A Study of Structures, Mechanisms, and Theories through Computational Chemistry

By

KA YI TSUI

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Dean Tantillo, Chair

_____

Neil Schore

_____

Lee-Ping Wang

Committee in Charge

2021

TABLE OF CONTENT

List of Schemes

List of Figures

List of Tables

Acknowledgement

The author is thankful to her mentor and committee chair, Dr. Dean Tantillo, for his academic guidance. Dr. Dean Tantillo has been an aspiration to the author's view in physical organic chemistry. It has been the author's pleasure to work with him.

The author is also thankful to the rest of her committee members Dr. Neil Schore and Dr. Lee Ping Wang. Dr. Schore was a supportive committee head for the author's qualification exam. Dr. Wang was also a committee member. They have provided the author helpful suggestions and feedbacks. The author thanks them for their time and willingness to be a part of this dissertation.

The author also thanks her collaborators and coworkers. They have provided the author meaningful discussions and helped the author to further her understanding in quantum chemistry.

Abstract

The purpose of this dissertation is to showcase the use of computational chemistry to understand and corroborate chemical observations from experiments.

This report will begin with highlighting the importance of considering the conformational space of an organic molecule. Instead of a static picture, molecules have degrees of freedom in space and they can equilibrate between different conformations. This flexibility can affect a molecule's reactivity and spectral properties. Chapter 2 and 3 focus the study of structures. Here, we used computational data to correlate to experimental observations, and subsequently guide decisions on synthesis.

Then, we move from the study of structures to mechanisms, highlighting the ability of using quantum chemistry to compare reaction energetics. For example, using transition state theory (TST), we can predict if a chemical transformation is kinetically feasible and which product is favored under the experimental conditions.

The final chapter describe a purely computational project which considers the two major assumptions involved in the previous chapters: 1) a chemical reaction obeys conventional TST; 2) the inclusion of implicit solvent environment is sufficient to simulate realistic solvent effect. This chapter will consider the limitations of these two assumptions using a bifurcating pericyclic model reaction, in which the inclusion of explicit solvent molecule affects product selectivity.

Chapter 1: Introduction

1.1 Quantum Chemistry

Atoms, the smallest unit of matter, is composed of a nucleus and one or more electrons. The whereabout and presence or lack of an electron is critical to how an atom or a molecule behaves, specifically in the context of organic chemistry, how organic compounds react with each other. However, these behaviors are not definitive. For example, in our macroscopic world, objects appear at a certain time and location and their motions obey the classical Newton's laws of mechanics. However, the same cannot be said when we zoom into nanoscale or below. Depending on the detection methods, a particle can also exhibit its wave character. For example, an electron, the capsule of electricity can be represented by a wavefunction, Ψ.[1,2] The Schrödinger equation is a linear partial differential equation that solves the wavefunction of a system, whose solution describes the quantized properties of a particle.[1-3] It is simplified as below:

$$H\Psi = E\Psi$$

The Hamiltonian (H) is an operator that takes functions to functions. Here it corresponds to the total energy of the system of interest, including the kinetic and potential energies. The wavefunction (Ψ) serves as the eigenvector, which when H operates on, to generate the eigenvalue E, the energy value of the system. When Ψ is squared, it represents the spatial probability of an electron for a single-electron system. For a many-electron system, however, it is more complicated because it is impossible to solve the wavefunction analytically. Therefore, computational methods have been developed to approximate the Hamiltonian for many-electron systems, such as the Hartree-Fock (HF) method[4] and the Linear Combination of Atomic Orbitals (LCAO)[5] procedure. The HF method can approximate the determination of the wavefunction and the energy of a many-electron system in a stationary state. By invoking the variational method, the solution of these equations yields the HF wavefunction and the energy of the system. The HF

method is a Self-Consistent Field (SCF) method, that the solution is required to be self-consistent with the initial field. Linear Combination of Atomic Orbitals (LCAO) is also a SCF technique, for calculating molecular orbitals. Here, the wavefunctions are the basis set of functions, which describe the electrons of an individual atom. Using this method, we can determine which portions of individual atomic orbitals, through quantum superposition provide molecular orbitals with the lowest total energy. However, it is important to note that during the steps of SCF calculation, the nuclei positions stay unchanged since the Born-Oppenheimer approximation is applied for separating the nuclear and electronic components of the Hamiltonian.

Throughout this report, to understand electronic structures and simulatmolecular dynamics, density-functional theory (DFT) methods are applied. Using functionals of spatially dependent electron densities, DFT allows us to compute properties of many-body systems. DFT is originated from two Hohenberg-Kohn (HK) theorems.[6] According to the first HK theorem, the ground-state properties (such as total energy) of a many-electron system can be determined by an electron density, which depends only on its three spatial coordinates. The second HK theorem defines an energy functional for a system which is minimized by its true ground-state electron density, i.e., the energy of the Hamiltonian reaches its absolute minimum. DFT has become a very popular and important tool in chemical calculations due to its relatively low computational costs and versatility.

In this report, we focus on computationally studying the minima and transition states of organic molecules during chemical reactions. This report will begin with conformational analyses of flexible organic molecules and how the thermodynamic stability among conformer can affect the compound's structural-activity relationship and spectral properties, especially for the case of nuclear magnetic resonance (NMR). Then, we will shift to the application of transition state theory (TST) to elucidate mechanisms for organic chemical reactions, highlighting carbocation and carbene chemistries. Chapter 2 to 5 demonstrated the use of computational chemistry to obtain

information such as predicted NMR shifts and reaction energetics to assist experiment. Here, we highlight the indispensability of computational chemistry since can provide information difficult to obtain from experiment and hence correlating experimental observational to guide decisions in synthesis.

However, there are challenges involved in using computational methods to study chemical reactions, such as choosing the appropriate method for a certain system and the computational cost if expensive methods were used. The assumptions we make when studying a chemical reaction can also be problematic. The final chapter of this report presents a pure theoretical project, which we question some of the assumptions previously involved. Our first major assumption involves 1) the reactions follow traditional transition state theory and 2) the use of implicit solvent environment is sufficient to describe realistic solvent effect.

In conclusion, this report showcases computational applications to solve different types of chemical problems, with a primarily focus on structure and reactivity. However, we also question the assumption involved in our computational applications and present a model reaction where these assumptions were challenged.

1.2 Transition State Theory

Conventional (nonvariational) transition state theory (TST) is applicable for calculating reactions rates for electronically adiabatic reactions in gas phase or condensed phases.[7-9] According to TST, the reaction rate of an elementary step can be studied by examining the transition state structure (TSS) which located near/on the saddle point of a potential energy surface. Here, reactant molecules are assumed to be in quasi-equilibrium with the transition state TSSs, which then convert into products. In order to assess the kinetic feasibility of a certain conversion, the activation energy is calculated, which equals to the average energy of all reactant

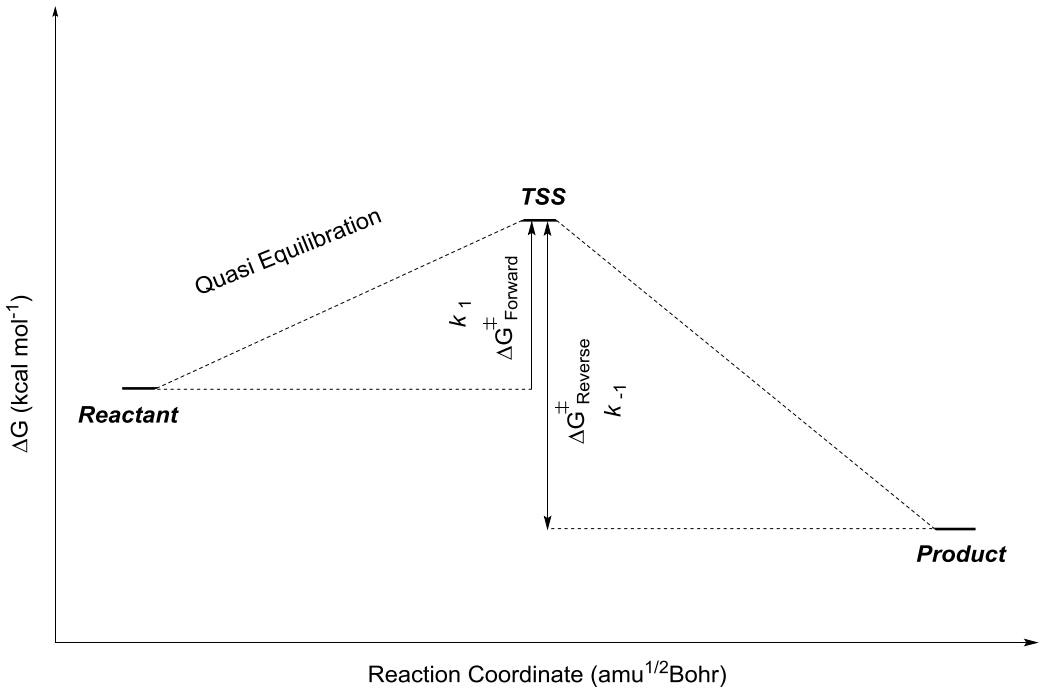


Figure 1. A generic free energy diagram. Y-axis is free energy in kcal mol^{-1} and X-axis is the reaction coordinate. The reactant and TSS are assumed to be in quasi equilibrium. Given the activation energy, reaction rate for a reaction in the forward or reverse direction can be calculated using the Eyring equation.

molecules subtracted by the average energy of all molecules at the transition state (Figure. 1). After obtaining the activation energy, we can solve the rate of reaction using the Eyring equation.

Computationally, this ties into our mechanistic studies where we performed stationary-point analyses to identify the TSS and its connecting minimum energy structures for a proposed mechanism. Applying TST, we will be able to predict whether a mechanism is kinetically favored under given experimental conditions. In the case of competing reactions under kinetic conditions, we can also predict chemical selectivity based on their activation barriers. Collaborating with experiments, this technique allows us to approximate reaction rate of a chemical reaction and product selectivity before the actual syntheses are performed, hence we can provide guidance to synthetic decision.

Traditionally, TST is applicable only for reactions that occur on a single Born-Oppenheimer potential energy surface. However, variational TST (VTST), has also been developed, combining with modern electronic structure theory, multi-dimensional tunneling methods, and statistical and quantum theory for anharmonicity.[8-21] VTST serves as a more powerful tool in predicting reactions rates and mechanisms.

In this report, chapter 4 and 5 present chemical reactions where conventional TST is applicable. However, the final chapter showcases a bifurcating reaction, where TST fails to predict product selectivity.

1.3 Implicit Solvation Models

Interactions of a solute with the solvent affect the energies of the ground and excited state of the molecule, generally due to long-range electrostatic interactions and/or short-range contributions. Implicit solvent models are commonly used in computational chemistry to calculate free energy of solvation. It describes a solvated molecule with atomistic detail placed in a molecule-sized electrostatic cavity surrounded by a dielectric medium that depicts the solvent.[22-25] The electrostatic component of the solvent effect can then be accounted from the solute with surrounding polarizable medium. Common continuum solvent models used in this report are the polarizable continuum model (PCM)[23,24] and its variations, such as the solvation model based on density (SMD).[25]

Although universally used and computationally inexpensive, they fail to account for specific short-range solvent-solute interactions, such as dispersion. There is also an artificial boundary between solvent and solute when only implicit solvent environment is included. On contrary, explicit solvent inclusions, either modelling the solvent fully atomistic or partly atomistic with an empirical mean field (QM/MM), are more computationally intensive. However, when

solvent-specific effects are important, explicit solvation models provide a more realistic physical description for studying solvation.[26]

In this report, implicit solvents are used to calculate solvation free energies in chapters 2 to 5. On contrary, chapter 6 showcases the systematic inclusion of explicit solvent molecules in our systems and its effects on product selectivity.

Reference

1) Trixler, F., *Curr. Org. Chem*., **2013**, *17*, 1758-1770

2) de Broglie, L., *Nobel Lecture*, **1929**, *12*, 244-256

3) Schrödinger, E., *Am. Phys.*, **1926**, 79, 361-376

4) Hartree, D. R., *Math. Proc. Camb. Philos. Soc.*, **1928**, *24*, *1*, 111

5) Mulliken, R. S., *Nobel Lecture*, *Science*, **1967**, *157*, *3784*, 13-24

6) Hohenberg, P., and Kohn, W., *Physical Review.*, **1964**, *136* (3B), B864–B871

7) Glasstone, S., Laidler, K., and Eyring, H.,*The theory of Rate Processes*, McGraw-Hill, New York, **1941**

8) Truhlar, D. G., Garrett, B. C., *Am. Rev. Phys. Chem.,* **1984**, *35*, 159-89

9) E. Wigner, *J. Chem. Phys.*, **1937**, *5*, 720

10) Bao, J.L., Truhlar, D. G, *Chem. Soc. Rev*., **2017**, 46, 7548-7596

11) Horiuti, J., *Bull. Chem. Soc*. *Jpn.*, **1938**, *13*, 210

12) J. C, Keck, *J. Chem. Phys.*, **1960**, *32*, 1035

13) J. C, Keck, *Adv. Chem. Phys.*, **1967**, *13*, 85

14) Garrett, B. C., Truhlar, D. G., *J. Chem. Phys.*, **1979**, *70*, 1593

15) Garrett, B. C., Truhlar, D. G., *J. Phys. Chem.*, **1979**, *83*, 1052

16) Truhlar, D. G., Garrett, B. C., *Acc. Chem.Res.,* **1980**, *13*, 440

17) Truhlar, D. G., Garrett, B. C., *Annu. Rev. Phys. Chem.,* **1984**, *35*, 159

18) Truhlar, D. G., Isaacson, A. D., Garrett, B. C., Generalized Transition State Theory, in *Theory of Chemical reaction Dynamic*s, ed. M. Baer, CRC Press, Boca Ranton, 1985, p. 65

19) Garrett, B. C., Truhlar, D. G., Transition State Theory, in Encyclopedia of Computational Chemistry, ed. Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., and Schafer, H.F., John Wiley & Sons, Chiichester, UK, 1998, vol. 5, p. 3094

20) Garrett, B. C., Truhlar, D. G., Variational Transition State Theory, in *Theory and Applications of Computational Chemistry: The First Forty Years,* ed. Dykstra, C. E., Frenking, G., Kim., K. S., and Scuseria, G. E., Elsevier, Amsterdam, **2005**, p.67

21) Ramos-Fernandez, A., Ellington, B. A., Garrett, B. C., and Truhlar, D. G., Hase, W.L., and Hynes, J. T., *J. Phys. Chem.*, **1983**, *87*, 2664

22) Skyner, R. E., McDonagh, J. L. Groom, C. R., van Mourik, T., and Mitchell, J. B. O., *Phys. Chem. Chem. Phys.*, **2015**, *17*, 6174

23) Miertus, S., Scrocco, E., Tomasi, J., *J. Chem. Phys.*, **1981**, *55*, 117

24) Miertus, S., Tomasi, J., *J. Chem. Phys.*, **1982**, *65*, 239

25) Marenich, A. V., Cramer, C. J., and Truhlar, D. G., *J. Phys. Chem. B*, **2009**, *113*, 6378-6396

26) Levy, R. M., Gallicchio, E., *Annu. Rev. Phys. Chem.*, **1998**, *49*, 531-567

Chapter 2: 1-Benzylspiro[piperidine-4,1'-pyrido[3,4-B]indole] 'Co-Potentiators' for Minimal Function CFTR Mutants

Portions of this work are reproduced with permission from *Eur. J. Med. Chem*., **2021**, *209*, 1112888. Results from syntheses and experimental structural characterizations were provided by Jung-Ho Sun, Jie S. Zhu, Elena Lipman, and Amy Chueng. Bioassays were performed by Puay-Wah Phuan. Computational studies, including conformational searches, RMSD calculations, and quantum NMR predictions were contributed by the author of this report.

2.1 Introduction

Cystic fibrosis transmembrane conductance regulator (CFTR) modulator therapy aiming to restore defective folding and channel gating of CFTR mutants has been a major recent focus in cystic fibrosis (CF) drug development. CFTR modulators include 'potentiators', which restore channel gating, and 'correctors', which rescue misfolded CFTR to improve cell surface presentation**.**[1] With available potentiators and correctors it is anticipated that up to 90 % of CF subjects may have an efficacious CFTR modulator therapy.[2]

However there remains an unmet need to develop therapeutics for the remaining about 10% of CF subjects.[2,3] We have described a novel approach that involves two potentiators, with the second potentiator called a 'co-potentiator', that function in synergy to increase chloride conductance for CFTR mutations found mainly in the nucleotide binding domain 2.[4-6] In our experiment, synthetic chemistry were performed to advance the spiro[piperidine-4,1-pyrido[3,4-b]indole] class of co-potentiators. Herein, structure-activity relationship studies were performed to advance the spiro[piperidine-4,1-pyrido[3,4-b]indole] class of co-potentiators.

2.2 Results

Figure 1 shows the synthesized analogs of the original hit spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **20** that were prepared in six rounds of synthetic study (synthesis rounds **r1** through **r6**). The co-potentiator activities (tested in an N1303K-CFTR expressing FRT cell model) of the synthesized analogs of **20** (10 μM/V_{MAX} 100%) are summarized at the end of the chemistry description in Table 1. The synthetic work began by preparing aryl analogs **1a-j** (synthesis round **r1**) by appropriately N^1-alkylating spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **10**. Of these first ten compounds, nine (**1b-j**) had a better EC_{50} than **20** and two (**1c** and **1j**) showed improved V_{max} (Table 1).

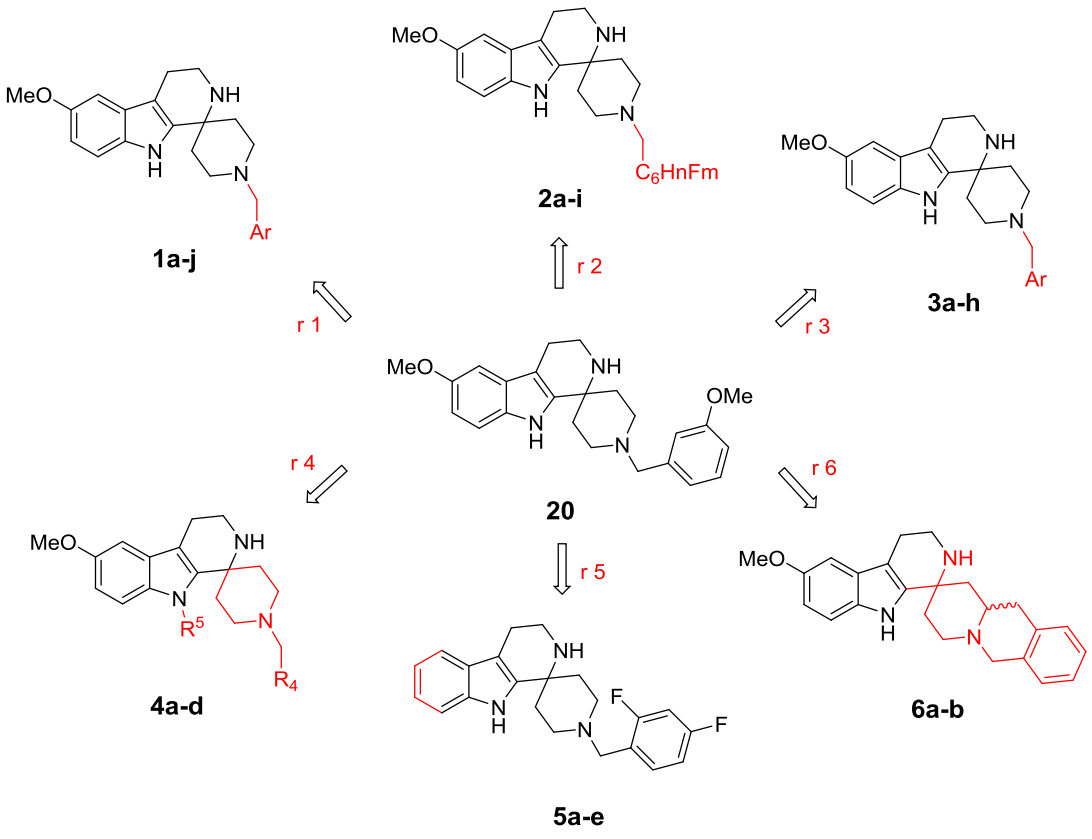
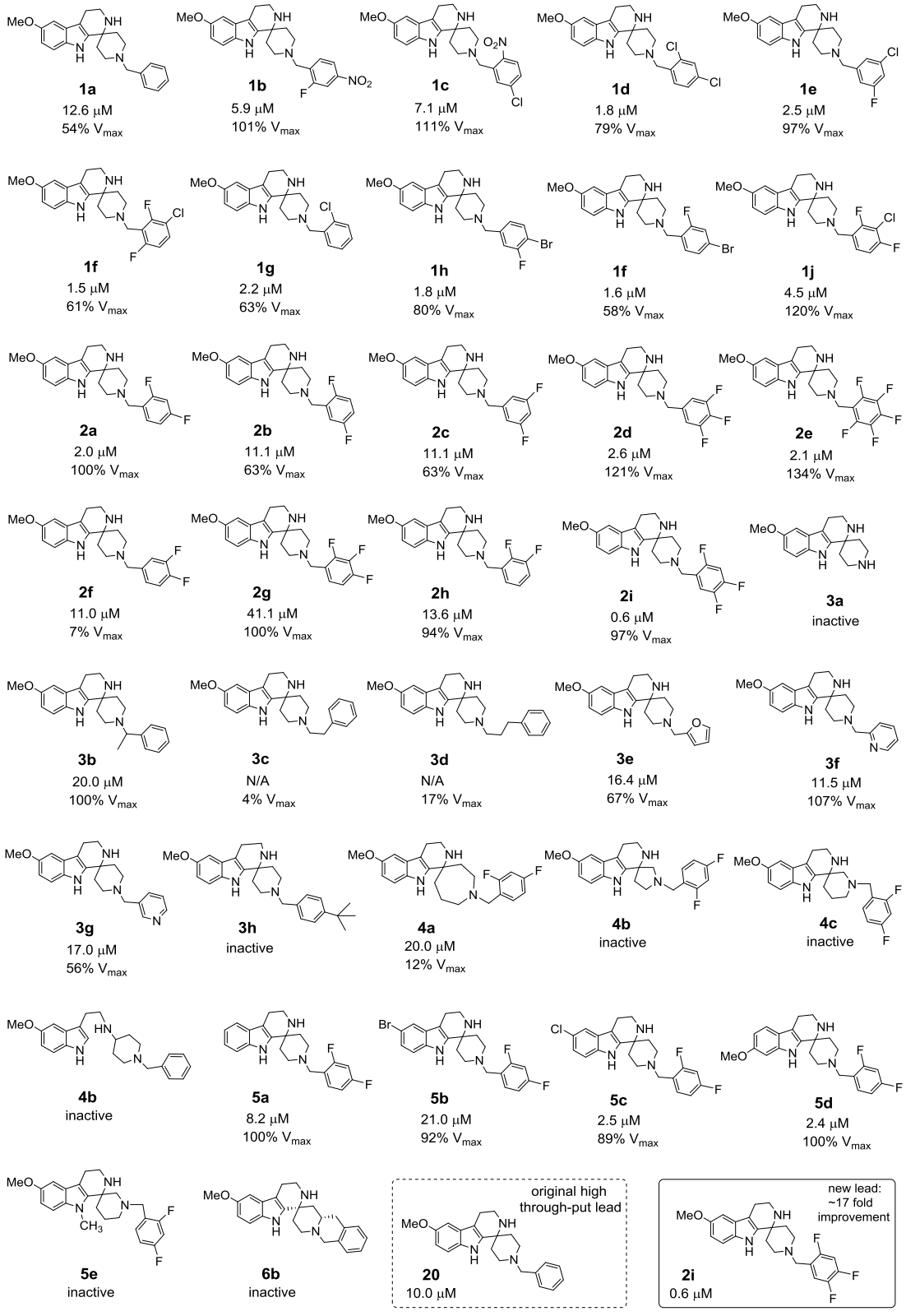


Figure 1. Structural variants 1-6 of the spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] scaffold.

