Computational Study of Mechanistic Pathways to Angiopterlactone Natural Products



Yining Lu
St Catherine's College
University of Oxford

A thesis submitted for the degree of
MSc by Research in Physical ad Theoretical Chemistry
Trinity 2024

Declaration

I, Yining Lu, declare that the Thesis I am submitting is entirely my own work except where indicated in the text, caption, footnote or bibliography. This Thesis has not been submitted in whole or in part for any other academic degree or professional qualification.

Yining Lu

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Fernanda Duarte, for her support throughout the course of this research.
Special thanks to my colleagues and fellow researchers at Oxford for their help throughout this journey. I am fortunate to have worked alongside such talented individuals.

Abstract

Computational chemistry has been widely used to study organic mechanisms and reactivities over the last few decades. More specifically, when there are complicated structures such as natural products, computational tools can be used to elucidate structures and therefore facilitate mechanistic studies. Building on previous works by the Lawrence group, who successfully synthesized (-)-angiopterlactone B and challenged the suggested biosynthesis proposal of (-)-angiopterlactone A as a biosynthetic precursor of (-)-angiopterlactone B, this work aims to further explore the underlying mechanism leading to the observed stereoselectivity. Based on our proposed mechanisms, we found that there are eight possible diastereoisomers of the final product. To determine the configuration of the final product and the most favorable pathway, a combination of DFT modeling, NMR calculations, and microkinetic models was applied to investigate this reaction.

First, NMR calculations confirmed the configuration of the final product obtained in the experiment, confirming that the Re-cis-syn-cis was the most likely obtained product.

Originally three different mechanistic pathways were proposed. However, the differences between the energy span and the activation barrier of the rate-determining step are too small for us to determine which one is the most favorable. Therefore, the microkinetic model was applied to compare the yield of products for the pathways, taking into account the concentration of different compounds in the solution. We conclude that the oxa-Michael/Michael/ring-contraction domino sequence is the most favorable mechanistic pathway to the Re-cis-syn-cis configuration of the product. Our results revealed that it is important to consider the concentration in a one-pot reaction where highly reversible reactions such as a Michael addition happen, and several different mechanistic pathways are competing with each other. We would expect that the approach used in this project can be extended to a wide range of biomimetic total synthesis that potentially involves complicated mechanisms, which have been mostly studied with DFT calculations before.

List of Abbreviations

Å . . . . . Ångstrom
ax . . . . . Axial
CC . . . . . Coupled cluster
CD . . . . . Circular dichroism
CI . . . . . Configuration interaction
CPCM . . . . . Conductor-like Polarizable Continuum Model
DCE 1,2-dichloroethane
δ . . . . . NMR shielding tensor
δE . . . . . Energetic span
DIA . . . . . Distortion / interaction analysis
DLPNO . . . . . Domain based local pair natural orbital
DMSO . . . . . Dimethylsulfoxide
DSO . . . . . Diamagnetic spin-orbit
E . . . . . Energy
ent . . . . . Enantiomer
eq . . . . . Equatorial
FC . . . . . Fermi contact
GGA . . . . . Generalised gradient approximation
GIAO . . . . . Gauge independent atomic orbital
G . . . . . Gibbs free energy
$\boldsymbol{H}$ . . . . . Hamiltonian operator
h . . . . . Planck's constant
HF . . . . . Hartree Fock
Hz . . . . . Hertz
INT . . . . . Intermediate
J . . . . . Coulping constant
k . . . . . Rate constant
K . . . . . Kelvin
k_B . . . . . Boltzmann's constant
kcal/mol . . . . . Kilocalories per mole
K_{eq} . . . . . Equilibrium constant
KS-DFT . . . . . Kohn Sham density functional theory
LSDA . . . . . Local spin density approximation
M . . . . . Michael addition
MAE . . . . . Mean absolute error
mol/L . . . . . Mole per liter
mol . . . . . Mole
MP2 . . . . . Second-order Møller-Plesset perturbation theory
NBO . . . . . Natural bond orbital
NMR . . . . . Nuclear magnetic resonance
NOESY . . . . . Nuclear Overhauser effect spectroscopy
OM . . . . . Oxa-Michael addition
PCM . . . . . Polarizable continuum mode

