alkaloids from Dendrobium Crepidatum Lindl. Ex Paxt. *Molecules* **2019**, *24* (17), 3071. https://doi.org/10.3390/molecules24173071.

(43) Hu, Y.; Ren, J.; Wang, L.; Zhao, X.; Zhang, M.; Shimizu, K.; Zhang, C. Protective Effects of Total Alkaloids from Dendrobium Crepidatum against LPS-Induced Acute Lung Injury in Mice and Its Chemical Components. *Phytochem.* **2018**, *149*, 12–23. https://doi.org/10.1016/j.phytochem.2018.02.006.

(44) Leete, E.; Riddle, R. M. Steric Inhibition of the Rotation of the Phenyl Groups in 2,6-Dimethyl-1-Phenylcyclohexanol and Dendrocrepine Detected by Means of Carbon-13 Nuclear Magnetic Resonance. *Tetrahedron Lett.* **1978**, *19* (52), 5163–5166. https://doi.org/10.1016/S0040-4039(01)85839-1.

(45) (a) Horii, Z.; Imanishi, T.; Yamauchi, M.; Hanaoka, M.; Parello, J.; Munavalli, S. Structure of Phyllantidine. *Tetrahedron Letters* **1972**, *13* (19), 1877–1880. https://doi.org/10.1016/S0040-4039(01)84740-7. (b) Lajis, N. H.; Guan, O. B.; Sargent, M. V.; Skelton, B. W.; White, A. H. Viroallosecurinine and Ent-Phyllanthidine From the Leaves of Breynia Coronata (Euphorbiaceae). *Aust. J. Chem.* **1992**, *45* (11), 1893–1897. https://doi.org/10.1071/CH9921893. (c) Yuan, W.; Zhu, P.; Cheng, K.; Meng, C.; Ma, L.; Wu, F.; Zhu, H. Callus of Securinega Suffruticosa , a Cell Line Accumulates Dextro Securinega Alkaloids. *Nat. Prod. Res.* **2007**, *21* (3), 234–242. https://doi.org/10.1080/14786410701189781.

(46) Park, K. J.; Kim, C. S.; Khan, Z.; Oh, J.; Kim, S. Y.; Choi, S. U.; Lee, K. R. Securinega Alkaloids from the Twigs of Securinega Suffruticosa and Their Biological Activities. *J. Nat. Prod.* **2019**, *82* (5), 1345–1353. https://doi.org/10.1021/acs.jnatprod.9b00142.

(47) Sun, Q.; Shen, Y.-H.; Tian, J.-M.; Tang, J.; Su, J.; Liu, R.-H.; Li, H.-L.; Xu, X.-K.; Zhang, W.-D. Chemical Constituents of Crinum Asiaticum L. Var. Sinicum Baker and Their Cytotoxic Activities. *Chem. Biodiversity* **2009**, *6* (10), 1751–1757. https://doi.org/10.1002/cbdv.200800273.

(48) Sarikaya, B. B.; Kaya, G. I.; Onur, M. A.; Viladomat, F.; Codina, C.; Bastida, J.; Somer, N. U. Alkaloids from Galanthus Rizehensis. *Phytochem. Lett.* **2012**, *5* (2), 367–370. https://doi.org/10.1016/j.phytol.2012.03.004.

(49) Wang, Z.-W.; Zhang, J.-P.; Wei, Q.-H.; Chen, L.; Lin, Y.-L.; Wang, Y.-L.; An, T.; Wang, X.-J. Rupestrisine A and B, Two Novel Dimeric Indole Alkaloids from Alstonia Rupestris. *Tetrahedron Lett.* **2021**, *87*, 153525. https://doi.org/10.1016/j.tetlet.2021.153525.

(50) Tantillo, D. J. Drawing Polycyclic Molecules. *ACS Omega* **2021**, *6* (36), 23008–23014. https://doi.org/10.1021/acsomega.1c03607.

(51) For more examples of this 'complication' and also other challenges, such as partial protonation of basic alkaloids in NMR experiments (ostensibly, chloroform containing residual hydrochloric acid, etc.) see Supporting Information.

(52) Malit, J. J. L.; Liu, W.; Cheng, A.; Saha, S.; Liu, L.-L.; Qian, P.-Y. Global Genome Mining Reveals a Cytochrome P450-Catalyzed Cyclization of Crownlike Cyclodipeptides with Neuroprotective Activity. *Org. Lett.* **2021**, *23* (17), 6601–6605. https://doi.org/10.1021/acs.orglett.1c01022.