Table 1. Bioassay results.



The EC_{50} and V_{MAX} of the *N*[1]-3-chloro-2,4-difluorobenzyl analog **1j** (4.5 μM/120%) motivated us to next prepare a series of fluorobenzyl-analogs (**2a-i**). These round **r2** targets (Figure 1) were again prepared by selectively *N*[1]-alkylation of spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **10**. Escalating the number of fluorine atoms on the benzyl moiety (e.g., **2a-c**/**2f**/**2h** with two fluorines versus **2d**/**2g**/**2i** with three and **2e** with five fluorines) improved the activity of these spirocycles. It was also noted that the positioning of the fluorine atoms on the benzyl moiety was important for maintaining both EC_{50} and V_{MAX} (Table 1: for example, compare **2a** versus **2b** and **2c**).

With results from synthetic rounds **r1** and **r2** in hand, the next objective (synthesis round **r3**, Figure 1) was to more broadly probe the requisite features of the *N*[1]-R^3-substituent by preparation of analogs **3a-h**. Notably, deletion of R^3 (i.e., **3a** where R^3 = H) gave an inactive compound and lengthening the *N*[1]—aryl tether [-$N(CH_2)_nR^3$ from n = 1 (**1a**) to n = 2 (**3c**) or n = 3 (**3d**)] greatly reduced activity. Similarly, adding a methyl substituent to the CH_2 tether in **1a** (12.6 μM/V_{MAX} 54%), giving **3b**, also reduced activity (**3b**: 20 μM/V_{MAX} 100%). Replacing the phenyl moiety of **1a** with a furan ring (**3e**: 16.4 μM/V_{MAX} 67%) reduced activity, while replacing the phenyl moiety of **1a** with a pyridyl ring to electronically mimic the perfluorophenyl moiety of compound **2e** was better tolerated with the nitrogen at position two (**3f**: 11.5 μM/V_{MAX} 107%), but poorly tolerated with the nitrogen at position three (**3g**: 17 μM/V_{MAX} 56%).

Having thoroughly investigated the *N1*-substituent in these spiro[piperidine-4,1'-pyrido[3,4-*b*]indoles], attention was next turned to a probe of the spiropiperidine moiety – synthesis round **r4** (Figure 1) targeting analogs **4a-d**. This work began with the preparation of spiroazepane, spiropyrrolidine, and spiropiperidine analogs where the piperidine ring of **2a** was transformed into an azepine **4a**, pyrrolidine **4b**, and *N*-offset piperidine **4c** rings. These three analogs of **2a** were prepared by reaction of 2-(5-methoxy-1*H*-indol-3-yl)ethan-1-amine (**8** where R^1 = methoxy at C5) with the appropriate ketoamine (**12**), directly giving the targeted **2a** analogs

4a-c. Interestingly, all three of these analogs were inactive. We next prepared piperidine ring-opened analogs truncated analog **4d**.

We questioned if the inherent chemical properties of the co-potentiators play a role in their activity, e.g., why **2a**, **2c** and **2e** are active but not **4a**-**4c**? We began with conformational searching using *Spartan10*.[7] For the more flexible structures **4a** and **4c**, multiple runs of conformational search were performed and the results pooled. The conformational search runs were systematic and used the Merck Molecular Force Field (MMFF). All resulting conformers were subjected to single-point energy calculations using PCM-(chloroform)-B3LYP/6-31+G(d,p).[8-11] Geometry optimizations were then performed on all conformers within 4 kcal mol^{-1} of the conformer with the lowest electronic energy, with the D3(BJ) dispersion correction[12,13] included. All quantum chemical calculations were performed using *Gaussian16*.[14] An implicit chloroform solvent was chosen for these calculations to simulate the polarity inside a general protein binding site.[15] Since **2a**, **2c**, **2e**, **4a**-**4c** contain fluorine atoms, the 6-31+G(d,p) basis set might not be sufficient for computing accurate energies; consequently, we also calculated energies for relevant conformers with PCM-(chloroform)-B3LYP-D3(BJ)//6-311+G(2d,2p). B3LYP/6-311+G(2d,2p) has been reported as a standard method for use with high electron affinity atoms, i.e., O and F. [16]

Our conformational analysis suggested a trend distinguishing between active **2a**, **2c**, and **2e,** and inactive **4a**-**4c**. Figure 2 shows the optimized geometries of the lowest energy conformers at B3LYP/6-31+G(d,p) for each structure. Most of the energetically relevant conformers of the active co-potentiators (i.e., **2a**, **2c**, and **2e**) have their benzene ring (*Ring i*) exposed for

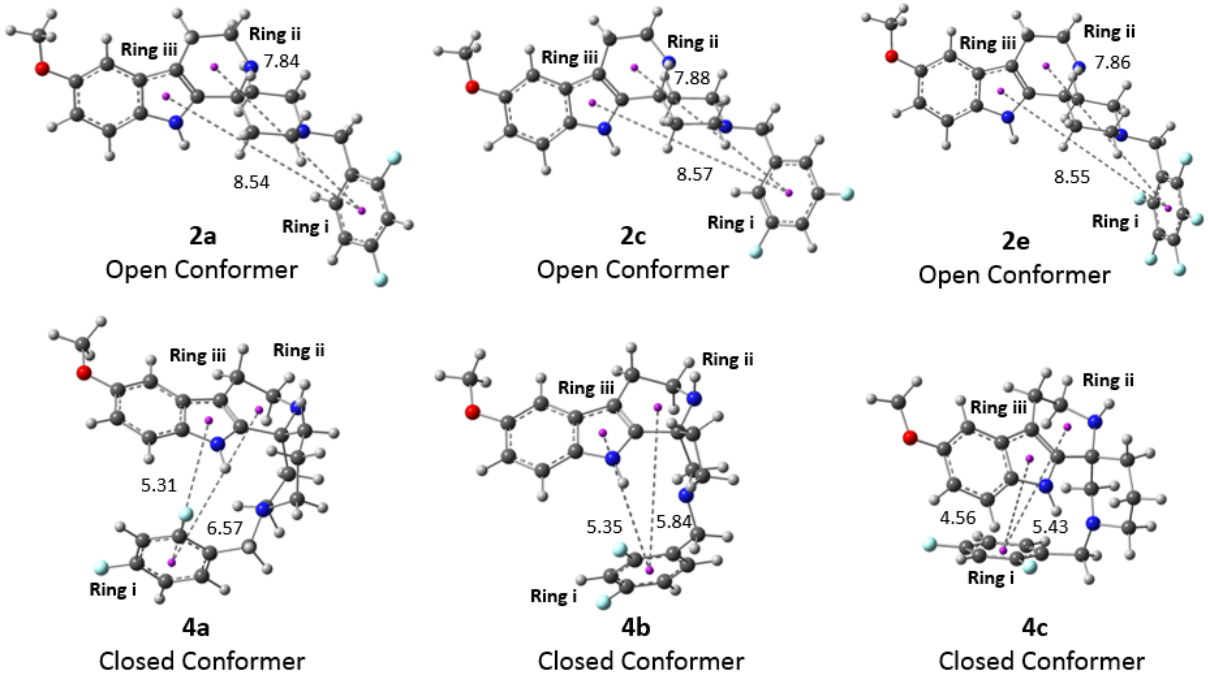


Figure 2. Optimized geometries of the lowest energy conformers of **2a**, **2c**, **2e** and **4a**–**4c**. Distances (Å) are measured from the centroid of *Ring i*, to those of *Ring ii* and *Ring iii*. The lowest energy conformers for **4a–c** preferred a close from conformation, where one face of *Ring i* is blocked from potential interactions with the protein interior.

interaction, which we define as the "Open" form (Figure 2). Assuming the protein's binding site accommodates the Open form better and has residues in the vicinity to interact with the phenyl group of the co-potentiators, the Open conformers would allow for various interactions between the π-system and the protein interior. In contrast, most of the energetically relevant conformers for **4a**-**4c** have one face of *Ring i* interacting with the rest of the molecule, precluding protein interactions with it; we refer to these conformers as "Closed" (Figure 2).

Figure 3 shows the superposition of all optimized conformers within 4 kcal mol^{-1} of the lowest energy conformer for **2a**, **2c**, **2e,** and **4a**-**4c**. We aligned the conformers and calculated their RMSD values using an RMSD calculating tool from VMD.[17] In this case, we considered

Table 2. Predicted Boltzmann distributions for Open vs. Closed forms for **2a**, **2c**, **2e,** and **4a**–**4c** at (PCM)-chloroform-B3LYP-D3(BJ)//6-31+G(d,p) (left) and (PCM)-chloroform-B3LYP-D3(BJ)//6-331+G(2d,2p) (right).

	Free Energy at PCM-(chloroform)-B3LYP-D3(BJ)//6-31+G(d,p)		Electronic Energy at PCM-(chloroform)-B3LYP-D3(BJ)//6-311+G(2d,2p)	
Structures	**Boltzmann Weight for Open form**	**Boltzmann Weight for Closed Form**	**Boltzmann Weight for Open form**	**Boltzmann Weight for Closed Form**
2a	99.5%	0.5%	89.6%	10.4%
2c	98.7%	1.3%	96.7%	3.3%
2e	98.9%	1.1%	94.5%	5.5%
4a	3.7%	96.3%	1.7%	98.3%
4b	6.1%	93.9%	7.7%	92.3%
4c	2.9%	97.1%	1.7%	98.3%

conformers to be unique if they have an RMSD value greater than 0.05 relative to other conformers. We then calculated the Boltzmann weighted averages for the Open and Closed conformers at room temperature. This allows us to determine the population of Open vs. Closed conformers for each structure, assuming all are in equilibrium.[18] Table 2 summarizes the Boltzmann weighted averages for **2a**, **2c**, **2e,** and **4a**-**4c**. The results of this analysis emphasize the difference in conformational preferences for active and inactive structures.

We also considered that these structures contain amines, which can be protonated in biological environments. Therefore, we performed conformational analysis on various protonated states for each structure as well. However, there were no consistent differences in conformational preferences for active and inactive structures. This analysis is complicated by the ability for amines to hydrogen bond with each other within each molecule, which may or may not occur in biologically relevant environs.

With exploration of the spiropiperidinepyrido aspects of the spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] scaffold in hand with synthesis rounds **r1-r4**, we next turned to synthesis round **r5** where the indole moiety of scaffold **9** was probed with analogs **5a-e**. Spiro[piperidine-

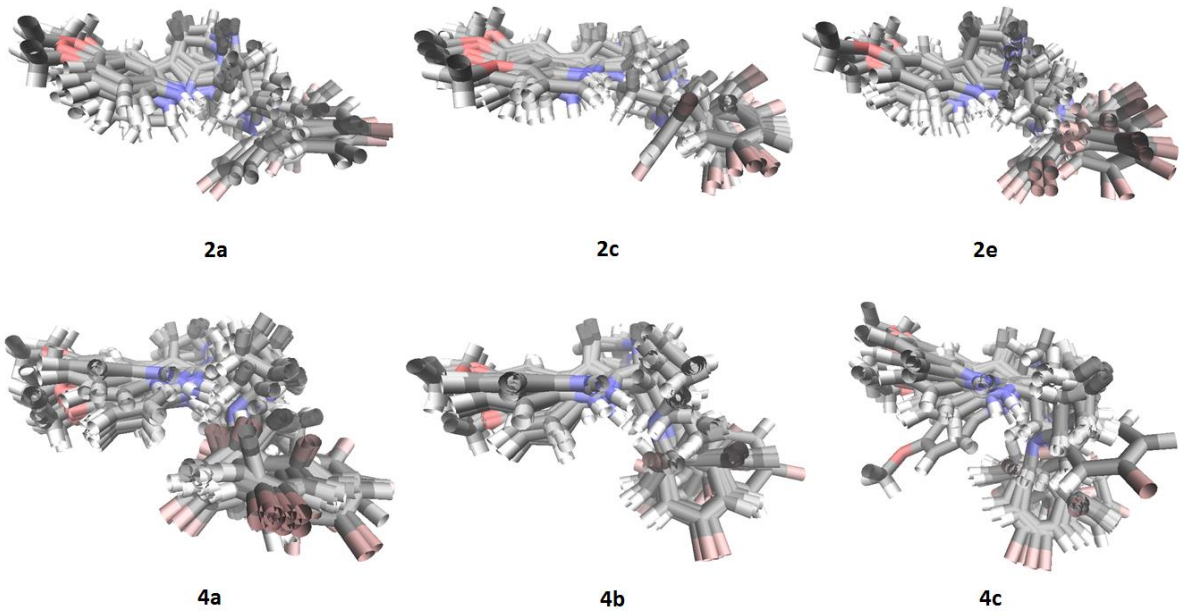


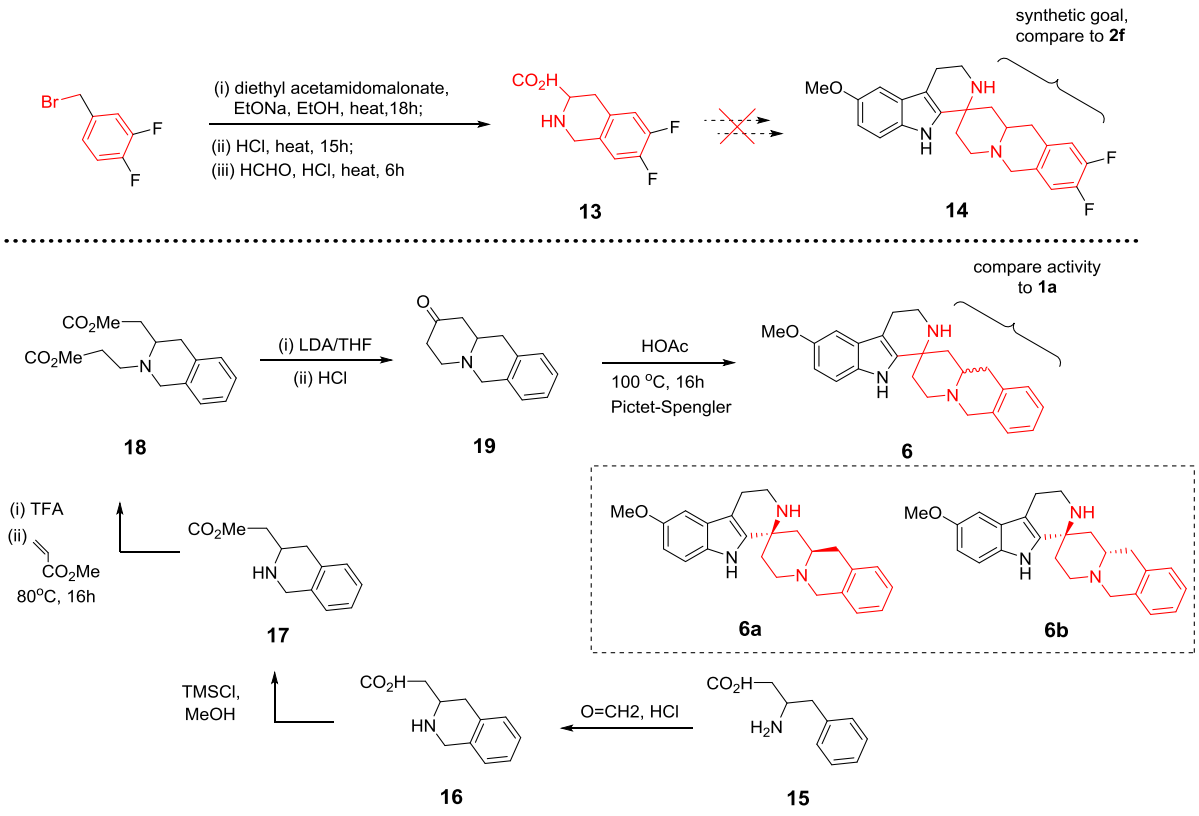
Figure 3. Superposition of optimized energetically relevant conformers for **2a**, **2c**, **2e**, **4a**-**4c**.

4,1'-pyrido[3,4-*b*]indoles] **5a-d** were prepared by reacting the appropriate tryptamine **8** with 1-(2,4-difluorobenzyl)piperidin-4-one. Whereas **5b** performed in a comparable manner to **20**, **5a** (8.2 μM/100%) showed mild improvement, and **5c** (2.5 μM/89%) and **5d** (2.4 μM/100%) were approximately 4–fold more potent. Compound **5e**, however, was inactive.

In light of the structure-activity insights gained with synthetic rounds **r1-r5**, one final synthesis round (**r6**) was undertaken where targeting constrained spiropiperidinepyrido analog **14** became a high priority in an attempt to address the question of whether there is a conformational requirement for activity vis-à-vis the N^1-(benzyl) moiety. To address this issue, we set out to prepare **14**, a constrained analog of compound **2f**. *C*-Alkylation of diethyl acetamidomalonate with 4-(bromomethyl)-1,2-difluorobenzene followed by ester hydrolysis/decarboxylation (Scheme 2) set the stage for a subsequent Pictet–Spengler reaction to delivered tetrahydroisoquinoline-3-carboxylic acid **13**.[19] Unfortunately, all attempts at the Pichet-Spengler failed – an apparent consequence of the electron deficient nature of the difluorophenyl ring. Due to this outcome, our

constrained analog study was modified to instead target constrained analog **6** (the nor-fluoro analog of **14**). Commercially available 3-amino-4-phenylbutanoic acid (**15**) successfully participated in a Pictet–Spengler reaction to give **16** and subsequent esterification delivered methyl 2-(1,2,3,4-tetrahydroisoquinolin-3-yl)acetate **17**. Michael addition to methyl acrylate delivered bis-ester intermediate **18**, which then underwent Dieckmann cyclization/decarboxylation to give ketone **19**. A second Pictet–Spengler reaction[20] of 2-(5-methoxy-1*H*-indol-3-yl)ethan-1-amine (**8** where R^1 = methoxy at C5) with **19** gave the spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **6b** as the sole product – i.e., spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **6a** was not obtained.

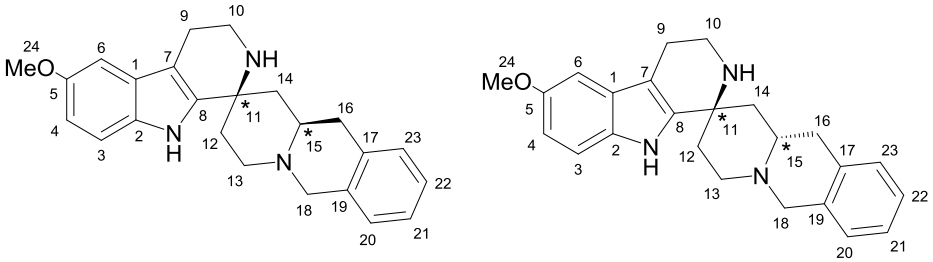
To confirm the product is indeed **6b**, we performed quantum chemical ^{1}H and ^{13}C NMR calculations on both **6a** and **6b** (Figure 4). Both **6a** and **6b** were optimized using SMD-(chloroform)-B3LYP-D3(BJ)//6-31+G(d,p).[12,13,21] Using the gauge-including atomic orbital



Scheme 1. Synthesis of constrained analog **6**.

(GIAO)[24] method, chemical shift calculations were performed with SMD-(chloroform)-mPW1PW91/6-311+G(2d,p)[23] on conformers that are within 4 kcal mol^{-1} relative to the lowest energy conformer. The chemical shifts of these energetically relevant conformers were then averaged using Boltzmann distributions. This is a common procedure for computational NMR studies.[22,24] Linear scaling (using slope= -1.0533, intercept= 186.524) and (slope= -1.0936, intercept= 31.802) (cheshirenmr.info) was applied to computed isotropic shieldings to arrive at the ^{13}C and ^{1}H shifts, respectively. Due to the similarity in the calculated ^{1}H shifts, coupling constants (Hz) were also calculated using the same methods to confirm the identity of some diagnostic ^{1}H peaks.

Table 3 shows the calculated chemical shifts for **6a** and **6b** and their deviations from the experimental shifts. Deviations within 6 ppm for ^{13}C and 0.3 ppm for ^{1}H shifts are considered acceptable.[22,24,25] Although the ^{13}C chemical shifts for **6a** and **6b** are quite similar, the computed chemical shift at position C13 of **6a** deviates greatly from the experimental value, whereas all ^{13}C shifts for **6b** are within the acceptable deviation range. Similarly, the deviations between the calculated ^{1}H shifts for **6b** and the experimental shifts are within the accepted range, but the calculated chemical shifts at positions H15, H16, and H18 for **6a** have deviations greater than 0.3 ppm from the experimental shifts. Both ^{13}C and ^{1}H chemical shifts for **6b** have lower mean



6a
[0.0]

6b
-2.33

Figure 4. Structures of **6a** and **6b**. Based on quantum calculations at SMD-(chloroform)-B3LYP-D3(BJ)//6-31+G(d,p). **6b** is 2.33 kcal mol^{-1} more thermodynamically favorable that **6a**.

absolute deviations (MAD) compared to those for **6a** as well. We also compared the free energies of the lowest energy conformers of **6a** and **6b**; that for **6a** is 2.3 kcal mol^{-1} higher than that for **6b** (Figure 4), consistent with formation of the thermodynamic product during synthesis.