PES . . . . . Potential energy surface
Φ . . . . . Wavefunction
PSO . . . . . Paramagnetic spin-orbit
R . . . . . Gas constant
RC . . . . . Ring contraction
ρ . . . . . Electron density
RI . . . . . Resolution of the identity
RMSE . . . . . Root-mean-squared error
RRHO . . . . . Rigid rotor-harmonic oscillator
SCF . . . . . Self consistent field
SMD . . . . . Solvent Model based on Density
T . . . . . Temperature
TBAF . . . . . Tetra-n-butylammonium fluoride
TDI . . . . . TOF-determining intermediate
TDTS . . . . . TOF-determining transition state
THF . . . . . Tetrahydrofuran
TMS . . . . . Trimethylsilyl
TOF . . . . . Turnover frequency
TS . . . . . Transition state

Contents

1 Introduction **2**

1.1 Angiopterlactones A and B: total synthesis and biosynthetic speculations . . . 2

1.1.1 Isolation and the structures of angiopterlactones . . . 2

1.1.2 Biosynthetic speculations and total synthesis . . . 4

1.2 Contribution of this Thesis . . . 8

1.3 Computational structure optimisation: Wavefunction theory and Density functional theory . . . 9

1.3.1 Wavefunction theory . . . 9

1.3.2 Computational methods for retrieving dynamic correlation . . . 11

1.3.3 Density functional theory . . . 15

1.3.4 Basis set . . . 17

1.3.5 Solvent effects . . . 18

1.4 Computational structure elucidation: Nuclear Magnetic Resonance Spectroscopy (NMR) . . . 20

1.4.1 Application of computational NMR studies . . . 20

1.4.2 Chemical Shifts . . . 22

1.4.3 Coupling constants . . . 24

1.5 Methods . . . 26

1.5.1 Conformational sampling . . . 26

1.5.2 Energetic span approximation . . . 28

1.5.3 Microkinetic model . . . 29

2 Computational Details **31**

2.1 Optimisation and benchmarking methods . . . 31

2.2 NBO analysis . . . 33

2.3 NMR . . . 34

2.4 pK_a . . . 35

2.5 Microkinetic model . . . 36

3 Mechanistic studies **37**

3.1 PK_a . . . 37

3.2 Pathways and energy diagrams . . . 39

3.2.1 Mechanistic pathways . . . 39

3.2.2 Pathway A . . . 41

3.2.3 Pathway B . . . 44

3.2.4 Pathway C . . . 51

3.2.5 All 3 pathways . . . 57

3.3 Analysis on electronic structures and noncovalent interactions . . . 60

3.3.1 Comparison of structures and electronic energies of different conformations . . . 60

3.3.2 NBO analysis . . . 61

3.3.3 Distortion Interaction analysis (DIA) . . . 62

3.4 The microkinetic model . . . 64

4 Structure benchmarking and structure elucidation using NMR **67**