The lowest energy conformer of **6b** resembles more closely the Closed conformers described above, but that of **6a** resembles the Open, i.e., where *Ring i* is available for interactions (Figure 5). Considering all conformers that are within 4 kcal mol^{-1} of the lowest energy conformers, Boltzmann distributions of open and closed forms of **6a** and **6b** were calculated (Figure 5). For **6a**, 99.3% of conformers are predicted to be Open, while only 23.2% of conformers are predicted to be Open for **6b**. Therefore, we expect **6b** to have low activity at best, if our hypothesis that conformational bias is related to activity is correct. As predicted, **6b** was inactive when tested experimentally.

Table 3. Calculated ^{13}C and ^{1}H chemical shifts for **6a** and **6b**, and their deviation from experiments.

Atom Label	Exp. ^{13}C δ (p.p.m)	Calculated 6a shift	Dev.	Calculated 6b shift	Dev.	Exp. ^{1}H δ (p.p.m)	Calculated 6a shift	Dev.	Calculated 6b shift	Dev.
1	128.10	129.08	0.98	130.04	1.94	---	---	---	---	---
2	130.74	130.14	0.60	129.69	1.05	---	---	---	---	---
3	111.62	109.53	2.09	109.80	1.82	7.06	7.19	0.14	6.79	0.26
4	100.46	103.28	2.82	102.43	1.97	6.78	6.72	0.06	6.60	0.18
5	153.98	153.72	0.26	153.66	0.32	---	---	---	---	---
6	108.47	103.35	5.12	103.45	5.02	6.94	6.92	0.02	6.92	0.02
7	111.46	109.05	2.41	105.97	5.49	---	---	---	---	---
8	140.49	142.89	2.40	143.23	2.74	---	---	---	---	---
9	23.26	24.11	0.85	24.08	0.82	2.71 2.71	2.52 2.59	0.19 0.12	2.55 2.62	0.16 0.09
10	43.76	39.56	4.21	40.94	2.82	3.08 3.08	2.94 2.98	0.14 0.10	3.07 3.21	0.01 0.13
11	52.91	53.93	1.02	53.82	0.91	---	---	---	---	---
12	26.37	34.96	1.41	37.20	0.83	2.00 1.88	2.02 1.79	0.02 0.08	2.25 1.72	0.25 0.15
13	50.40	43.27	**7.13**	48.26	2.13	2.94 2.62	3.06 2.38	0.12 0.24	2.90 2.66	0.04 0.04
14	39.28	40.98	1.70	38.35	0.93	2.25 1.88	2.29 1.53	0.04 0.34	2.04 1.88	0.21 0.15
15	56.05	52.83	3.22	56.32	0.27	2.88	3.26	**0.38**	2.76	0.12
16	36.64	29.75	5.89	35.51	0.13	2.62 2.71	2.41 4.02	0.21 **1.31**	2.38 2.46	0.24 0.25
17	133.42	135.31	1.89	136.74	3.31	---	---	---	---	---
18	57.58	56.39	1.19	56.81	0.77	3.52 3.86	3.69 4.22	0.17 **0.36**	3.50 3.86	0.00 0.02
19	133.80	135.63	1.83	136.99	3.19	---	---	---	---	---
20	126.02	126.24	0.22	126.36	0.34	7.06	7.15	0.09	7.24	0.18
21	125.81	125.98	0.17	125.82	0.01	7.11	7.10	0.01	7.14	0.03
22	126.40	126.37	0.02	126.86	0.46	7.14	7.12	0.02	7.18	0.04
23	127.67	128.94	1.27	129.60	1.93	7.06	7.16	0.10	6.99	0.06
24	51.35	53.17	1.82	53.14	1.79	3.86	3.67	0.19	3.56	0.30
MAD			2.11		1.71			0.19		0.12

*Abbreviation: MAD, mean absolute deviation; and Dev., deviation. Deviations of less than 6 ppm for ^{13}C and less than 0.3 ppm for ^{1}H shifts are considered acceptable.[22,24,25] Chemical shifts exceeding the accepted deviations are bolded above. Coupling constants (J) were also calculated to confirm the identity of complex proton peaks.

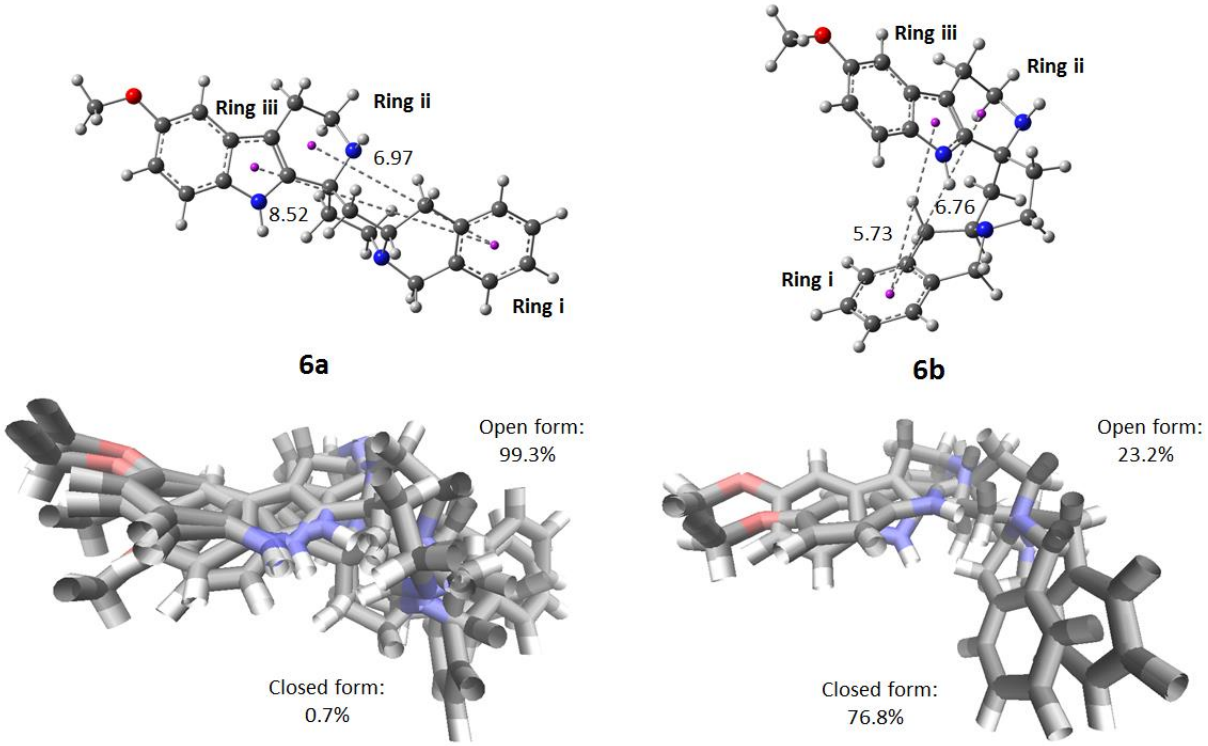


Figure 5. Lowest energy conformers optimized at SMD-(chloroform)-B3LYP//6-31+G(d,p) with D3(BJ) for **6a** and **6b** respectively (top); the distance between centroids of *Rings i* and *ii* and *Rings i* and *iii* were measured. Superimposed images of energetically relevant conformers for **6a** and **6b** (bottom); **6a** has more open form conformers in equilibrium.

2.3 Conclusion

This study extends our prior work that identified a novel class of CFTR modulators, termed co-potentiators, that act in synergy with existing potentiators such as VX-770 and GLPG1837 to increase chloride channel function of CFTR mutants.[4-6] The objective of this study was to optimize of spiro[piperidine-4,1-pyrido[3,4-b]indoles] co-potentiators previously identified in a high-throughput screen.[6] A straightforward, high yielding synthetic strategy was developed to investigate structure-activity relationships. In total, 37 spiro[piperidine-4,1-pyrido[3,4-b]indoles] were synthesized to investigate the consequences of altering the indole and benzyl moieties. Analogs of the original spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] compound **20** were prepared in

six rounds of synthetic studies (synthesis rounds **r1** through **r6**) starting from commercial tryptamines **8**. These were condensed with *N*-alkylated piperidin-4-ones **12** to directly deliver the targeted spiro[piperidine-4,1'-pyrido[3,4-*b*]indole] **9** by a classical Pictet–Spengler reaction. When preparing R^2 analogs of **9**, it proved more efficient to prepare unalkylated analog **10** [acetic acid-mediated condensation of tryptamine **8** with piperidin-4-one **11**] and then subsequently perform selective N^1-alkylation of **10** (R^2-X in CH_2Cl_2 + K_2CO_3) to give **9**.

In summary, spiro[piperidine-4,1-pyrido[3,4-b]indoles] represent an evolvable CFTR co-potentiator scaffold with nanomolar potency that in synergy with existing potentiators activates certain CF-causing CFTR mutations. The expansion of spiro[piperidine-4,1-pyrido[3,4-b]indole] structure activity relationships may yield insight into the binding site of co-potentiators through computational or structural approaches, as well as subsequent rational compound optimization.

Reference

1) Elborn, J.S., *Lancet*, **2016**, *388*, 2519-2531

2) Burgener, E.B. and Moss, R. B,.*Curr. Opin. Pediatr.*, **2018**, *30*, 372-377

3) Kym, P.R., Wang X., Pizzonero M. and Van der Plas, S. E., *Prog. Med. Chem.* **2018**, *57*, 235-276

4) Haggie, P.M., Phuan, P. -W., Tan, J. A., Xu, H., Avramescu, R. G., Perdomo, D., Zlock, L., Nielson, D. E., Finkbeiner, W. E., Lukacs, G. L. and Verkman, A. S., *J. Biol. Chem.,***2017***, 292*, 771-785

5) Phuan, P.-W., Son J. -H., Tan, J. A., Li, C., Musante, I., Zlock, L., Nielson, D. W., Finkbeiner, W. E., Kurth, M. J., Galietta, L. J., Haggie, P. M., and Verkman, A. S., *J. Cyst. Fibros.,* **2018**, *17*, 595-606

6) Phuan, P.-W., Tan, J. A., Rivera, A. A., Zlock, L., Nielson, D.W., Finkbeiner, W. E., Haggie, P. M, and Verkman, A. S., *Sci. Rep.* **2019**, *9*, 17640

7) Spartan, Version10 (Wavefunction, Ind.: Irvine, CA, USA **2010**)

8) Becke, A.D., *J. Phys. Chem.,* **1993**, *98*, 5648-5652

9) Lee, C.J., Wang, W., and Parr, R. G., *Phys. Rev. B: Condens. Matter. Phys.,* **1988**, *37*, 785-789

10) Miehlich, B., Savin, A., Stoll, H., and Preuss, H., *Chem. Phys. Lett.,***1989**,*157*,200-206

11) Cammi, R., Mennucci, B., and Tomasi, J. *In computational chemistry;* J. Leszczynski, Editor. **2003**, World Scientific Publishing Co. Pte. Ltd.: Singapore.

12) Grimme, S., Antony, J., Ehrlich, S., and Krieg, H., *J. Chem. Phys.,* **2010**, *132*, 154104

13) Grimme, S., Ehrlich, S., and Goerigk, L. *J. Comp. Chem.,* **2011***, 132*,1456

14) Gaussian 16, Revision C.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R.; Hasegawa, J.; Ishida, M., Nakajima, T., Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N., Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P., Burant, J. C.; Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M.; Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J., Gaussian, Inc., Wallingford CT, **2016**

15) Lee, J.K. and Tantillo, D. J., *Adv. Phys. Org. Chem.* **2003**, *38*,183-218

16) Jursic, B.S., *J. Chem. Phys.,* **1996**, *104*, 4151

17) Humphrey, W., Dalke, A., and Schulten, K., *J. Mol. Graphics*, **1996**, *14*, 33-38

18) Sharp, K. and Matschinsky, F.,Translation of Ludwig Boltzmann's paper "on the relationship between second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissen Classe. Abt II, LXXVI 1877, pp 373-435*. Entropy,* **2015**, *17*,1971-2009

19) Xu, X., Ge, R., Li, L., Wang, J., Lu, X., Xue, S., Chen, X., Li, Z., and Bian, L., *Eur. J. Med. Chem.,* **2018***, 143*,1325-1344

20) Harrison, J.R., O'Brien, P., Porter, D. W., and Smith, N. M., *J. Chem. Soc.,* **1999***, 24*, 3623-3631

21) Marenich, A.V. and Truhlar, D. G., *J. Phys. Chem. B,* **2009***, 113*, 6378-6396

22) Grimblat, N. and Sarotti, A. M., *Chem. Eur. J.*, **2016**, *22*,12246-12261

23) Adamo, C. and Barone, V., *J. Chem, Phys.,* **1998***, 108*, 664-675

24) Lodewyk, M.W., Siebert, M. R., and Tantillo, D. J., *Chem. Rev.*, **2012**, *112*, 1839-1862

25) Willoughby, P.H., Jansma, M. J. and Hoye, T. R. *Nat. Protoc.,* **2014***, 9*, 643-660

Chapter 3: Reconsidering the Structure of Serlyticin-A

Portions of this work are reproduced with permission from *J. Nat. Prod.*, **2019***, 82* (12)*,* 3464–3468. Syntheses and experimental structural characterizations were performed by Robert J. Tombari. Computational studies, including conformational searches and quantum NMR predictions were contributed by the author of this report.

3. 1 Introduction

Serlyticin-A is a secondary metabolite first isolated from the fermentation broth *Serratia ureilytica* using squid pen as the sole carbon source. A previous study by Kuo *et al.* demonstrated that it has antioxidative and antiproliferative properties. However, the proposed chemical structure of serlyticin-A is likely incorrect based on the thermodynamic instability of its three contiguous heteroatom-heteroatom bonds. Here, we use quantum chemical calculations to predict ^{1}H and ^{13}C chemical shifts for serlyticin-A, and demonstrate a discrepancy between the calculated and experimental chemical shifts. Based on our calculations, we propose several reasonable alternative structures for serlyticin-A. Considering the known antioxidant and antiproliferative activity of hydroxamic acids as well as their stability and prevalence in natural products of bacterial origin, we believe that serlyticin-A is most likely 3-indolylacetohydroxamic acid (**4**). We provide our rationale for this assignment as well as experimental data for pure 3-indolylacetohydroxamic acid obtained via *de novo* synthesis. This study highlights the power of computational NMR shift prediction to revise chemical structures for natural products like serlyticin-A.

The identification of a novel natural product with antioxidant properties is highly beneficial to both the food and pharmaceutical industries. Previous efforts in this area have utilized a variety of organisms with plants being a major source of antioxidant compounds.[1-4] Despite the fact that shellfish waste is a rich source of phenolic antioxidants, there have been relatively few attempts to isolate novel antioxidant natural products from this source.[5,6] In 2012, serlyticin-A was identified

by Kuo and co-workers as one of the compounds in the fermentation broth of squid pen waste, and they demonstrated that it has antioxidative and antiproliferative properties[5]. While serlyticin-A certainly has the potential for medicinal and food science applications, we suspect that the reported structure is incorrect for several reasons. First, the structure contains an unstable O–O bond, which based on our calculations, would dissociate in solution. Second, the characterization data reported for serlyticin-A, including 1D and 2D NMR, mass spectrometry, UV-vis, and IR, is consistent with multiple potential structures. For example, 1D and 2D NMR data confirm the presence of an indole moiety but does not necessarily support the assignment of the proposed N–O–O–N substructure. Moreover, the reported mass spectrometry data does not match the exact mass of **1**. While select tabulated numerical data was provided in the original report, the actual spectra were not.[7] These factors coupled with the lack of X-ray crystallography data contribute to the ambiguity in the structure of serlyticin-A. While it is an accepted practice to elucidate the chemical structures relying on information obtained from NMR, IR, and UV-Vis, it is not uncommon for these structures to be misassigned, leading to the expenditure of additional resources in chemistry.[8-10] Here, we report the use of quantum chemical calculations of NMR chemical shifts[18-20] to determine that the original structure of serlyticin-A may have been misassigned, and we provided several plausible alternative structures.

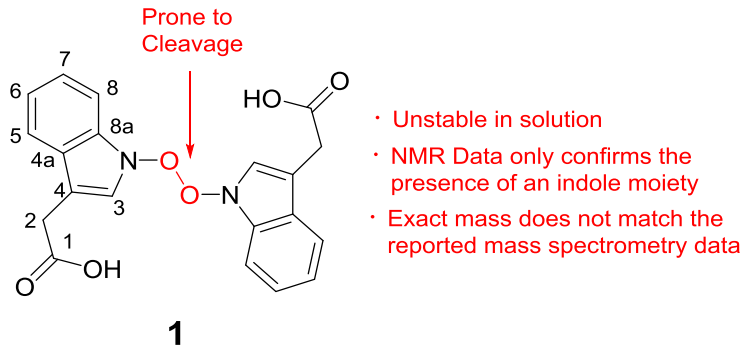


Figure 1. The previously proposed structure of serlyticin-A (left).

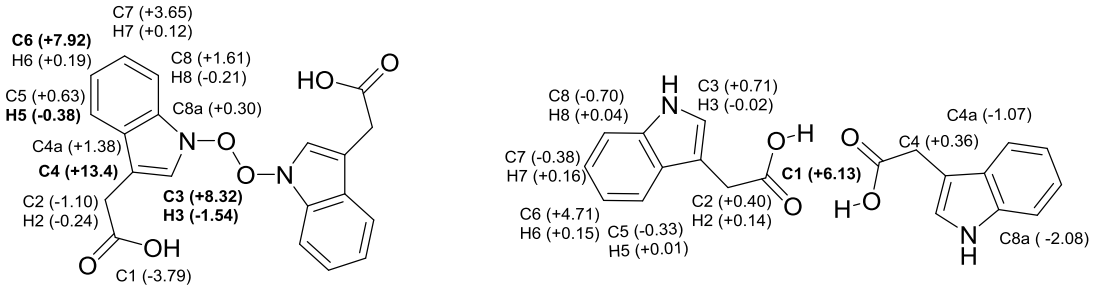
3.2 Methods

Quantum chemical calculations were performed using *Gaussian09*.[11] Structural optimizations and frequency calculations were performed with both B3LYP/6-31+G(d,p)[12] in the gas phase and PCM-(MeOH).[13,14] Both sets of optimized structures were subjected to NMR calculations using the gauge-including atomic orbital (GIAO) method.[19] Results from both approaches are consistent with each other. Shown in the text are the results from NMR calculations on optimizations performed with PCM-(MeOH). Since the NMR experiments reported by Kuo and co-workers were performed on a sample dissolved in deuterated methanol, our ^{13}C and ^{1}H shifts were calculated with PCM-(MeOH)-mPW1PW91//6-311+G(2d,p).[15] Chemical shifts were linearly scaled using scaling factors obtained from cheshirenmr.info (slope= −1.0754 for ^{1}H and −1.0399 for ^{13}C; intercept= 31.8463 for ^{1}H and 186.5993 for ^{13}C).[16,18]

In order to sufficiently sample conformational space, multiple systematic conformational search runs were performed on each compound using *Spartan10*.[17] NMR calculations were only performed on the conformers within 3 kcal mol^{-1} of the lowest energy conformer. The geometries of all conformers for each structure were confirmed to be minima (no imaginary frequencies). The chemical shifts of these contributing conformers were averaged using Boltzmann-weighting. This procedure is well precedented for computational NMR stuies[18-20]. In addition, an optimization constraining the fragmentable O–O bond of the originally proposed structure for serlyticin-A, **1** (Figure 1) was performed to assess its thermostability and its tendency to fragment in solution. Natural Bond Orbital (NBO)[21] analysis calculations were performed on all contributing conformers of **1** to obtain their Wiberg bond orders.[22]

3.3 Results

Our investigations began with the structural optimization of **1**. We observed a long O–O bond (2.10 Å) in all conformers of **1** optimized in the gas phase and solution. Computed Wiberg bond orders for the O–O bond were less than 0.32 for all conformers within 3 kcal mol^{-1} of the lowest conformer, suggesting that **1** is prone to fragment in both gas phase and solution. Optimization of these conformers while constraining the O–O bond to 1.48 Å (a reasonable distance for a covalent O–O bond) led to structures that are at least 15 kcal mol^{-1} higher in free energy than their unconstrained versions. Additionally, the calculated ^{13}C and ^{1}H NMR shifts for **1** deviate greatly from the reported NMR shifts (Figure 2). For the constrained structure of **1**, although the ^{13}C shifts are close to the experimental shifts, several ^{1}H shifts deviate greatly from the experimental shifts (Figure 2). Given the accuracy of such calculations[18-20], serlyticin-A is likely



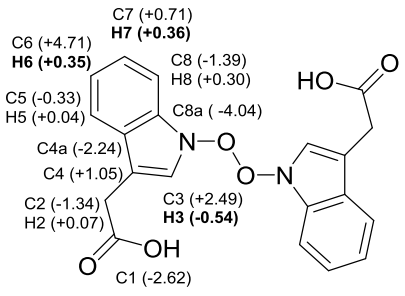


Figure 2. Deviations on calculated NMR shifts for **1**, **1** with constrained O-O bond, and dimer of indole-3-Acetic Acid. Each structure was subjected to computational NMR calculations. Deviations of less than 5–6 ppm for 13C and less than 0.3 ppm for 1H shifts are considered acceptable.[18-20] Deviations exceeding these limits are bolded.

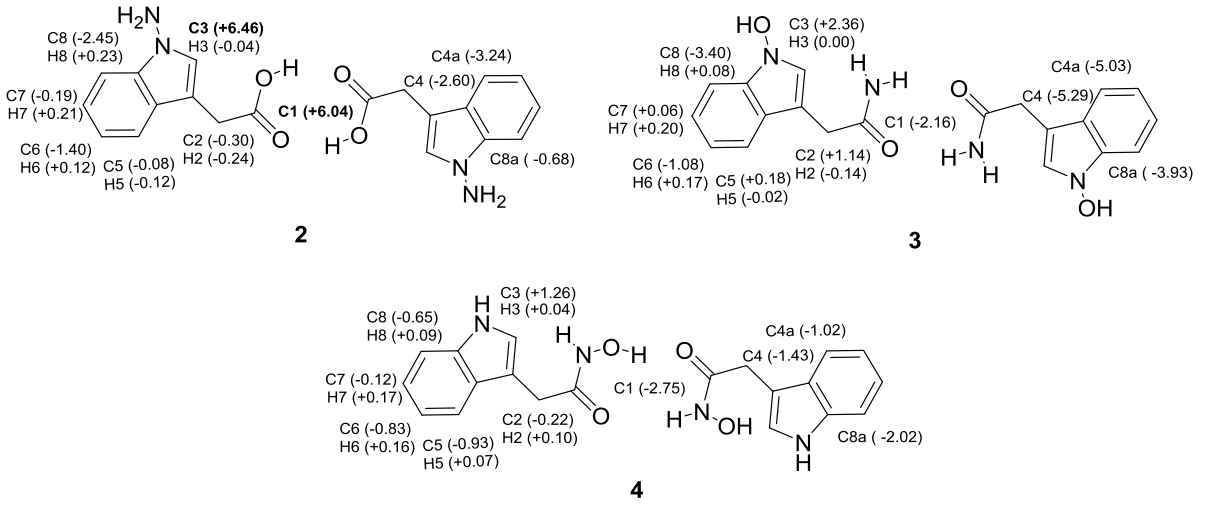


Figure 3. Structures examined for serlyticin-A structural correction. Each structure was subjected to computational NMR calculations. Deviations of less than 5–6 ppm for 13C and less than 0.3 ppm for 1H shifts are considered acceptable.[18-20] Deviations exceeding these limits are bolded.

misassigned. Kuo *et al.* reported ^{1}H-^{1}H COSY (correlation spectroscopy) and HMQC (heteronuclear multiple bond coherence) correlations, that are consistent with serlyticin-A containing an indole-3-acetic acid moiety. This led us to consider structural alternatives possessing this group. Our predicted ^{13}C chemical shifts for the indole-3-acetic acid dimer (Figure 2) matched the experimental shifts very well with only one calculated ^{13}C chemical shift (C1) falling outside the acceptable range. Therefore, derivatives of this structure were examined further. Such structures included hydrazine, **2**, hydroxylamine **3**, and hydroxamic acid **4** (Figure 3). All of these structures have computed chemical shifts that are consistent with the reported shifts for serlyticin-A, though **4** is the closest match (Figure 4). Additional structures considered but were not good matches. Serlyticin-A has demonstrated antioxidant properties in a DPPH radical scavenging assay.[5] Compounds **2–4** all have the potential to serve as effective radical scavengers due to the presence of weak N–O or N–N bonds. However, we deemed **4** to be the most likely structure of serlyticin-A for several reasons. First, to the best of our knowledge, N-amino indole natural products have never been reported. Second, while N-hydroxy indole natural products are

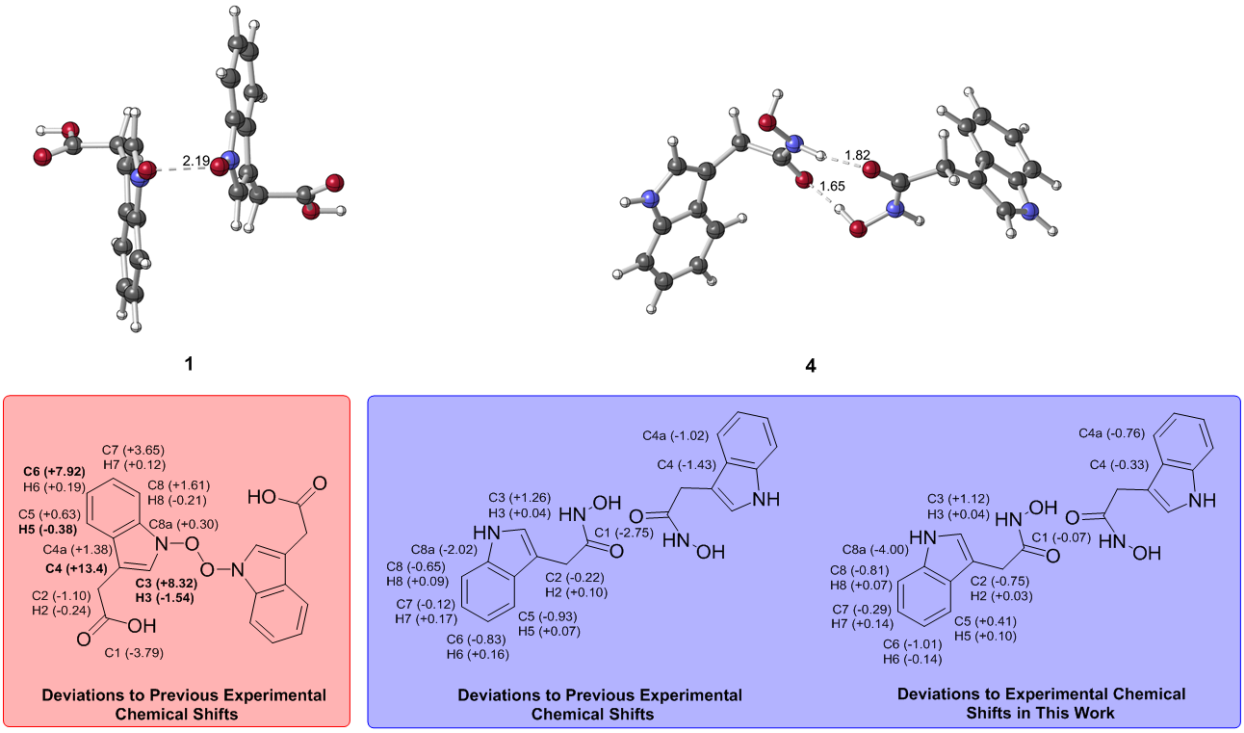


Figure 4. Deviations on calculated shifts of **1** and its reported experimental shifts (left). Deviations on calculated shifts of **4** and the previously reported experimental shifts and experimental shifts in this work (middle and right, respectively). Deviations of less than 5–6 ppm for ^{13}C and less than 0.3 ppm for ^{1}H shifts are considered acceptable.[18-20] Deviations exceeding these limits are bolded.

known, these are exceedingly rare and often difficult to isolate.[23] Examples of notable N-hydroxyindole-containing natural products include stephacidin B, versicoamide F, nocathiacin-I, thiazomycin, coproverdine, notoamide G, and N-hydroxy-β-carbolines.[24-32] The overwhelming majority of N-hydroxyindole-containing natural products possess substitution at the 2-position of the indole, likely a consequence of the tendency of indoles to be oxidized at C2. Finally, **4** contains a hydroxamic acid group—a structural motif that is found in numerous natural products of bacterial origin.[33-36] Hydroxamic acids are excellent siderophores and bacteria often produce them to sequester iron from their environment. In fact, **4** itself has been shown to coordinate metals.[37] Furthermore, many hydroxamic acids possess anti-proliferative properties, like that observed for serlyticin-A, due to their ability to potently inhibit histone deacetylases (HDACs).[38,39] Next, we synthesized **4** from the methyl ester of indole acetic acid by converting the ester into the

hydroxamic acid.[40] We have characterized the product and the data agree with previous studies.[41,42] While Kuo *et al.* described serlyticin-A as a yellow powder with IR absorption bands of 3366 cm^{-1} and 1713 cm^{-1} corresponding to a hydroxyl group and carbonyl moiety, respectively, we found compound **4** to be a tan powder with IR absorption bands at 3422, 3218, 1636, and 740 cm^{-1}. Unfortunately, given the lack of raw characterization data in the original report[5], we cannot be sure if these differences are meaningful. Kuo and co-workers reported UV absorption bands (solvent not indicated) for serlyticin-A at 244, 261, and 299 nm, implying that the molecule could contain an indole moiety. We analyzed **4** by UV-Vis absorption in MeOH and found absorption bands at 218, 230, and 280 nm, a reasonable match to the reported values assuming different solvents were used. Structure **1**, with the molecular formula $C_{20}H_{16}N_2O_{6,}$ has a calculated exact mass of 380.1008. The reported ESI-MS $[m + H]^+$ *m/z* was 381.2629, a difference of > 0.15 from the calculated m/z of **1** $[m + H]^+$. It is possible that the mass reported by Kuo and co-workers reflected a noncovalent dimer of the natural product. This type of dimerization is commonly detected in ESI-MS,[43,44] and would be anticipated for a hydroxamic acid such as **4**. The dimer of **4** has a calculated *m/z* $[2m + H]^+$ of 381.1563, which is closer to the reported mass for serlyticin-A than the calculated *m/z* for **1**. The ^1H and ^{13}C NMR chemical shifts for synthesized **4** agree with the calculated shifts (Figure 4). Overall, the chemical and biological data reported for serlyticin-A[5] are consistent with a compound such as **4**, but comparison with an authentic sample will be necessary to confirm this hypothesis. Regardless, a variety of factors make it highly unlikely that **1** is the correct structure of serlyticin-A.

3.4 Conclusions

In the field of organic chemistry, structural misassignments are problematic.[8] Here, we have demonstrated that the originally proposed structure for serlyticin-A is likely incorrect, a

conclusion based on 1) computations that point to the likelihood that the central O–O bond in the proposed structure would readily break, 2) deviations between the reported and calculated chemical shifts for **1**, and 3) the unprecedented nature of such an N-oxidized indole natural product lacking substitution at C2. Furthermore, the high prevalence and known antioxidant/antiproliferative properties of naturally occurring hydroxamic acids is consistent with our proposal that serlyticin-A is likely 3-indolylacetohydroxamic acid (**4**) or a closely related compound.

Reference

1) (a) Inatani, R., Nakatani, N., Fuwa, H., and Seto, H., *Agric. Biol. Chem.,* **1982**, *46*, 1661-1666 (b) Chang, S. S., Ostric-Matijasevic, B., Hsieh, O. L., and Huang, C. L., *J. Food Sci.*, **1977**, 42, 1102-1106 (c) Cuvelier, M. E., Berset, C., and Richard, H., *J. Agric. Food Chem,* **1994**, *42*, 665-669

2) Farombi, E. O., Britton, G., and Emerole, G. O., *Food Res. Int.*, **2000,** *33*, 493−499

3) Canadanovic-Brunet, J. M., Djilas, S. M., Cetkovic, G. S., Tumbas, V. T., Mandic, A. I., and Canadanovic, V. M., *Int. J. Food Sci.Technol*., **2006**, *41*, 667−673

4) Diouf, P. N., Stevanovic, T., and Cloutier, A., *Food Chem.*, **2009**, *113*, 897−902.

5) Kuo, Y. H., Hsu, H. C., Chen, Y. C., Liang, T. W., and Wang, S. L., *J. Agric. Food Chem.*, dx.doi.org/10.1021/jf302481n

6) Seymour, T. A., Li, S. J., and Morrissey, M. T., *J. Agric. Food Chem.*, **1996**, *44*, 682-685

7) Attempts to contact the authors of ref. 1 requesting an authentic sample and/or original spectra went unanswered.

8) Nicolaou, K. C., and Snyder, S. A., *Angew. Chem. Int. Ed.*, **2005**, *44* (7), 1012-1044

9) Elyashberg, M., Blinov, K., Molodtsov, S., Smurnyy, Y., Williams, A. J., and Churanova, T., *J. Cheminformatics.,* **2009**, 1:3

10) Elyashberg, M., Williams, A. J., and Blinov, K., *Nat. Prod. Rep.*, **2010**, *27*, 1296–1328

11) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., et al. Gaussian 09, Revision D. 01; Gaussian, Inc. Wallingford, CT, USA, **2010**

12) (a) Becke, A. D., *J. Chem. Phys.*, **1993**, *98,* 5648−5652 (b) Lee, C., Yang, W., and Parr, R. G., *Phys. Rev. B: Condens. Matter. Mater. Phys.*, **1988**, *37*, 785−789 (c) Miehlich, B., Savin, A., Stoll, H., and Preuss, H., *Chem. Phys. Lett.*, **1989**, *157*, 200−206

13) Cammi, R., Mennucci, B., Tomasi, J., Leszczynski, J., Ed.; *World Scientific Publishing Co. Pte. Ltd.: Singapore*, **2003**; Vol. 8

14) Cheeseman, J. R., Trucks, G. W., Keith, T. A., and Frisch, M. J., *J. Chem. Phys.*, **1996**, *104*, 5497−5509

15) Adamo, C., and Barone, V., *J. Chem. Phys.*, **1998**, *108*, 664−675

16) Scaling factors were obtained from chesirenmr.info.

17) Spartan, Version10 (Wavefunction, Ind.: Irvine, CA, USA **2010**)

18) Lodewyk, M. W., Siebert, M. R., and Tantillo, D. J., *Chem. Rev.*, **2012**, *112*, 1839–1862

19) Grimblat, N., and Sarotti, A. M., *Chem. Eur. J.*, **2016**, *22*, 12246–12261

20) Willoughby, P.H., Jansma, M. J., and Hoye, T. R., *Nat. Protoc.,* **2014**, *9*, 643–660.

21) NBO 7.0. E. D. Glendening, J, K., Badenhoop, A. E., Reed, J. E., Carpenter, J. A., Bohmann, C. M., Morales, P., Karafiloglou, C., Landis, R., and Weinhold, F., Theoretical Chemistry Institute, University of Wisconsin, Madison (**2018**)

22) (a) Weinhold, F., *J. Comp. Chem.*, **2012**, *33*, 2363-2379 (b) Reed, A. E., Weinstock, R. B., and Weinhold, F., *J. Chem. Phys.*, **1985**. *83*, 735-746 (c) Glendening, E. D., and Weinhold, F., *J. Comp. Chem.*, **1998**, *19*, 610-627

23) Somei, M., *Adv. Heterocycl. Chem.,* **2002**, *82*, 101-155

24) Qian-Cutrone, J., Huang, S., Shu, Y. Z., Vyas, D., Fairchild, C., Menendez, A., Krampitz, K., Dalterio, R., Klohr, S. E., and Gao, Q., *J. Am. Chem. Soc.*, **2002**, *124*(49),14556-14557

25) Liu, L., Wang, L., Bao, L., Ren, J., Bahadur Basnet, B., Liu, R., He, L., Han, J., Yin, W. B., Liu, H., and Versicoamides F-H., *Org Lett.*, **2017**, *19*(4), 942-945

26) Singh, S. B., Herath, K., Yu, N. X., Walker, A. A., and Connors, N., *Tetrahedron. Let.*, **2008**, *49*, 6265–6268

27) Zhang, C., Herath, K., Jayasuriya, H., Ondeyka, J. G., Zink, D. L., Occi, J., Birdsall, G., Venugopal, J., Ushio, M., Burgess, B., Masurekar, P., Barrett, J. F., and Singh, S. B., *J. Nat. Prod.* **2009**, *72*(5), 841-7

28) Urban, S., Blunt, J. W., and Munro, M. H., *J. Nat. Prod.*, **2002**, *65*(9), 1371-1373

29) Tsukamoto, S., Kato, H., Samizo, M., Nojiri, Y., Onuki, H., Hirota, H., and Ohta, T., *J. Nat. Prod.,* **2008**, *71*(12), 2064-2067

30) Susanna, T. S., Chan, A., Pearce, N., Page, M. J., Kaiser, M., and Copp, B. R., *J. Nat. Prod.*, **2011**, *74*, 1972–1979

31) Santos A. K., Machado L. L., Bizerra A. M., Monte F. J., Santiago G. M., Braz-Filho, R., and Lemos, T. L., *Nat. Prod. Commun.*, **2012**, *7*(6):729-730

32) Costa, E. V., Pinheiro, M. L., Xavier, C. M., Silva, J. R., Amaral, A. C., Souza, A. D., Barison, A., Campos, F. R., Ferreira, A. G., Machado, G. M., and Leon, L. L., *J. Nat. Prod.*, **2006**, *69*(2):292-4

33) Neilands, J. B., *Science.*,**1967**, *156*(3781), 1443-1447

34) Challis, G. L., *Microbiology*., **2008**, *154*, 1555–1569

35) Miller, M. J., *Chem. Rev.*, **1989**, *89*, 1563-1579

36) Tsuji, N., Kobayashi, M., Nagashima, K., Wakisaka, Y., and Koizumi, K., *J. Antibiot.*, **1974**, *49*, 14691

37) Arrabal, M. J., González, P. V., Gámez, C. C., Misiego, A. S., and de la Peña, A. M., *Analyst*, **1994**, *119*, 1537-1540

38) Marks, P. A., Richon, V. M., and Rifkind, R. A., *J. Natl. Cancer. Inst.*, **2000**, *92* (15)

39) Marks, P. A., *Expert Opin. Investig. Drugs.*, **2010**, *19*(9), 1049-1066

40) Holson, E., and Olson, D., U.S. Patent 9745613B2, **2017**

41) Cohen, W., and Erlanger, B. F., *J. Am. Chem. Soc.*,**1960**, *82*(15), 3928-3934

42) Boularot, A., Giglione, C., Petit, S., Duroc, Y., Alvesde Sousa, R., Laure, V., Cresteil. T., Dardel. F., Artaud., I., and Meinnel., T., *J. Med. Chem.,* **2007**, *50*, 10-20

43) Pan, H. A., *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 3555–3560

44) Ding, J., and Anderegg, R. J., *J. Am. Soc. Mass Spectrom.*, **1995**, *6*, 159

Chapter 4: Structure and Computational Basis for Backbone Rearrangement in Marine Oxasqualenoids

Portions of this work are reproduced with permission from *J. Org. Chem.*, **2021***, 86*(3) 2437–2446. Extraction, structural characterizations, and computational NMR predictions were contributed by Francisco Cen-Pacheco and Adrián J. Santiago-Benítez. Computational mechanistic studies were performed by the author of this report.

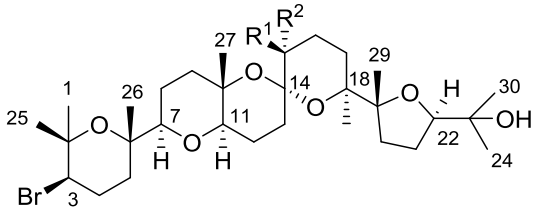
4.1 Introduction

Red algae of the genus *Laurencia* are one of the most prolific sources of secondary metabolites known so far.[1,2] Indeed, more than a thousand different compounds have been isolated from this genus. In addition, a vast array of notable bioactivities has been reported for these molecules.[3] Triterpene polyethers, also known as oxasqualenoids, represent one of the most interesting groups of metabolites isolated from *Laurencia*. Their structural diversity and complexity derive from different cyclization reactions are resulted from a chemical reaction cascade. Oxasqualenoids likely arise from a common precursor, (10R,11R) squalene epoxide, isolated from *L. okamurai*.[4] The later would be further oxidized into (6S,7S,10R,11R,14R,15R,18S,19S) squalene tetraepoxide as a common intermediate. Next, different squalene cyclizations have been proposed to proceed by a sophisticated multistep carbocation cascade, generating a large number of different compounds. However, a significant number of intricate structures are difficult to explain by this classic sequential cascade, therefore generating a great challenge to establish their biosynthetic pathway.[5-8]

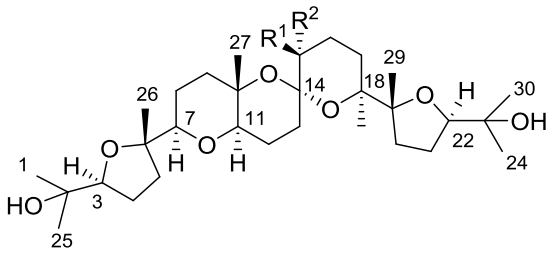
4.2 Results and Discussion

Herein, we describe the isolation and structure elucidation of six unreported oxasqualenoids, named laurokanols A-E and yucatecone, from *L. viridis* (Figure 1). Their discovery is significative from a biosynthetic point of view. Thus, the novel tricyclic core of

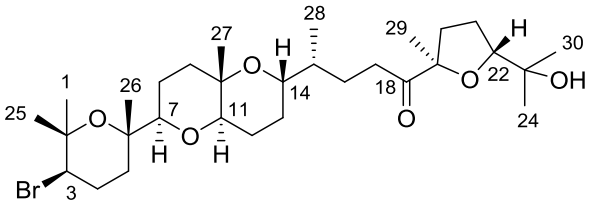
laurokanols can be considered a key biogenetic intermediate to explain the intriguing configuration observed in yucatecone. Density-functional theory (DFT) calculations were used to understand their biosynthetic pathway. Laurokanol A (**1**) showed a pseudomolecular ion at *m/z* 609.2751/611.2755 $[M+Na]^+$ (calcd 609.2731/611.2744) in its HRESIMS spectrum, which correspond to a molecular formula of $C_{30}H_{51}O_6Br$. Analysis of COSY and TOCSY spectra, confirmed the existence of five ^1H—^1H spin systems:H3—H25, H7—$H_2$9, H11—$H_2$13, $H_3$28—H18 and $H_2$20—H22. A comparison of its NMR data with those reported for the lead compound dehydrothyrsiferol (**7**) indicated that differences were located in the two spin systems within the C11—C18. All fragments were linked using the HMBC correlations of the methyl groups C1, C24, C25, C26, C27, C28, C29 and C30. Thus, the presence of a 3-bromo-2,2,6-trimethyl-oxane ring, together with a dioxabicyclo[4,4,0]decane core and a 2-hydroxy-isopropyl-oxolane ring, was verified. All of these structural motifs were previously found in other oxasqualenoids isolated from *L. viridis*.[3] However, a distinctive spiroketal functionality in **1** was established demonstrated by the correlations of the ketal quaternary carbon C—14 with $H_2$12, $H_2$13, H15, $H_2$16, H18 and $H_3$28.



1 Laurokanol A; $R^1 = CH_3$; $R^2 = H$
2 Laurokanol B; $R^1 = CH_2Cl$; $R^2 = OH$
3 Laurokanol C; $R^1 = OH$; $R^2 = CH_2Cl$
4 Laurokanol D; $R^1 = OH$; $R^2 = CH_2Br$



5 Laurokanol E



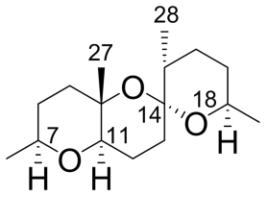
6 Yucatecone

Figure 1. Structure of oxasqualenoids **1**-**6**.

The dipolar correlations between H_3-1/H-3, H-3/H-5α, H_3-25/H_3-26, H-7/H-11, H-8β /H_3-27 and H-9α/H-11 confirmed the relative configuration relationship within the A and B oxane rings as 3*R**,6*S**,7*R**,10*S**,11*R**. In addition, correlations between H-12β/H_3-27, H-13β/H_3-27 and H-13β/H_3-28 indicated that the orientation of the [6,6]-spiroketal ring can only be explained by 14*R** and 15*R** (Figure 1). The relative configuration for the tetrahydropyran rings of the [6,6]-spiroketal system, was proposed from the analysis of $^3J_{HH}$ values determined in TOCSY-1D experiment. were measured in benzene-*d*6, due to signal overlapping in $CDCl_3$.