4.1 Benchmarking of geometries . . . 67

4.2 NMR . . . 74

5 Conclusions **76**

5.1 Computational results and analysis . . . 76

5.2 Outlook . . . 78

Appendices **79**

A Appendix: Absolute energies **79**

B Bibliography **83**

1 Introduction

1.1 Angiopterlactones A and B: total synthesis and biosynthetic speculations

1.1.1 Isolation and the structures of angiopterlactones

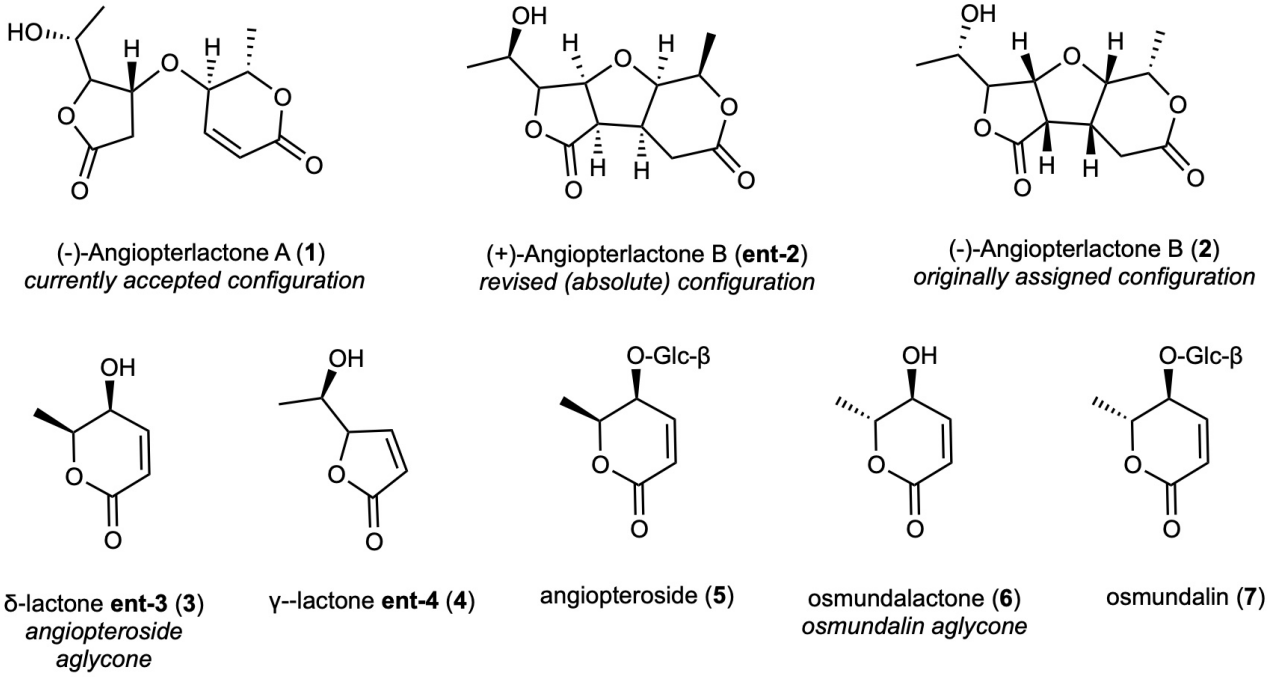


Figure 1 Lactones isolated from *Angiopteris* and *Osmunda* ferns.[1,2]

We are interested in specific types of stereoisomers of the angiopterlactone natural products, (-) angiopterlactone A (**1**) and (+)-angiopterlactone (**2**) (Figure 1) due to their bioactivities and the enantioselectivity in total synthesis.[1–4] These angiopterlactone natural products are rooted and isolated from *Angiopteris caudatiformis* Heiron, an ancient fern species from China, and their bioactivity and structural complexity of being a tricyclic ring are noteworthy.[3,4] For instance, the rhizome (a subterranean plant stem giving rise to shoots and roots) of this fern, named "ji ma" in Chinese, can be used as ingredients to treat tuberculosis, venomous snake bite, and bleeding wounds in folk medicine.[3] Since there are more than four stereocenters in these natural products, how to control the enantioselectivity of the total synthesis becomes an interesting problem for chemists to solve. Relevant lactones which are

structurally similar to fragments on the natural product are listed in the second row of Figure 1, with two stereocenters on each of them. Starting from these small molecules, biosynthetic speculations of the angiopterlactone natural products can be made. Therefore, our goal is to explore the mechanistic pathways of the synthesis of the angiopterlatone products in Figure 1.

Here are the our aims of the synthesis of these natural products and computational studies:

- Overcome the enantioselectivity challenges in the total synthesis of (-) angiopterlactone A (1) and (+)-angiopterlactone (2) by proposing possible mechanisms from structurally related lactone fragments such as ent-3 and ent-4, which contain multiple stereocenters.
- Investigate the biosynthetic origin of angiopterlactones and understand stereochemical development of the total synthesis, by exploring the mechanistic pathways.
- Determine the most favorable mechanistic pathway by applying computational tools such as DFT.

The synthesis of angiopterlactone A (**1**) and B (**2**) (Figure 1) has been of synthetic chemists' interest since it was first isolated by Zou *et al.* in 2009. Together with the other lactones, they are considered the most complex as a group of lactone metabolites isolated from *Angiopteris* and *Osmunda* ferns.[3,5] However, isolating the lactones appears to be challenging. Zou *et al.* isolated from 10 kg of *Angiopteris caudatiformis* 230 mg of (-)-angiopterlactone A (**1**), but only 20 mg of (+)/(-)-angiopterlactone B (**2**)/(**ent-2**) were isolated.

Zou *et al.* utilized 1D and 2D NMR spectroscopy to determine the geometry connectivity of angiopterlatone A (**1**) and angiopterlacone B (**2/ent-2**). To determine the relative configuration of the angiopterlactones (**1, 2**), nuclear Overhauser effect spectroscopy (NOESY) experiments were conducted (Figure 2) In **1**, a NOESY correlation between H3' and H5 was assigned by the authors as a cis relationship between them.[3] The ether linkage between H3' and H5 is rotatable, and since the

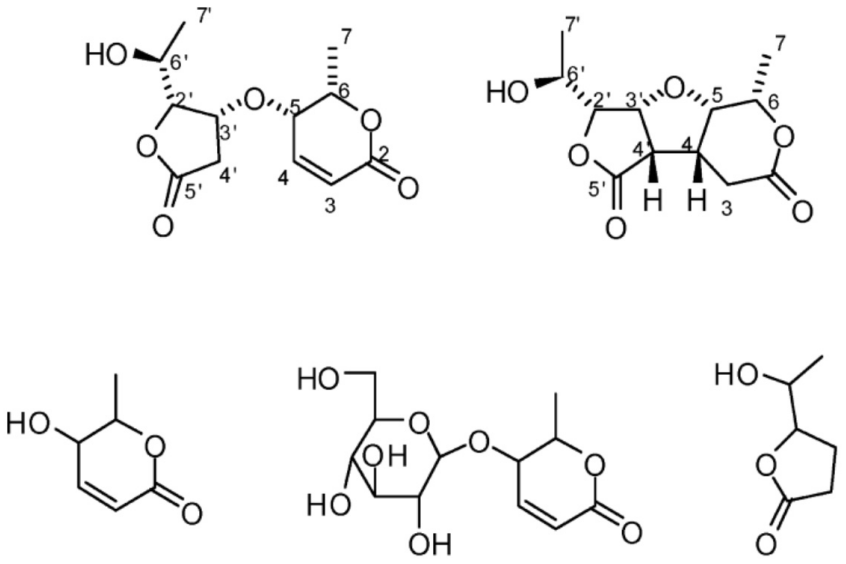


Figure 2 The angiopterlactones **(1, 2)** and co-isolated compounds (adpated from *J. Nat. Prod.* 2009, **72**, 921–924)

NOESY correlation method is not completely reliable across two rotatable bonds, it is possible that the relative configuration assigned here might not be correct.[2,6] Zou *et al.* then went on to assign the other protons of (-)-angiopterlactone (**1**) using the NOESY method: there is a cis relationship between H2' and H3', as well as between H5 and H6. To complete the absolute configuration of **1**, they determined the absolute stereochemistry of C6, C6' and C3': using the circular dichroism (CD) excitation chirality method, C6 was assigned as S, and C3' was assigned as R.[7] For C6', the modified Mosher's method was utilized to determine its stereochemistry, and result was S.[3]

However, the stereochemistry of the other co-isolated compounds (Figure 2) was not provided in the original paper, thus making it hard to tell the structural connections between the isolated compounds. The biosynthetic origin of the angiopterlactones was yet not clear, given the fact that various stereoisomers of **3** have been reported in previous literature.[5,8–10]