To confirm the previous conclusions about the relative configurations at C14 and C15, ^{1}H and ^{13}C chemical shifts calculations using density functional theory (DFT) were performed.[10-12] Thus, computational models of the four possible candidate diasteroisomers (**1a-1d** in Figure 2) of the tricyclic core of **1** were built and conformational searches were performed for each one using a hybrid MCMM/low-mode sampling scheme and the MMFF94 force field. Afterward, conformers within a 12 kJ mol^{-1} threshold of the best minimum found were geometrically optimized and their Boltzmann weighted magnetic shielding constants were calculated at the B3LYP-D3/6-31G(d,p) level.[19] Statistical analysis was done by linear regression of the calculated values (δ_{scaled} = (σ_{calc} - intercept)/slope) against experimental ones to correct possible bias introduced by systematic errors. The results were evaluated using three parameters, the correlation coefficient (R^2), the corrected mean absolute error (CMAE), and the maximum absolute error (MaxErr).[13] According to our results, isomer **1a** gave the best outcomes for all parameters (Figure 2). In addition, the DP4+ probability was calculated,[14] giving a best match for isomer **1a** with a probability of 98.8% probability for the configuration. Finally, a comparison between the chemical shifts of the tetrahydrofuran rings of laurokanol A (**1**) and dehydrothyrsiferol (**7**) resulted in almost identical values suggesting identical relative configurations (19*R**, 22*R**).

7*S*,10*S*,11*R*,14*R*,15R,18*R* (**1a**)
7*S*,10*S*,11*R*,14*R*,15S,18*R* (**1b**)
7*S*,10*S*,11*R*,14*S*,15R,18*R* (**1c**)
7*S*,10*S*,11*R*,14*S*,15S,18*R* (**1d**)



Isomer	MaxErr		CMAE		R^2	
	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H
1a	3.39	0.35	1.31	0.11	0.997	0.971
1b	3.45	0.54	1.76	0.21	0.994	0.922
1c	5.51	0.56	2.39	0.24	0.990	0.860
1d	4.26	0.76	2.41	0.28	0.990	0.808

Figure 2. Model substructures (C6—C19) used to perform DFT calculations of the four C14—C15 possible diasteroisomers of laurokanol A (**1**). Statistical parameters obtained for the correlations of experimental vs calculated ^{1}H and ^{13}C chemical shifts. MaxErr is the maximum absolute error. CMAE is the corrected mean absolute error and $\mathbf{R^2}$ is the correlation coefficient.

Four closely related metabolites, laurakonols B-E (**2**-**5**), were also found in this chemical study. Comparison of their NMR data indicated that all of them (**1**-**5**) share the same carbon backbone, but that **2**-**4** incorporated additional functional groups (—OH, —Br and —Cl) at C14 and C15. Differently, laurokanol E (**5**) does not incorporate any halogen atom in its structure and ring A was contracted to an oxolane ring with an isopropanol append, instead of the bromine-containing pyran observed for **1**-**4** (Figure 1).

Yucatecone (**6**) turned out to be an isomer of the lead compound dehydrothysiferol (**7**), according to its molecular formula $C_{30}H_{51}O_6Br$ (ESI-HRMS *m/z* ion peak $[M+Na]^+$ at 609.2772 and 611.2760) and NMR data. The combined study of COSY, HSQC and HMBC spectra confirmed the presence of all the characteristic cyclic system observed in dehydrothyrsiferol (**7**). However, NMR data of **6** revealed differences in the C11—C19 moiety, such as the substitution of the typical olefinic protons $H_2$28 (δ_H 4.88/5.05) by a methyl group (δ_H 0.88), the absence of the proton H18 by oxidation of hydroxyl group on C18 (δ_C 215.7 ppm) and a change in the chemical shifts of H11 and H14.

The dipolar correlations observed in the NOESY experiment of the yucatecone (**6**), in particular those observed between $H_3$1/H3, H3/H5, $H_3$25/$H_3$26, H7/H11, H7/H8α and $H_3$27/H8β, supports that **6** possesses the same relative configuration within the A and B rings as in **7**. However, the relative configuration of C14 was initially established as *R** from the observed correlations between H14 and $H_3$27. The epimerization of C14 represents the first example of this situation in all known thysiferol derivatives so far. This intriguing change led us to further structural studies to confirm the relative configurations at C14 and C15. Thus, we used a *J*-based NMR configurational analysis (JBCA).[10] The measured values for $^3J_{H14,H15}$ = 4.6 Hz, $^3J_{C28,H14}$ = 3.8 Hz, $^3J_{C13,H15}$ = 3.3 Hz, $^3J_{C16,H14}$ = 2.2 Hz and $^2J_{C14,H15}$ = -4.8 Hz, measured from homodecopling experiments, HSQC-HECADE and *J*-HMBC spectra indicated unequivocally the presence of a conformational equilibrium anti—gauche$^-$ with configuration *erythro* (Figure 3). Finally, the oxidation of **7** with pyridinium chlorochromate in anhydrous CH_2Cl_2 yielded the oxidate compound, 18-ketodehydrothyrsiferol, that show almost identical ^{1}H and ^{13}C chemical shift values that **6**, within the C18—C24 portion, supporting that the new compound shares the same configuration in that moiety.

The complex structures of these novel oxaesqualenoids are difficult to explain by the classic sequential cascade cyclizations of squalene-10-11 oxide.[5-7] Thus, laurokanols **1**-**5** are the first examples of polyether triterpenes possessing a tricyclic core including a [6,6]-spiroketal

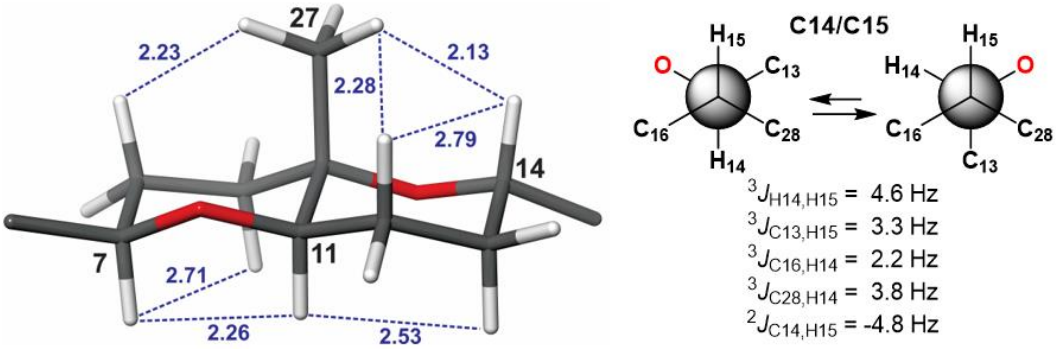
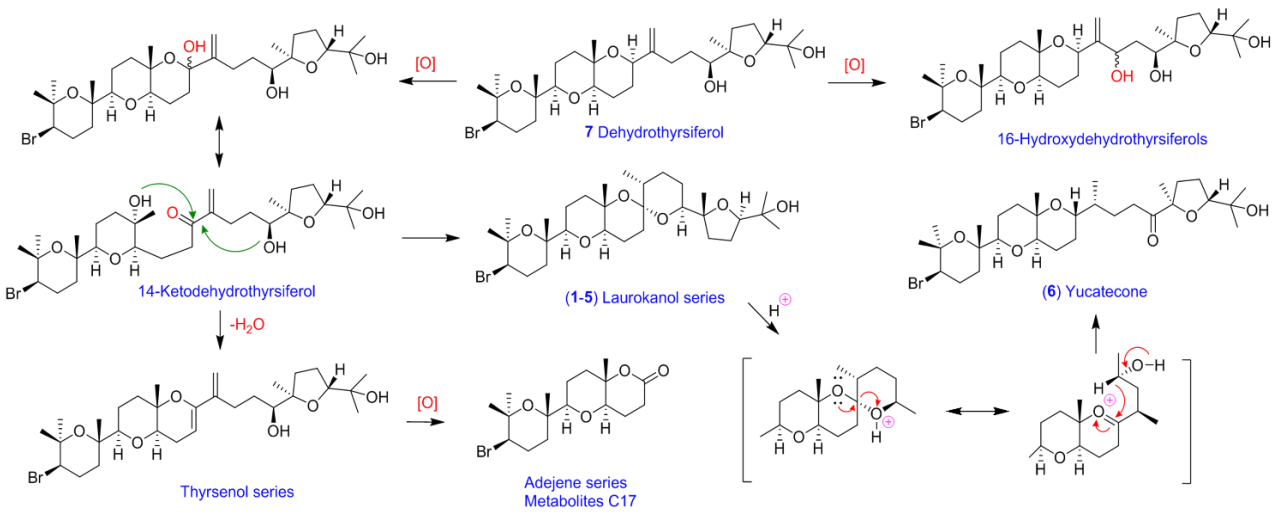


Figure 3. Key dipolar correlations observed within the bicyclic core of yucatecone (**6**) and a summary of the *J*-based NMR configurational analysis around the C14-C15 bond.

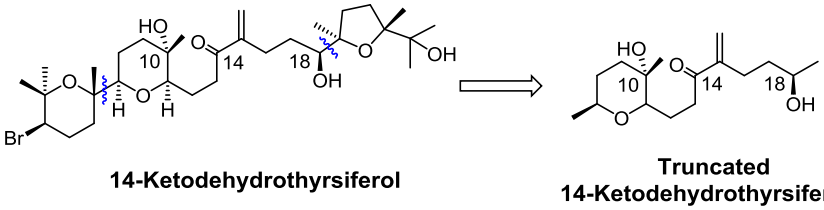
system and **6** shows an intriguing epimerization in the asymmetric position C14. Therefore, we propose a biosynthetic pathway starting from dehydrothirsiferol (**7**), the most abundant secondary metabolite found in *L. viridis*. Allylic oxidation of the most abundant metabolite dehydrothyrsiferol (**7**) would be the key step on the construction of other oxasqualenoids such as 16-hydroxidehydrothyrsiferol and its C16 epimer as well as the intermediate 14-hydroxy derivative of 14-ketodehydrothyrsiferol (**8**).[5,7] The keto form of the last metabolite, would evolve by dehydration to yield the thyrsenol series of compounds and the oxidative cleavage between C14 and C15 in the thyrsenols would yield the adejene series of compounds comprising a C17 backbone.[7] Alternatively, ketalization of the carbonyl C14 of 14-ketodehydrothyrsiferol, would lead to the formation of the laurokanol series of compounds in a similar way to that observed in avermectin and corozalic acid.[16,17] Yucatecone would be synthetized from protonation of the oxygen atom at C-18 that results in the fragmentation of the spiroketals system in laurokanols with the concomitant formation of the oxonium intermediate, followed by hydride migration from C18 to C14 (Scheme 1).

To support the previous general hypothesis, we investigated the mechanisms using computational quantum chemical calculations performed with *Gaussian09*.[18] Structural optimizations and frequency calculations were performed with PCM-(water)-B3LYP(D3)//6-31+G(d,p)[19-21] and PCM-(water)-M06-2X//6-31+G(d,p).[22] All computed transition state structures (TSS) were confirmed with the correct imaginary frequency and by Intrinsic reaction coordinate (IRC) calculations.[23] Minima were confirmed by the absence of imaginary frequency. In order to obtain accurate relative free energy, conformational search using Spartan10[24] were performed on each flexible structure. Merck Molecular Force Field (MMFF) were used in our systematic search. Subsequently, the resulting conformers were subjected to single-point calculation at PCM-(water)-B3LYP(D3)//6-31+G(d,p). Conformers within 4 kcal mol^{-1} to the lowest energy conformer were



Scheme 1. General biosynthetic pathway proposed for compounds **1**-**6**.

then optimized. To shorten computational cost, we also truncated the starting material, 14-ketodehydrothyrsiferol, assuming the truncated parts do not interfere with the reactions of interest (Scheme 2). Further mechanistic investigations were based on this truncated starting material. We began with modelling the cyclization of 14-ketodehydrotyrsiferol. The carbonyl carbon at position C14 is susceptible to attack by the hydroxyl (OH) groups at either C10 or C18, leading to pathway A (Figure 4) or pathway B (Figure 5). The attack can also occur at the *si* or *re* face at C14. Starting with pathway A, the formation of INTA(*S*) is 10.7 kcal mol^{-1} uphill and has a barrier of 14.1 kcal mol^{-1} (Figure 4). Whereas the formation of INTA(*R*) is only 2.73 kcal mol^{-1} uphill has a barrier of 11.9 kcal mol^{-1} (Figure 4).



Scheme 2. Truncation of 14-Ketodehydrothyrsiferol. To reduce computational cost, parts of the structure that are irrelevant to mechanisms of interest are replaced by methyl groups.

Although, both reactions are kinetically feasible at room temperature, attacking C14 by OH at C10 is favored at the *re* face. Subsequently, solvent-assisted dehydration occurs to form INT2A. Then, deprotonation of INT2A affords thyrsenol. However, INT2A can also cyclize to form INT3A(*R*) and INT3A(*S*) with barriers of 7.4 and 12.9 kcal mol^{-1} respectively. Deprotonation of INT3A(*S*) forms INT3B(*S*), which undergoes reduction form laurokanol (Figure 4). The formation of INT3A(*R*) is both kinetically and thermodynamically favored at PCM-(water)-M06-2X//6-31+G(d,p), but unviable at PCM(water)-B3LYP(D3)//6-31+G(d,p). Therefore, we failed to compare the kinetics of forming INT3A(*R*) vs INT3A(*S*). After reduction, INT3A(*S*) and INT3A(*R*) lead to the formation of lauroknaol and laurokanol epimer respectively, with laurokanol thermodynamically favored for 5.37 kcal mol^{-1} over its epimer. We also considered the formation of laurokanol by pathway B, which started with the C14 of the starting material attacked by OH at C18, forming INTB(*R*)/(*S*) (Figure 5). The formation of INTB(*R*) is 13.1 kcal mol^{-1} uphill and has a barrier of 15.2 kcal mol^{-1}. Whereas the formation of INT1B(*S*) has a barrier of only 8.62 kcal mol^{-1} and is 13.1

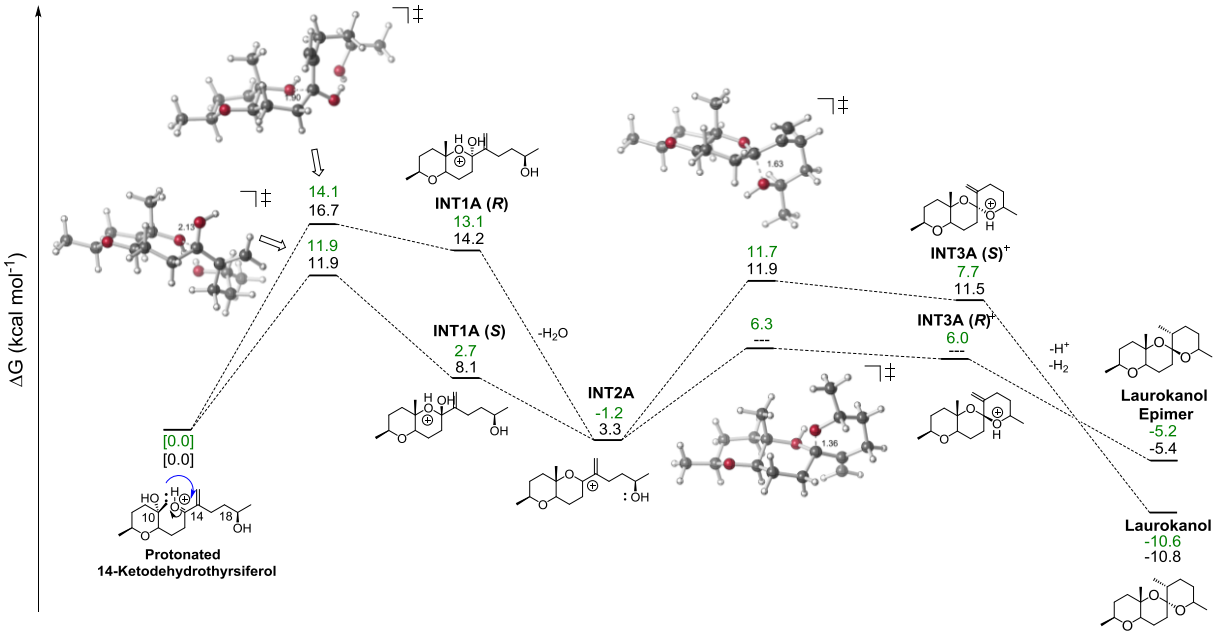


Figure 4. Free energy diagram of pathway A. Energies are calculated using PCM-(Water)-M06-2X//6-31+G(d,p) (green) and PCM-(Water)-B3LYP(D3)//6-31+G(d,p) (black). The protonated form of 14-ketodehydrothyrsiferol, is set as the relative zero. Units are in kcal mol^{-1}. INT3A(*R*)$^{+}$ failed to optimize as minima using PCM-(Water)-B3LYP(D3)//6-31+G(d,p). Three-dimensional representations for TSSs calculated at PCM-(Water)-M06-2X//6-31+G(d,p) are shown.

kcal mol^{-1} uphill (Figure 5). The barriers to form INT1B(*R*) and INT1B(*S*) are similar and attack at either face of C-14 is kinetically feasible. Dehydration of INTB(*R*)/(*S*) forms INT2B, which then cyclizes to form INT3B(*R*)$^+$. However, INT3B(*S*)$^+$ is not a minimum on the potential energy surface using both methods.

Therefore, there is not viable TSS connecting it and INT2B. We optimized INT3B(*S*)$^+$ with its C14 – O18 constrained to 1.58 Å (Figure 5). Although the estimated free energies of INT3B(*S*)$^+$ are similar to INT3B(*R*)$^+$, it is unlikely that INT3B(*S*)$^+$ is formed during the reaction, since C14 – O18 is prone to bond dissociation. Deprotonation of INT3B(R)$^+$ leads to INT3B(*R*), which undergoes reduction to afford the epimer of laurokanol instead. Therefore, we suspected that laurokanol is likely originated from pathway A instead. Subsequently, protonated laurokanol can ring-open to form INT4 with a small barrier of 0.35 kcal mol^{-1} (Figure 6). We were not able to

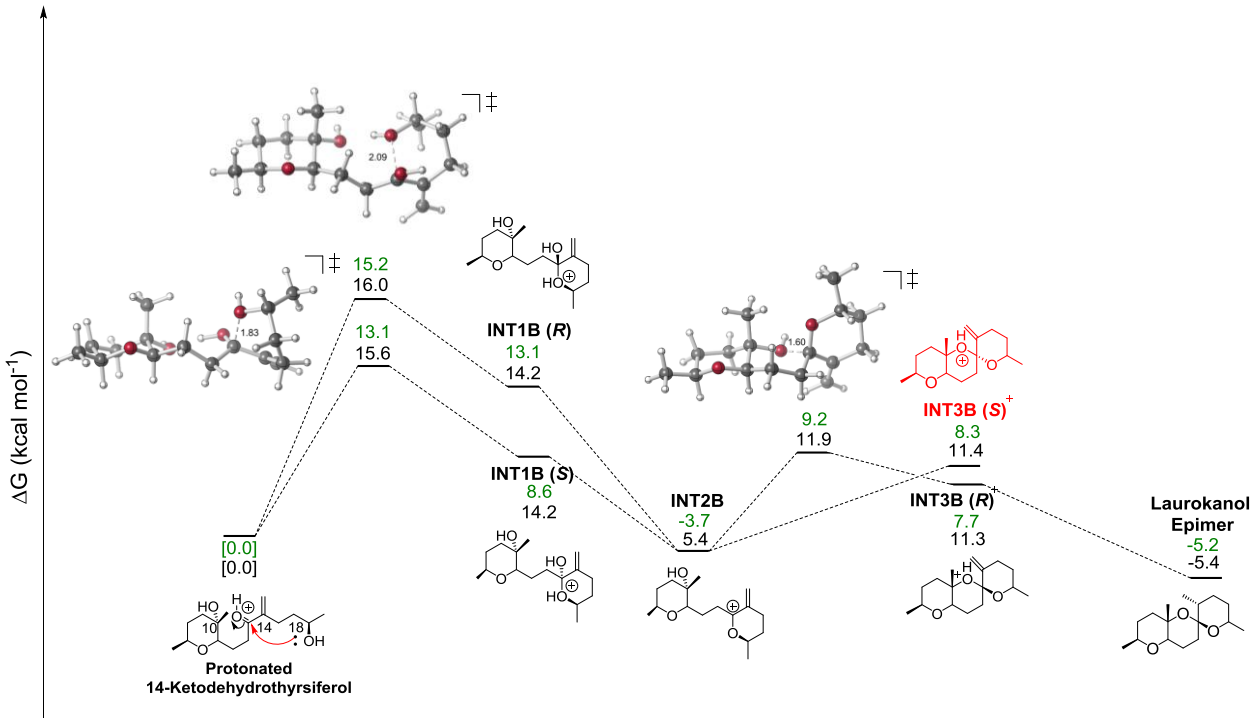


Figure 5. Free energy diagram of pathway B. Energies are calculated using PCM-(Water)-M06-2X//6-31+G(d,p) (green) and PCM-(Water)-B3LYP(D3)//6-31+G(d,p) (black). The starting material, 14-ketodehydrothyrsiferol, is set as the relative zero. Units are in kcal mol^{-1}. INT3B(*S*)$^+$ failed to optimize as minima using both methods, hence its free energies are resulted from constrained optimization with C14 – O18 restricted to 1.58 Å. Three-dimensional representations for TSSs calculated at PCM-(Water)-M06-2X//6-31+G(d,p) are shown.

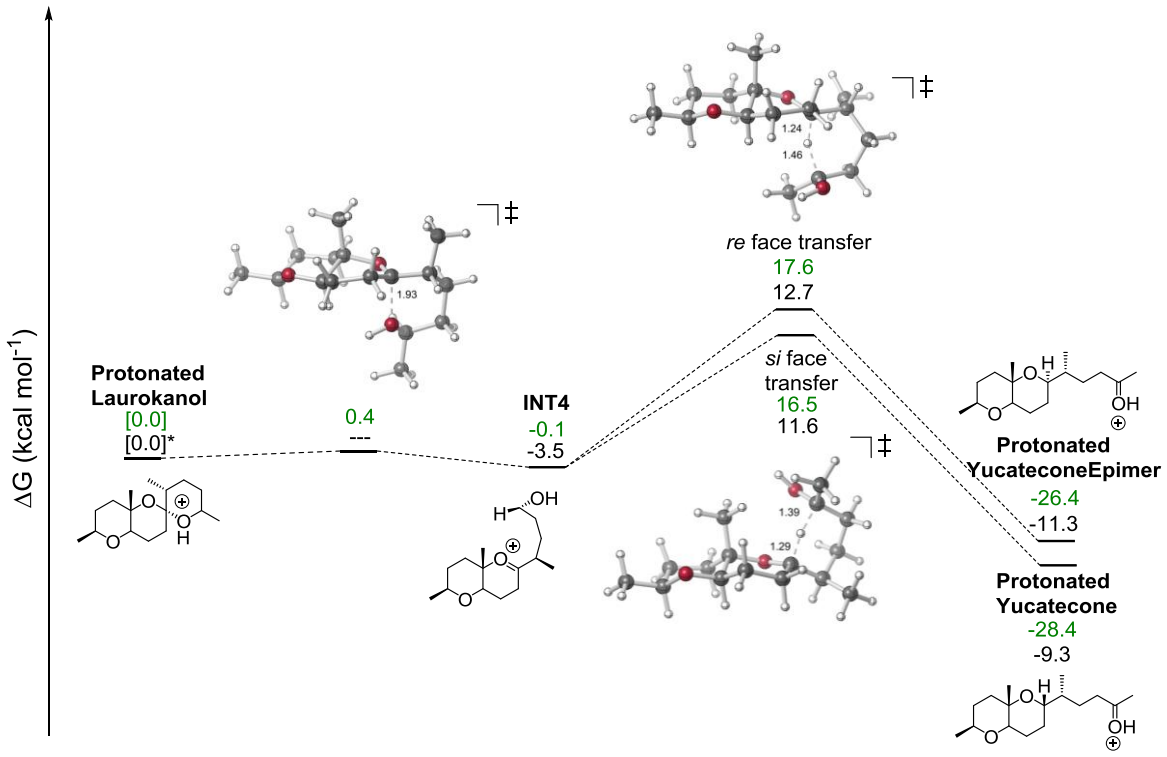


Figure 6. Free energy diagram of yucatecone formation. Energies are calculated using PCM-(water)-M06-2X//6-31+G(d,p) (green) and PCM-(water)-B3LYP(D3)//6-31+G(d,p) (black). Laurokanol is set as the relative zero and units are in kcal mol^{-1}. Protonated laurokanol is not a minimum at PCM-(water)-B3LYP(D3)//6-31+G(d,p). Three-dimensional representations for TSSs calculated at PCM-(water)-M06-2X//6-31+G(d,p) are shown.

optimize protonated laurokanol as a minimum at PCM-(water)-B3LYP(D3)//6-31+G(d,p). This is likely due to the flatness of the potential energy surface, i.e. the susceptibility of ring-opening in laurokanol upon protonation. The hydride at C18 can transfer to the *sp*^{2} hybridized C14 of INT4 at either the *si* or *re* face, with barriers of 16.6 and 17.7 kcal mol^{-1} respectively. Although the formation of both yucatecone and its epimer are kinetically feasible, the formation of yucatecone is favored thermodynamically, agreeing with the absence of its epimer. In conclusion, we have demonstrated the connection between the biosynthesis of thyrsenol and laurokanol, and subsequently yucatecone, where they shared a common precursor, INT1A(*R*)/(*S*), which originated from the nucleophilic attack from OH at C10 in 14-ketodehydrothyrsiferol. However, attack from OH at C18 is unlikely to be productive since INT2B only cyclizes to form INT3B(*R*),

which has the wrong configuration at C14. The formation of INTB(*R*)/(*S*) also have higher barriers than that of INTA(*R*)/(*S*), which is consistent to the absence of its related metabolites.

Reference

1) Harizani, M., Ioannou, E., and Roussis, V., *In Progress in the chemistry of organic natural products*., **2016**, *102*, 91–252

2) Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R., *Nat. Prod. Rep.*, **2019**, *36*, 122-173

3) Fernández, J. J., Souto, M. L., and Norte, M., *Nat. Prod. Rep.*, **2000**, *17*, 235-246

4) Kigoshi, H., Ojika, M., Shizuri, Y., Niwa H., and Yamada, K., *Tetrahedron Lett.*, **1982**, 5413-5414

5) Manríquez, C. P., Souto, M. L., Gavín, J. A., Norte, M., and Fernández, J. J., *Tetrahedron*, **2001**, *57*, 3117-3123

6) Cen-Pacheco, F., Villa-Pulgarin, J. A., Mollinedo, F., Norte, M., Daranas, A. H., and Fernández, J. J., *Eur. J. Med. Chem.,* **2011**, *46*, 3302-3308

7) Cen-Pacheco, F., Nordstrom, L., Souto, M. L., Norte, M., Fernandez, J. J., and Daranas, A., *Mar. Drugs*., **2010**, *8*, 1178-1188

8) Morimoto, Y., *Org. Biomol. Chem.*, **2008**, *6*, 1709-1719

9) Cen-Pacheco, F., Mollinedo, F., Villa-Pulgarín, J. A., Norte, M., Fernandez, J. J., Hernandez Daranas, A., *Tetrahedron*, **2012**, *68*, 7275−7279

10) Cen-Pacheco, F., Rodríguez, J., Norte, M., Fernández, J. J., and Daranas, A. H., *Chem. Eur. J.*, **2013**, *19*, 8525-8532

11) Domínguez, H. J., Crespín, G. D., Santiago-Benítez, A. J., Gavín, J. A., Norte, M., Fernández, J. J., and Daranas, A. H., *Mar. Drugs*., **2014**, *12*, 176-192

12) Lodewyk, M. W., Siebert, M. R., and Tantillo, D. J., *Chem. Rev.*, **2012**, *112*, 1839-1862

13) Saielli, G., Nicolaou, K. C., Ortiz, A., Zhang, H., and Bagno, A., *J. Am. Chem. Soc.,* **2011**, *133*, 6072-6077

14) Smith, S. G., and Goodman, J. M., *J. Am. Chem. Soc.*, **2010**, *132* (37), 12946-12959

15) Norte, M., Fernandez, J. J., Souto, M. L., Gavin, J. A., and Garcia-Gravalos, M. D., *Tetrahedron*, **1997**, *53*, 3173-3178.

16) Ikeda, H., and Omura, S., *Chem. Rev*., **1997**, *97*, 2591-2609

17) Napolitano, J. G., Norte, M., Fernandez, J. J., and Daranas, A., *Chem. Eur. J.,* **2010**, *16*, 11576-11579

18) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., et al. Gaussian 09, Revision D. 01; Gaussian, Inc.: Wallingford, CT, USA, **2010**

19) (a) Becke, A. D., *J. Chem. Phys.*, **1993**, *98*, 5648−5652. (b) Lee, C., Yang, W., and Parr, R. G., *Phys. Rev. B: Condens. Matter Mater. Phys.*, **1988**, *37*, 785−789 (c) Miehlich, B., Savin, A., Stoll, H., Preuss, H., *Chem. Phys. Lett.*, **1989**, *157*, 200−206

20) (a) Grimme, S., Antony, J., Ehrlich, S., and Krieg, H., *J. Chem. Phys*., **2010**, *132*, 154104 (b) Grimme, S., Ehrlich, S., and Goerigk, L., *J. Comp. Chem.*, **2011**, *32*, 1456

(21) Cammi, R., Mennucci, B., Tomasi, J., Leszczynski, J., Ed.; *World Scientific Publishing Co. Pte. Ltd.: Singapore*, **2003**; Vol. 8

(22) (a) Zhao, Y., and Truhlar, D. G., *Theor. Chem. Acc.*, **2008**, *120*, 215-241 (b) Walker, M., Harvey, J. A., Sen, A., Dessent, C. E. H., *J. Phys. Chem. A.,* **2013**, *117*, 47, 12590-12600

(23) (a) Gonzalez, C., and Schelgel, H. B., *J. Phys. Chem.*, **1990**, *94*, 5523-5527 (b) Fukui, K., *Acc. Chem. Res.*, **1981**, *14*, 363-368

(24) Spartan, Version10 (Wavefunction, Ind.: Irvine, CA, USA **2010**)

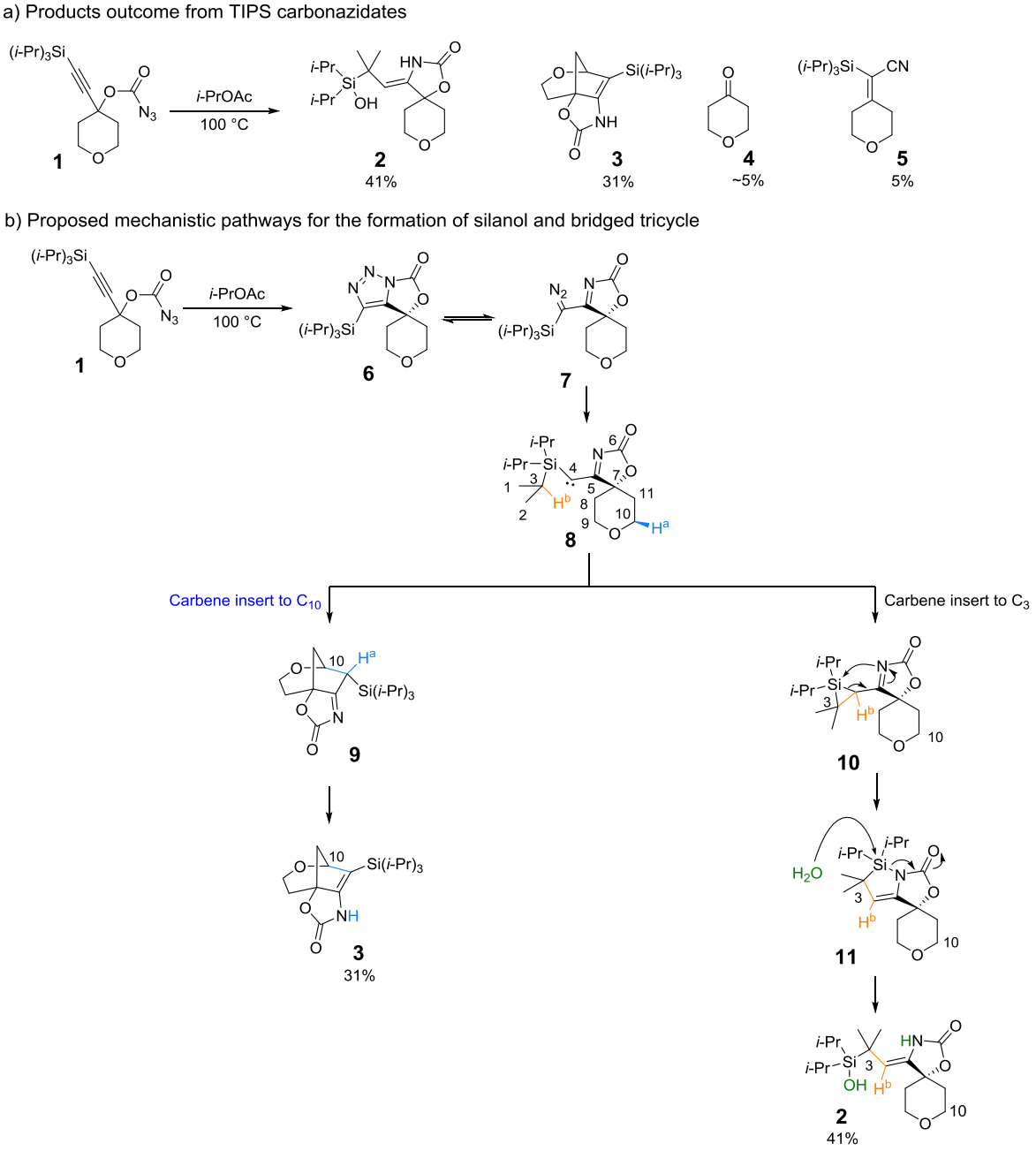
Chapter 5: Cascade Reactions of TIPS Carbonazidate

Portions of this work are reproduced with permission by Qinxuan Wang, who has contributed to the syntheses and experimental structural characterizations for this project. Computational mechanistic studies were performed Croix J. Laconsay and the author of this report.

5.1 Introduction

During the development of Huisgen cyclization/carbene cascade reactions from carbonazidates from a previous study[1], our collaborators found that a silyl alkynyl carbonazidate like **1** (Scheme 1a) formed an unusual silanol product **2** as the major product in 41% yield. They have isolated the anticipated bridged tricyclic product **3** in 31% yield, which illustrated that the carbonazidate could be used as an effective starting material for the construction of bridged azacycles. The reaction also generated two different side products sideproducts: tetrahydropyranone **4**, and silyl vinyl nitrile **5**. The unexpected outcome requires further mechanistic investigation. They did not observe silanol product **2** in the NMR of the crude reaction material; instead, azasilacyclopentene **11** (Scheme 1b) was determined to be the reaction product before an aqueous workup caused silanol formation by hydrolysis of the Si-N bond. Therefore, they proposed the mechanism for the formation of silanol **2** and bridged tricycle **3** shown in Scheme 1b. Carbonazidate **1** can first undergo Huisgen cyclization to form triazole **6**. Then α-imino carbene **7** can be generated via ring opening of triazole **6** followed by extrusion of dinitrogen. After the carbene generation, tertiery C–H bond insertion on an isopropyl silyl group (H^a) leads to the formation of silacyclopropane **10**, which could rearrange in a pseudosigmatropic fashion to azasilacyclopentene **11**. The weak Si–N bond in azasilacyclopentene can undergo rapid hydrolysis and give silanol **2**. Alternatively, a C–H bond insertion could occur with the ether-activated methylene (H^b), which built the bridged iminyl-tricycle **9** and undergo proton transfer to form the bridged enamine-tricycle **3**.

5.2 Results and Discussion



Scheme 1 Cascade reaction of TIPS carbonazidates; a) product outcome from TIPS carbonazidate **1**; b) proposed mechanisms for formations of **2** (by C-H insertion into C_{10}—H^a) and **3** (by C-H insertion into C_3—H^b).The carbon backbone is numbered in **8**.

To verify our proposed mechanisms, density functional theory (DFT) calculations were performed using *Gaussian16*.[2] Without geometrical constraints, optimization and frequency calculations for all structures were performed at the CPCM-(dichloromethane)-ωB97XD//6-311++G(d,p)[3,4] level for all intermediates and transition state structures (TSS). Individuals TSSs were confirmed by both a relevant imaginary frequency and intrinsic reaction coordinate (IRC) calculations.[5] All single and triplet structures were optimized with unrestricted spin and the key word guess=(mix,always). Our calculations predicted that the ground state for **8** is a triplet, with **8 (T)** being 6.41 kcal mol^{-1} lower in energy than **8 (S)**. The triplet state is likely stablilized by the electropositive Si and the neighbouring π-donating system. However, we are unsure about the accuracy of the S-T gap, which has been a known chllange for theoretical methods due to the energetic proximity between the two states and the difficulties in fair comparison of the open- and closed-shell systems.[6,7] Here, we selected a DFT method, since it includes correlation within an SCF formalism and has less spin contamination.[7] The performance of DFT methods on calculating the S-T gap also have been accessed in previous studies.[8] Here, our analysis set the free energy of **8 (T)** the relative zero. There are several sites in **8** where the carbene center C_4 can react intermolecularly. First, we discuss the C_4 insertion into C_{10}—H^a with a singlet barrier of 7.39 kcal mol^{-1} (Scheme 1 and Figure 1). This leads to the formation of a bridged iminyl-tricyclic intermediate **9 (S)**. Solvent-assisted proton transfer of **9 (S)** leads to the enamine-tricyclic product **3 (S)**, which is 54.3 kcal mol^{-1} themodynamically downhill relative to **8 (T)**. Alternatively, C_4 can insert to C_3—H^b with a singlet barrier of 18.9 kcal mol^{-1}, which results in the formation fo a silacyclopropane intermediate **10** (Scheme 1 and figure 1). Through a pseudopericylclic [1,3] sigmatropic shift, **10 (S)** forms **11 (S)** with a barrier of 11.3 kcal mol^{-1} (Figure 1). These two transitions can also occur on the triplet surface, with both barriers being kientically feasible (31.2 kcal mol^{-1} and then 14.1 kcal mol^{-1}). However, the singlet pathway will likely outcompete the triplet pathway due to its lower barriers. Subsequently, water adds to **11 (S)**, followed by a solvent-

assisted tautomerization, forming **2 (S)**. The formation of **2 (S)** is 55.8 kcal mol^{-1} favored than **2 (T)**.

Base on the calculated barriers, our results also predicted that the formation of **2 (S)** is thermodynamically favored over **3 (S)**, but formation of **3 (S)** is kinetically favored over **2 (S)**. We have also considered the formation of **3** on the triplet surface but the triplet versions of all intermediates invovled have consistently higher free energies. The TSS also exceed the limit to be kinetically feasible. Thus, it is likely that **8 (T)** has to be excited to **8 (S)** to form **3 (S)**. However, we also cannot eliminate the possibility of surface crossing throughout the mechanisms.

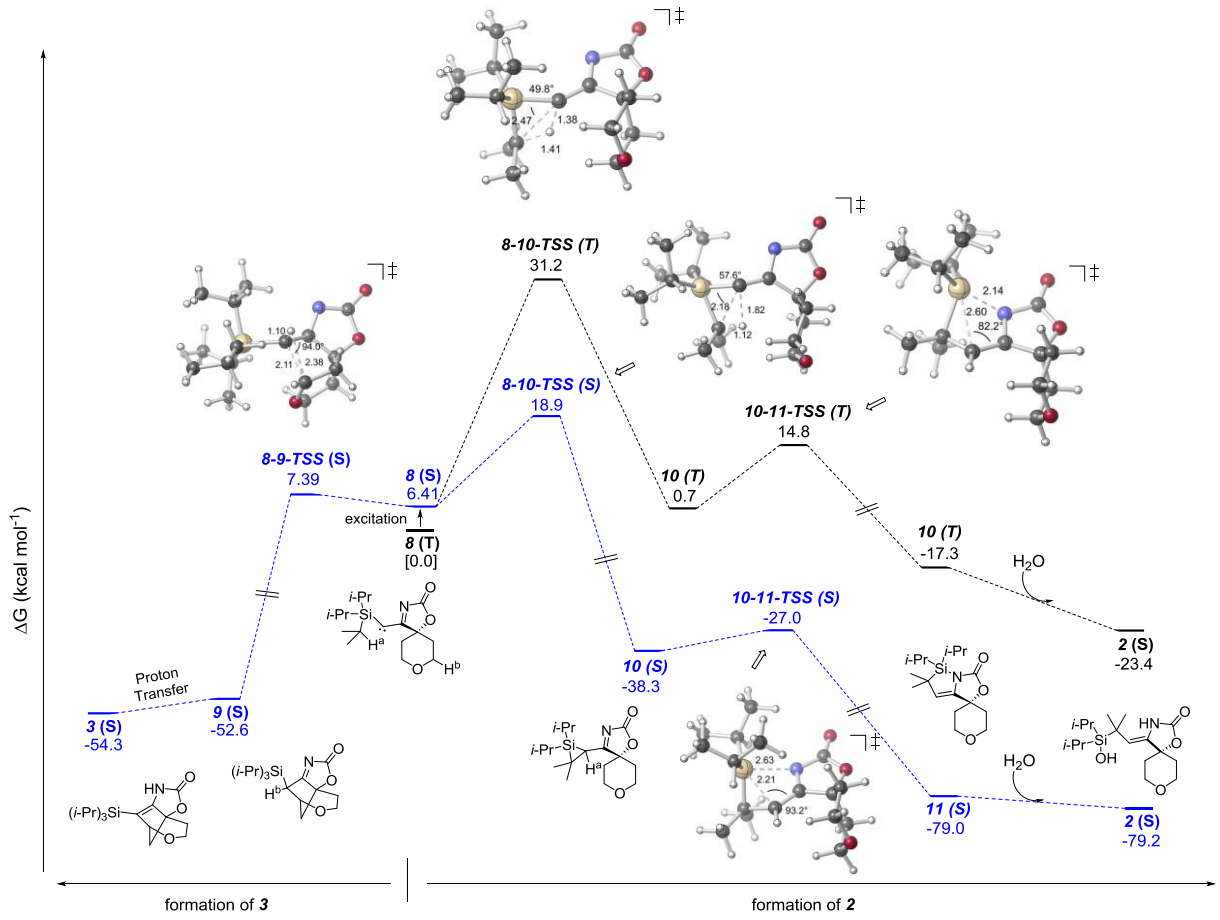
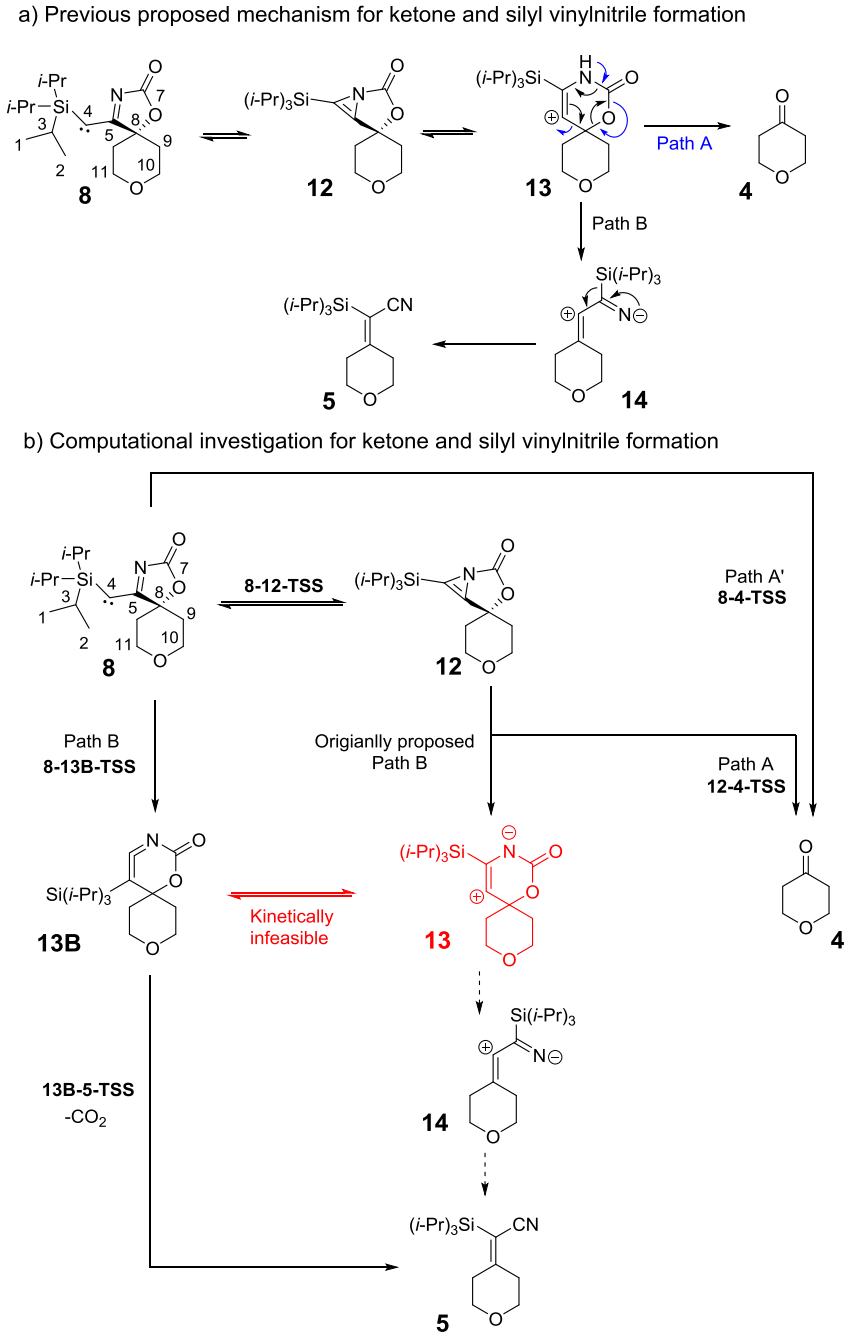


Figure 1. Free energy diagram for mechanisms forming **2 (S)** and **2 (T)** (right) and **3 (S)** (left). Mechanisms occur on the singlet surface are in blue. Energies are reported in kcal mol^{-1} but not drawn to scale. Only singlet TSS and intermediates are reported for the formation of **3**. Triplet versions of these structures were also considered but their calculated barriers exceeds the limit of kinetic feasibility under the given reaction temperature.



Scheme 2. a) Possible mechanistic pathways for the ketone **4** and vinyl nitrile **5** formation; b) pathways investigated in our calculations. The carbon backbone is numbered in **8**.

The formation of ketone **4** and silyl vinylnitrile **5** is interesting. In our original communication, we proposed the formation of a highly strained bicyclic arizine intermediate **12**

via electrocyclization of α-imino carbene **8** (Scheme 2a). An internal C–N bond fragmentation led to the ring expanded product **13**. Two pathways could follow: a) C–C bond fragmentation to provide ketone **4**, or b) decarboxylation via retro-hetero Diels-Alder reaction followed by a 1,2-silyl migration to yield the silyl vinylnitrile **5** Staring from path A, we were able to optimize a concerted TSS from **12 (S)** that leads to **4 (S)** and **4a (S)** but the barrier is 55.1 kcal mol^{-1}, which exceeds the limit of being kinetically feasible under the reaction temperature (Figure 2). We also failed to identify any alternative mechanism from **12 (S)** to **4 (S)**. Alternatively, according to our computational results, **8 (S)** can fragement into **4 (S)** and **4a (S)** in a single step with a barrier of 12.9 kcal mol^{-1} (Figure 2, top). Therefore, **4 (S)** is likely formed directly from **8 (S)** instead of **12 (S)**. **4(a)** is an interesting discovery due to its ethyne isocyanate substructure. Alkynyl isocyanates have been previously synthesized and characterized,[9] and are also invovled in different fields in chemistry, such as polymer chemistry, atmospheric chemistry, and combination of interstellar species.[10] Here, we expect **4a** to react further and possibly contributes to the formation of other side products.

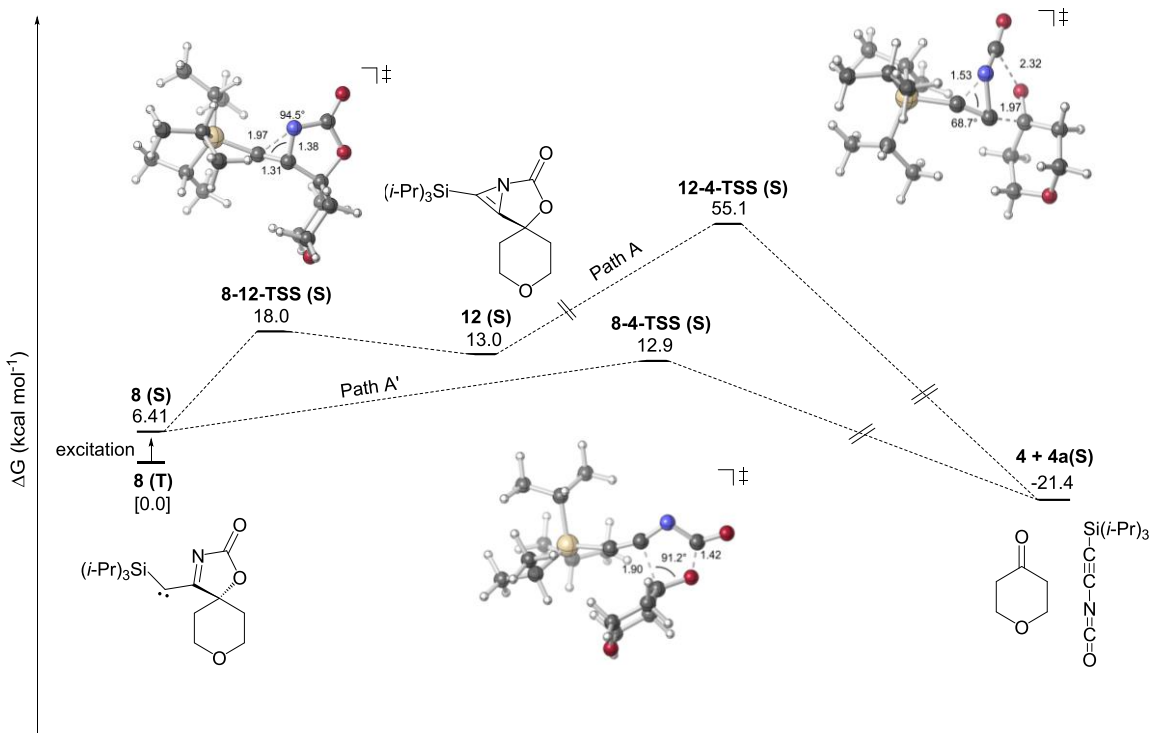


Figure 2. Free energy diagram for mechanisms forming **4 (S)** Free energies are reported in kcal mol^{-1} but not drawn to scale. Only singlet TSS and intermediates are reported. Triplet versions of these structures were also considered but their calculated barriers exceeds the limit of kinetic feasibility under the given reaction temperature.

For path B, however, we failed to identify the any kinetically transitions states from **8** or **12** to **13** by scanning the potential energy surface. This leads us to propose an alternative mechanism to the original path B. Through a 1-2 shift, **8 (T)** forms **13B (T)** endergonically with a high barrier of 72.0 kcal mol^{-1} (Figure 3). Although we failed to optimized any viable transition states on the singlet surface from **8** to **13B**, we suspect singlet-triplet surface crossing possibly accounts for the formation of **13B (S)**, which is 22.5 kcal mol^{-1} more favorable than **13B (T)**. We have analysed the IRC calculation from **8 (T)** to **13B (T)** and performed sinlget single-point calculation on each geometry along the reaction coordinate. These singlet-point calculations show a crossing on the potential energy surface and a non-zero S^2 values from **8 (S)** to **13 B (S)** along the reaction coordinate (Figure 3). Such spin contamination is charactersistic of surface crossing. In an highly exogeonic step, **13B (S)** decorboxylate to **5 (S)** with a barrier of 13.1 kcal mol^{-1}. However, this

transformation on the triplet surface is not kientically feasible due to a high barrier. We are currently invesitgating to identify a more kinetically favorable mechanism for this pathway. Compare to the formation of **2** and **3**, the overall barriers froming **4** and **5** are higher. Our computional predictions agree with the observed experimental yields.

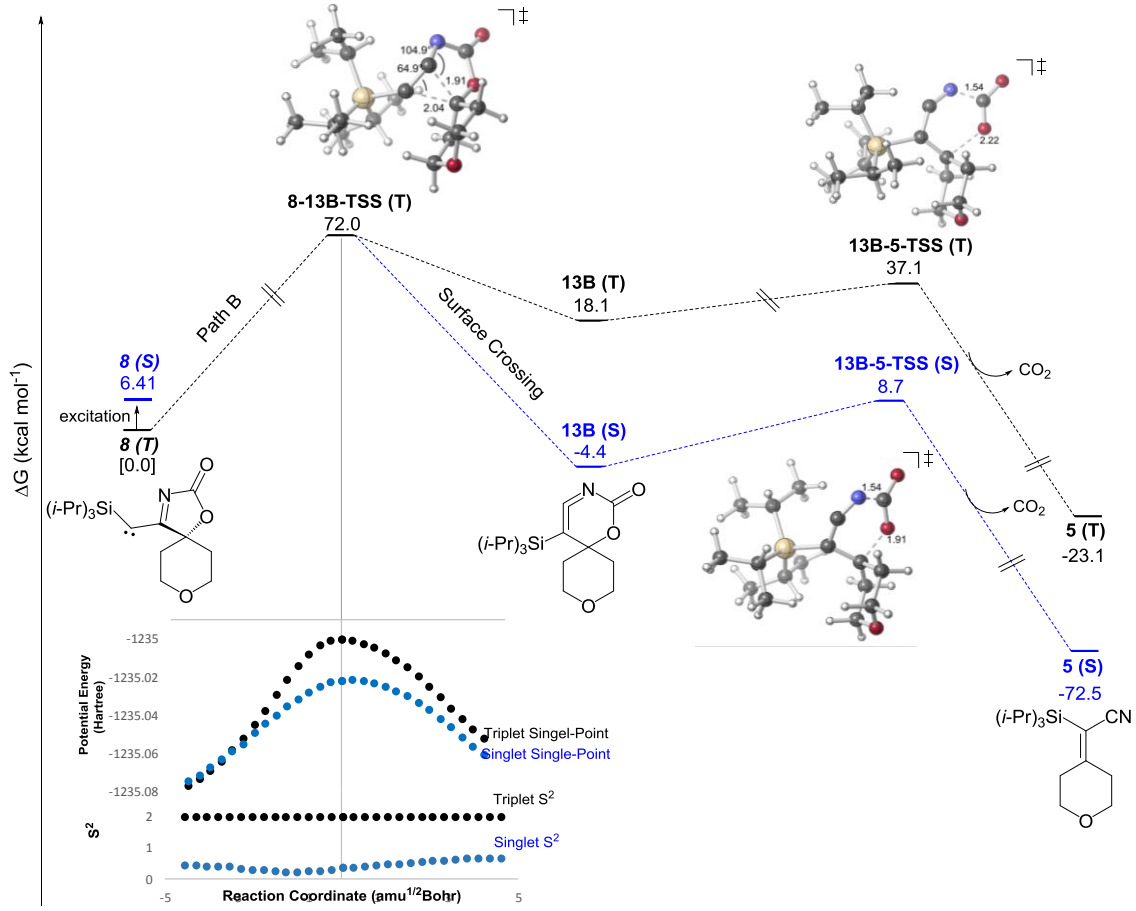


Figure 3. Free energy diagram for mechanisms forming **5 (S).** Triplet and singlet free energies are shown in black and blue respectively. Triplet and singlet potential energies and S^2 values along the IRC are shown in indent.

Reference

1) Shih, J. -L., Jansone-Popva, S., Huynh, C., and May, J. A., *Chem. Sci.*, **2017**, *8*, 7132

2) Gaussian 16, Revision C.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R.; Hasegawa, J.; Ishida, M., Nakajima, T., Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N., Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P., Burant, J. C.; Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M.; Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J., Gaussian, Inc., Wallingford CT, **2016**

3) Cammi, R., Mennucci, B., Tomasi, J., Leszczynski, J., Ed.; *World Scientific Publishing Co. Pte. Ltd.: Singapore*, **2003**; Vol. 8

4) Mardirossiana, N., and Head-Gordon, M., *Phys. Chem. Chem. Phys.*, **2014**, *16*, 9904-9924

5) (a) Fukui, K., *Acc. Chem. Res.*, **1981**, *14*, 363-368; (b) Gonzalez, C., and Schlegel, H. B., *J. Phys. Chem.*, **1990**, *94*, 5523-5527

6) (a) Davidson, E. R., *In Diradicals; Borden, W. T., Ed.; Wiley-Interscience: New York*, **1982**, 73 (b) Schaefer, H. F. III, *Science.*, **1986**, *231*, 1100

7) (a) Baker, J., Scheiner, A., and Andzelm, J., *J. Chem. Phys. Lett.*, **1993**, *216*, 380; (b) Cramer, C. J., Dulles, F. J., Falvey, D. E., *J. Am. Chem. Soc.*, **1994**, *116*, 9787

8) Henkel, S., Costa, P., Klute, I., Sokkar, P., Fernandez-Oliva, M., Thiel, W., Sanchez Garcia, E., and Sander, E., *J. Am. Chem. Soc.*, **2016**, *138*, 1689-1697

9) Qin, Y., Lu, B., Rauhut, G., Hagedorn, M., Banert, K., Song, C., Chu, X., Wang, L., and Zeng, X., *Angew. Chem. Int. Ed.*, **2019**, *58*, 17277-17281

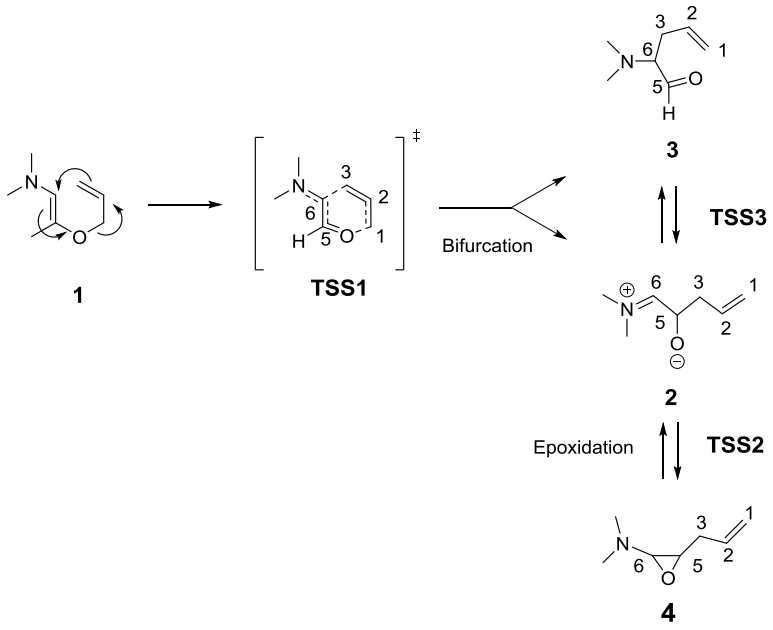
(10) (a) Vrtilek, J. M. Gottlieb C. A., Langer, W. D., Thaddeus, P., and Wilson, R. W., *Astrophys. J.* **1985**, *296*, L35-L38 (b) Pasinszki, T., and Havasi, B., *Phys. Chem. Chem. Phys.*, **2003**, *5*, 259-267 (c) Lattanzi, V., Gottlieb, C. A., Thaddeus, P., Thorwirth, S., and McCarthy, M. C., *Astrophys. J.,* **2010**, *720*, 1717-1720 (d) Fourre, I., Matz, O., Ellinger, Y., Guillemin, J. C., *A&A.*, **2020**, *639*, A16

Chapter 6: Systematic Study of Explicit Solvent Molecular Dynamics

6.1 Introduction

Solvent can have crucial effects on reactivity of organic molecules. In quantum chemical simulations, implicit solvent models are widely applied when solvent effects are important in the reaction of interest. Implicit solvent models represent a solvated molecule inside a molecule-sized electrostatic cavity surrounded by a dielectric medium, which depicts the solvent.[1-5] In the PCM (Polarizable Continuum Model) method the solute is inserted to a cavity of a continuum medium that is fully characterized by its dielectric constant.[2] CPCM (Conductor-Like Polarizable Continuum Model) is a variation of the PCM method, which treats the solvent as conductors, not dielectric and it is commonly used for systems in aqueous environment.[3] In this case, electrostatic interactions are calculated numerically. The SMD model is another variation of the PCM method that is universally used. It is based on the quantum mechanical charge density of a solute molecule interacting with a continuum representation of the solvent.[4]

However, to describe more realistic and accurate solvent-solute interactions, explicit solvent models are preferred, which is focused in this report. The reaction of interest is a pericyclic sigmatropic shift. Based on a previous studies[6,7], this type of concerted and cyclic rearrangement can plausibly lead to two distinct products, i.e. reaction is suspected to undergo post transition state bifurcation (PTSB) (Scheme 1). If the bond C3—C6 is formed during **TSS1**, the 3,3-sigmatropic shift product **3** is resulted. Moreover, a zwitterionic 2,3-shift product **2** is also plausible if the C3—C5 is formed, which can be stabilized by a polar solvent environment. However, without such stabilization, **2** will likely not be a minimum on the potential energy surface (PES). In this



Scheme 1. Mechanism for conversions of **1** to **2**, **3**, and **4**. **TSS1** is suspected to be bifurcating. **TSS2** converts between **2** and **4**, while **TSS3** converts between **2** and **3**.

case, **4** is formed instead. In this report, we systematically investigate the minimum amount of explicit solvent molecules required to reach a converged result of product selectivity of a bifurcating reaction.

6. 2 Results and Discussion

Quantum chemical calculations were performed using *Gaussian16*.[8] Structural optimizations and frequency calculations were performed at B3LYP/6-31-G(d)[9] in the gas phase, PCM-(water), and SMD-(water). [2,4] All structures were confirmed by frequency analysis, with all TSSs having one correct imaginary frequency and minima having none.

To begin our investigation, we have optimized structures in Scheme 1 in gas phase and in a list of CPCM implicit solvents with varying dielectric constants (ε) (Table 1). For comparison, the free energy of **1** is set to relative zero. Intrinsic reaction coordinate (IRC) calculations[10] were performed on **TSS1**, linking **TSS1** to **1** and **3**. For all cases shown in Table 1, **2** is not a minimum

on the PES. We also failed to find a direct TSS connecting **1** to **4**, but the relative energy of **4** is shown in Table 1. To test if **TSS1** bifurcates under gas phase and a list of CPCM implicit solvents, quasi-classical dynamic calculations were performed, using the program *Progdyn*. Using the normal modes from a standard Gaussian frequency calculation, *Progdyn* samples an initial structure from a given thermodynamic ensemble, then computes a trajectory by the Velocity-Verlet algorithm.[11] Since we are interested in using **TSS1** as the starting point to see what product(s) the trajectories lead to, this report focuses only on downhill dynamics. Table 2 shows the dynamic results when **TSS1** is optimized in gas phase and different CPCM implicit solvents. Majority of the trajectories leads to the formation of **3** for all cases. There are a small portion (< 10%) of recrossing from **3** to **3**. In all scenarios in Table 1, geometries resembling **2** and **4** were not observed in any of the trajectories in our dynamic simulations.

We proceed to question if the addition of explicit solvent molecules can change the dynamic results. We first added two explicit water molecules to the system with also the inclusion of SMD-(water) and CPCM-(water), respectively. We were then able to optimize **2** as a minimum on the PES. This is likely due to the stabilizing interactions of the explicit solvent molecules with the charge separation in **2**. We were also able to optimize **TSS2** and **TSS3** (Figure 1).

Table 1. Free energies calculated at B3LYP/6-31G(d) in gas phase and CPCM implicit solvents.

	ε	Activation Barrier	ΔG 1 → 3	ΔG 1 → 4
Gas Phase	---	23.6	-11.7	1.91
Toluene	2.38	22.7	-13.4	1.77
DCM	8.93	22.7	-13.3	1.79
1-Butanol	~17	21.9	-13.3	1.69
Methanol	32.7	21.8	-13.3	1.68
DMSO	46.7	21.8	-13.3	1.67
Water	80.1	18.1	-13.0	-2.03

Table 2. Downhill dynamic results at B3LYP/6-31G(d) in gas phase and CPCM implicit solvents.

	1 to 2	1 to 3	1 to 4	1 to 1 Recrossing	3 to 3 Recrossing	3 to 1 Recrossing
Gas Phase	0 (0%)	220 (96.9%)	0 (0%)	0 (0%)	7 (3.1%)	0 (0%)
Toluene	0 (0%)	95 (99.0%)	0 (0%)	0 (0%)	1(1.0%)	0 (0%)
DCM	0 (0%)	87(95.6 %)	0 (0%)	0 (0%)	4 (4.4%)	0 (0%)
1-Butanol	0 (0%)	89 (95.7%)	0 (0%)	0 (0%)	4 (4.3%)	0 (0%)
Methanol	0 (0%)	89 (95.7%)	0 (0%)	0 (0%)	4 (4.3%)	0 (0%)
DMSO	0 (0%)	83 (92.2%)	0 (0%)	0 (0%)	7 (7.8%)	0 (0%)
Water	0 (0%)	86 (95.6%)	0 (0%)	0 (0%)	4 (4.4%)	0 (0%)

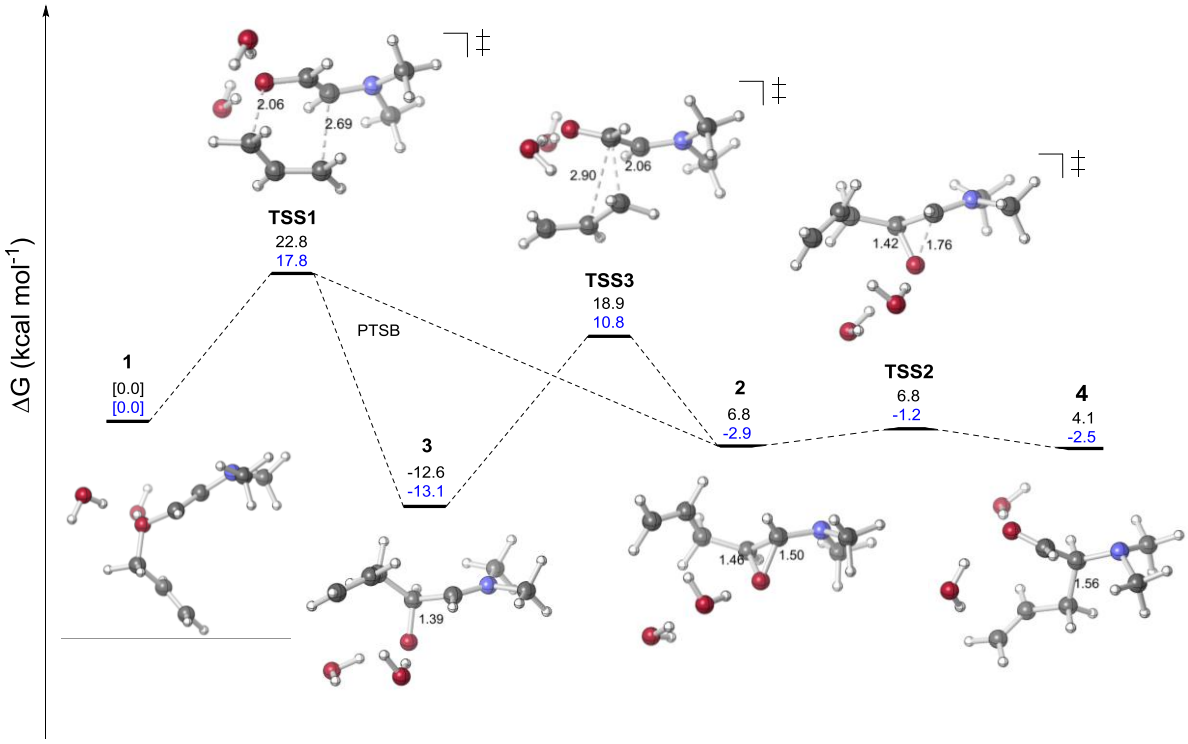


Figure 1. Free energy diagram at CPCM-(water)-B3LYP//6-31G(d) (black) and SMD-(water)-B3LYP//6-31G(d) (blue), with the inclusion of 2 explicit water molecules. Units are in kcal mol^{-1}.

Table 3. Comparison of downhill dynamic results including 2 vs 0 explicit water molecules.

	1 to 3	**1 to 2**	**1 to 4**	**1 to 1**	**Other Recrossing**
2 Explicit Water (cpcm)	69 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
0 Explicit Water (cpcm)	86 (96%)	0 (0%)	0 (0%)	4 (4%)	0 (0%)
2 Explicit Water (smd)	81 (87.1%)	1 (1.1%)	0 (0 %)	1(1.1%)	10 (10.7 %)
0 Explicit Water (smd)	104 (97%)	0 (0%)	0 (0%)	0 (0%)	3 (3%)

Note: The categories for "Other Recrossing" refers to **3** to **3**, **3** to **1**, **2** to **3**, and **3** to **2**.

After demonstrating **2** is a minimum with inclusion of explicit solvents, we sought to test if **2** can be originated from **TSS1** by running dynamic simulations. Table 3 compares the results from our dynamic calculations including two vs zero explicit water molecules. The top two entries of the dynamic results when CPCM-(water) is used, whereas SMD-(water) is used with the bottom two entries. Only one trajectory from **TSS1** to **2** were observed in the case when two explicit solvent molecules and SMD-(water) were included. Based on these results, it is inconclusive whether **TSS1** actually bifurcates. Therefore, we have decided to systematically increase the number of explicit water molecules to our system. Table 4 shows the free energies calculated for our system with the inclusion of 0, 2, 4, 6, and 8 explicit water molecules at CPCM-(water)-B3LYP//6-31G(d). Again, **1** is set as the relative zero. When only implicit solvent is included **2** is not a minimum of the PES, however, with the number of explicit water molecules included, the formation of **2** is favored over **4**. When six or more explicit water included to our calculations, **4** is no longer a minimum on the PES.

Table 4. Free energies calculated at CPCM-(water)-B3LYP//6-31G(d) Including 2, 4, 6, and 8 explicit water molecules.

	Activation Barrier	**ΔG 1 → 3**	**ΔG 1 → 2**	**ΔG 1 → 4**
8 Explicit Water	17.3	-13.5	-9.48	---
6 Explicit Water	21.3	-12.7	-2.10	---
4 Explicit Water	22.4	-10.2	2.53	1.75
2 Explicit Water	22.8	-12.6	6.85	4.14
0 Explicit Water	18.1	-13.0	---	-2.03

Table 5. Downhill dynamic results at CPCM-(water)-B3LYP//6-31G(d) Including 2, 4, 6, and 8 explicit water molecules.

	1 to 3	1 to 2	1 to 1	Other Recrossing
8 Explicit Water	59 (77.6%)	2 (2.6%)	0 (0%)	15 (19.7%)
6 Explicit Water	60 (92.3%)	0 (0%)	0 (0%)	5 (7.4%)
4 Explicit Water	58 (94.0 %)	0 (0%)	0 (0%)	4 (6.0 %)
2 Explicit Water	69 (100%)	0 (0%)	0 (0%)	0 (0%)
0 Explicit Water	86 (96.0 %)	0 (0%)	4 (4.0 %)	0 (0%)

Note: The categories for "Other Recrossing" refers to **3** to **3**, **3** to **1**, **2** to **3**, and **3** to **2**.

IRC analysis only shows **TSS1** connecting to **1** and **3** in all cases shown in Table 3. This is, however, expected with the case of a PTSB, since IRC generates a reaction path based on the steepest decent method. To confirm the connection between **TSS1** and **2**, we perform further dynamic calculations on our systems with additions of 4, 6, and 8 explicit water molecules respectively, also including a background CPCM-(water) implicit solvent environment. Table 5 shows the results from our dynamic calculations. After increasing the number of explicit solvent molecules, however, we still observed no trajectories from **TSS1** to **2**. We then questioned if a different background implicit solvent environment can affect the dynamic results. Therefore, we repeated explicit solvent calculations using SMD-(water) instead. With the SMD model, full solute electron density is used without defining partial atomic charges.[3] The observable solvation free energy is separated into two main components: 1) the bulk electrostatic contribution arising from a self-consistent reaction field (SCRF) treatment and 2) the cavity-dispersion solvent-structure

term contributed by the short-range interactions between the solute and solvent molecules in the first solvation shell.[4]

Figure 2 shows the optimized structure for TSS1 at CPCM-(water)-B3LYP//6-31G(d) and SMD-(water)-B3LYP//6-31G(d) with the inclusion of 2, 4, 6, and 8 explicit water molecules respectively. Table 6 shows the SMD-(water)-B3LYP//6-31G(d) free energies calculated for structures in our system. The energetics of **2** and **4** agree with our previous CPCM-(water) results: **2** is not a minimum when no explicit water was added, whereas **4** was not a minimum when six or more water molecule were included. Again, IRC calculations from **TSS1** only connects to **1** and **3**. Table 7 compares the dynamic results when 0, 2, 4, 6, and 8 explicit solvent molecules were introduced to our system, including background SMD-(water). Unlike our previous dynamic results in Table 5, when SMD-(water) was used as the background implicit solvent, the percentage of trajectories from **TSS1** to **2** rises as the number of explicit water molecules increases.

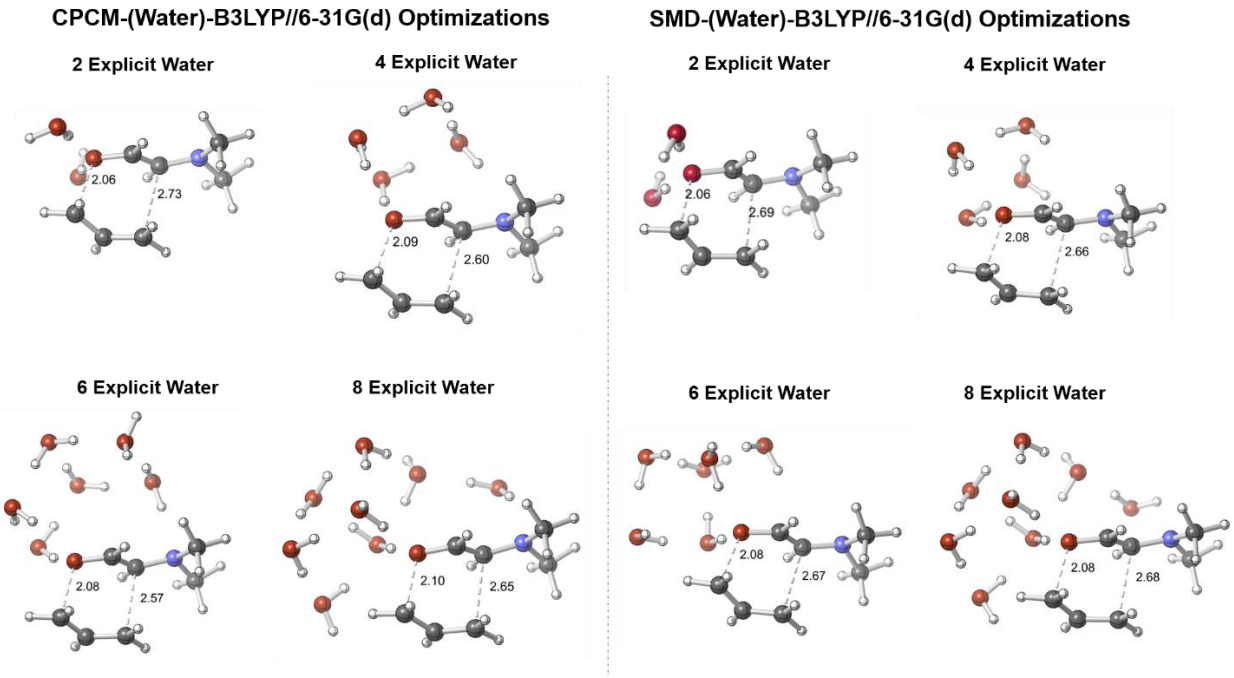


Figure 2. Optimized geometries of **TSS1** with addition of 2, 4, 6, and 8 explicit water molecules. CPCM-(water)-B3LYP//6-31G(d) (left) and SMD-(water)-B3LYP//6-31G(d) (right) **TSS1** structures are shown.

Table 6. Free energies calculated at SMD-(water)-B3LYP//6-31G(d) Including 2, 4, 6, and 8 explicit water molecules.

	Activation Barrier	ΔG 1 → 3	ΔG 1 → 2	ΔG 1 → 4
8 Explicit Water	15.8	-16.3	-13.7	---
6 Explicit Water	15.2	-16.0	-9.12	---
4 Explicit Water	22.2	-13.2	-3.15	-0.31
2 Explicit Water	17.8	-13.1	-2.94	-2.52
0 Explicit Water	20.4	-14.5	---	-0.57

Note: The categories for “Other Recrossing” refers to **3** to **3**, **3** to **1**, **2** to **3**, and **3** to **2**.

Table 7. Downhill dynamic results at SMD-(water)-B3LYP//6-31G(d) Including 2, 4, 6, and 8 explicit water molecules

	1 to 3	1 to 2	1 to 1	Other Recrossing
8 Explicit Water	39(75%)	5 (9.6%)	0 (0%)	8 (15.4%)
6 Explicit Water	58 (85.3%)	4 (5.9%)	2(2.9%)	4 (5.9%)
4 Explicit Water	94 (91.2%)	4 (3.9%)	0 (0%)	5 (4.9%)
2 Explicit Water	81 (87.1%)	1 (1.1%)	1(1.1%)	10 (10.7 %)
0 Explicit Water	104 (97.0 %)	0 (0%)	0 (0%)	3 (3.0 %)

Note: The categories for “Other Recrossing” refers to **3** to **3**, **3** to **1**, **2** to **3**, and **3** to **2**.

Although our calculations for the 6 explicit water and 8 explicit water systems have resulted in less trajectories due to an exponentially increased computational cost, we observed the trajectories from **TSS1** to **2** earlier on. With enough computationally resource to run more parallel dynamic calculations for these two cases, we expect to see an increased number of trajectories from **TSS1** to **2**. In addition, for the 8 explicit water system, we noticed a high percentage of recrossing trajectories. Recrossing can occur when the TSS is located at a flat region on the energy surface or the input TSS in the dynamic calculation deviates from the variational TSS. To investigate the reason of large percentage recrossing shown in Table 7, we subjected the optimized **TSS1** geometries to variational reaction path calculations, using the program *Polyrate*.[12] This allows us to identify the variational TST on the free energy surface.[13] Figure 3 shows the variational reaction path results at SMD-(water)-B3LYP//6-31G(d) including 2, 4, 6, and 8 explicit water molecules. On the X-axis, the zero point represents the input optimized TSS

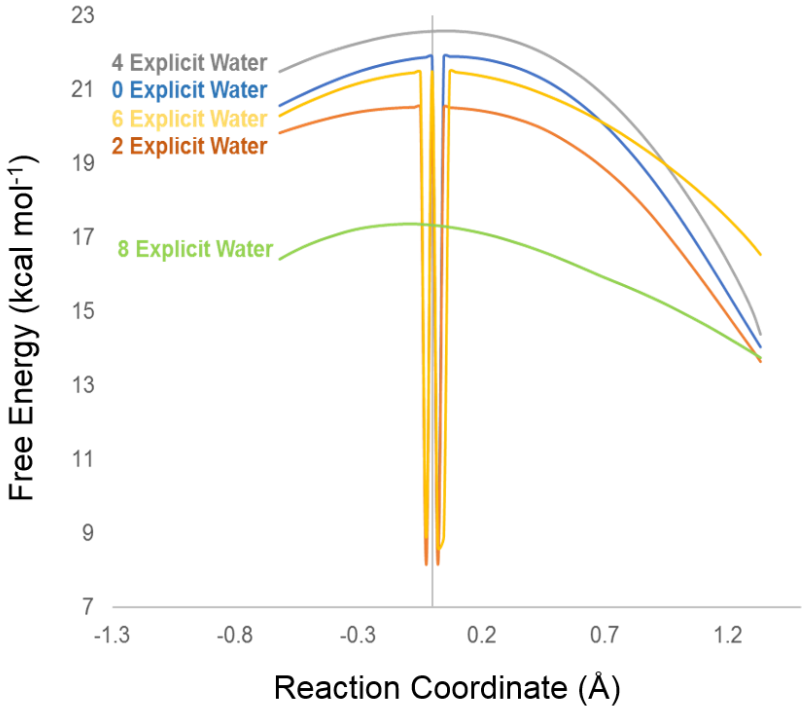


Figure 3. Polyrate reaction path results at SMD-(water)-B3LYP//6-31G(d) including 2, 4, 6 and 8 explicit water molecules. Y-axis is relative free energy in kcal mol^{-1}. X-axis is the reduced reaction coordinate in Å. X=0 represents the input optimized TSS geometries. For the 0, 2, and 6 explicit water systems, the dip at =0 represents the artifacts from the *Polyrate* calculations.

geometry from a standard Gaussian optimization calculation. For the cases with 0, 2, 4, and 6 explicit water systems. The input TSS either matches with or is positioned closely to the variational TSS. However, for the 8 explicit water system, the input TSS deviates from the variational TSS, which appears to be positioned earlier on the reaction coordinate. We aligned the geometries from these two TSS and calculated their RMSD values using an RMSD calculating tool from VMD.[14] The RMSD between these two TSSs is 0.037, indicating that they are geometrically different. This can explain the large amount (15.4%) of product recrossing for the 8 explicit water system shown in Table 7. We then performed downhill dynamic calculations on the identified VTST. Table 8 shows the dynamic results when the VTST is used as an input instead. As expected, the number of product recrossing trajectories decreased. The number of trajectories from **1** to **3** increased, however, the number of trajectories from **1** to **2** remain similar to the previous dynamic results.

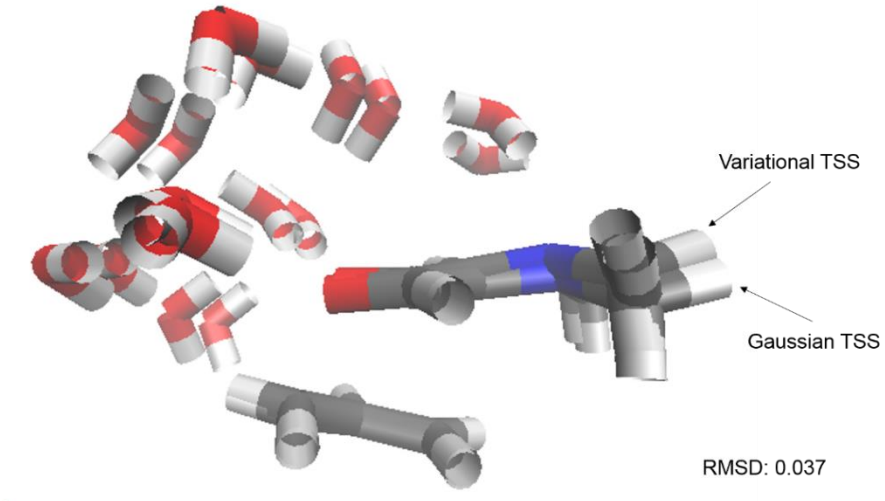


Figure 4. Superposition of Gaussian optimized **TSS1** and variational **TSS1** for the 8 explicit water system. RMSD (0.037) is calculated using VMD.

Table 8. Comparison of downhill dynamic results at SMD-(water)-B3LYP//6-31G(d) of Gaussian optimized **TSS1** and Variational **TSS1** (8 explicit water system).

	1 to 3	1 to 2	1 to 1	Other Recrossing
8 Explicit Water (Gaussian TSS)	39 (75.0%)	5 (9.6%)	0 (0%)	8 (15.4%)
8 Explicit Water (Variational TSS)	63 (91.3 %)	4 (5.8%)	0 (0%)	2 (2.9%)

Note: The categories for "Other Recrossing" refers to **3** to **3**, **3** to **1**, **2** to **3**, and **3** to **2**.

6.4 Conclusion

In this report, using a bifurcating model reaction, we have demonstrated the necessity for explicit solvent inclusion for certain types of reaction. Although implicit solvent models are universally used in the field of computational chemistry and their computational cost is relatively low, for reactions when atomistic solvent-solute interactions are crucial, explicit solvent is a better option. However, how many is enough? From our downhill dynamic simulation, we observed changes in product ratio when the number of explicit solvent molecules increases systematically. Therefore, the minimum number of explicit solvent molecules to include is not arbitrary. Ideally, we would keep increasing the number of explicit solvent molecules by increments and test if the dynamic results converge. That will likely be the future work for this project if it's continued. It is also worth noticing the difference when SMD was used instead of CPCM as the background implicit solvent model. In this report, SMD appears to be a better option. However, it is worth testing if the dynamic results change for CPCM-(water) if 10 or more explicit solvent molecules.

Reference

1) Tomasi, J., Mennuci, B., and Camm, R., *Chem. Rev.,* **2005**, *105*, 2999-3093

2) (a) Miertus, S., Scrocco, E., and Tomasi, J., *Chem. Phys.*,**1981**, *55*, 117-129 (b) Barone, V., Cammi, R., and Tomasi, J., *Chem. Phys. Lett.,* **1996**, 255, 327-335 (c) Cossi, M., Scalmani, G., Rega, N., and Barone, V., *J. Chem. Phys*., **2002**, *117* 43-54 (d) Baron, V., Improta, R., and Rega, N., *Thoer. Chem. Acc.*, **2004**, *111*, 237-245

3) (a) Klamt, A., and Schüürmaann, G., *J. Chem. Soc., Perkin Trans. 2*, **1993**, 799 (b) Andzelm, J., Kölmel, C., and Klamt, A., *J. Chem. Phys.*, **1995**, *103*, 9312-9320 (c) Barone, V., and Cossi, M., *J. Phys. Chem. A*, **1998**, *102*, 1995-2001 (d) Cossi, M., Rega, N., Scalmani, G., and Barone, V., *J. Comput. Chem.*, **2003**, *24*, 669-681 (e) Takano, Y., and Houk, K. N., *J. Chem. Theory. Comput.*, **2005**, *1*, 70-77

4) Marenich, A. V., Cramer, C. J., and Truhlar, D. G., *J. Phys. Chem. B*, **2009**, *113*, 6378-6396

5) Eilmes, A., *Theor. Chem. Acc.*, **2014**, *133*, 1538

6) Hare, S. R., Li, Ang., and Tantillo, D. J., *Chem. Sci.*, **2018**, *9*, 8937

7) Hare, S., and Tantillo, D. J., *Pure Appl. Chem.*, **2017**, *89*, 679-698

8) Gaussian 16, Revision C.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R.; Hasegawa, J.; Ishida, M., Nakajima, T., Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N., Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P., Burant, J. C.; Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M.; Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J., Gaussian, Inc., Wallingford CT, **2016**

9) (a) Becke, A. D., *J. Chem. Phys*., **1993**, *98*, 5648−5652 (b) Lee, C., Yang, W., Parr, R. G., *Phys. Rev. B: Condens. Matter Mater. Phys.*, **1988**, *37*, 785−789 (c) Mieh-lich, B.; Savin,

A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. Chem. Phys. Lett. 1989, 157, 200−206

10) (a) Gonzalez, C., and Schelgel, H. B., *J. Phys. Chem.*, **1990**, *94*, 5523-5527 (b) Fukui, K., *Acc. Chem. Res.*, **1981**, *14*, 363-368

11) Aziz, H. R., and Singleton, D. A., *J. Am. Chem. Soc.,* **2017**, *139*(16) 5965-5972

12) Zheng, J., Bao, J. L., Meana-Pañeda, R., Zhang, S., Lynch, B. L., Corchado, J. C., Chuang, Y. -Y., Fast, P. L., Hu, W. -P., Liu, Y. -P., Lynch, G. C., Nguyen, K. A., Jackels, C. F., Fernandez Ramos, A., Ellingson, B. A., Melissas, V. S., Villà, J., Rossi, I., Coitiño, E. L., Pu, J., Albu, T. V., Ratkiewicz, A., Steckler, A., Garrett, B. C., Isaacson, A. D., and Truhlar, D. G., Polyrate–version 2017, University of Minnesota, Minneapolis, **2018**

13) Garett, B. C., Redmon, M. J., and Steckler, R., Truhlar, D. G., Gordon, M. S., Baldridge, K. K., and Bartol, D., *J. Phys. Chem.*, **1987**, *92*, 1476

14) Humphrey, W., Dalke, A., and Schulten, K., *J. Mol. Graphics*, **1996**, *14*, 33-38