1.1.2 Biosynthetic speculations and total synthesis

From the isolated structures (Figure 2), it was speculated that angiopterlactone A (**1**) might be a biosynthetic precursor of angiopterlactone B (**2**), with **1** undergoing

an intramolecular Michael addition to form **2**.[3] Considering the other co-isolated structures with **1** and **2**, we think it is likely that the six-member ring (**3**) and the five-member (**4**) ring on the angiopterlactone (**1, 2**) are relevant to the co-isolated lactones. The speculation here is that the five-member and six-member lactones undergo an intramolecular ox-Michael addition to form **1**. For the reaction mechanism, we also propose that **3** undergoes a ring contraction to form **4**, with evidence in previous literature showing that similar lactones to **3** undergo a ring contraction to form a γ-lactone.

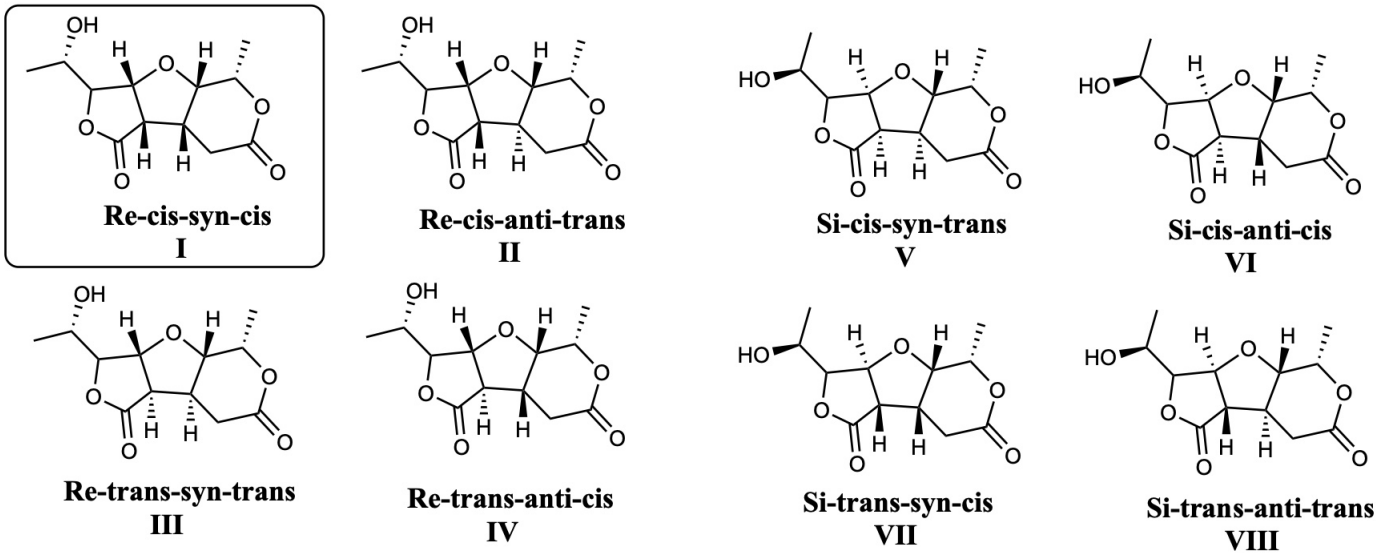


Figure 3 Eight possible diastereoisomers of angiopterlactone B , with Re/Si representing face attack in terms of prochirality. The circled compound **I** is (-)-angiopterlactone B (**2**).

Therefore, a biomimetic domino ring-contraction/oxa-Michael/Michael dimerization sequence was proposed. In the Michael addition, three new contiguous stereogenic centers are formed in a single step.[1] There are eight possible diastereoisomers of different configurations for angiopterlactone B, based on our biosynthetic speculation (Figure 3,4).

The first total synthesis of **2** was reported by Lawrence *et al.*, in 2017 (Figure 5).[1] A mechanism involving dimerization in an oxa-Michael addition for the synthesis was proposed (Figure 4), assuming **1** is the precursor of **2**. This biomimetic synthesis involves a K_2CO_3-mediated dimerisation of δ-lactone **ent-3**. The total synthesis

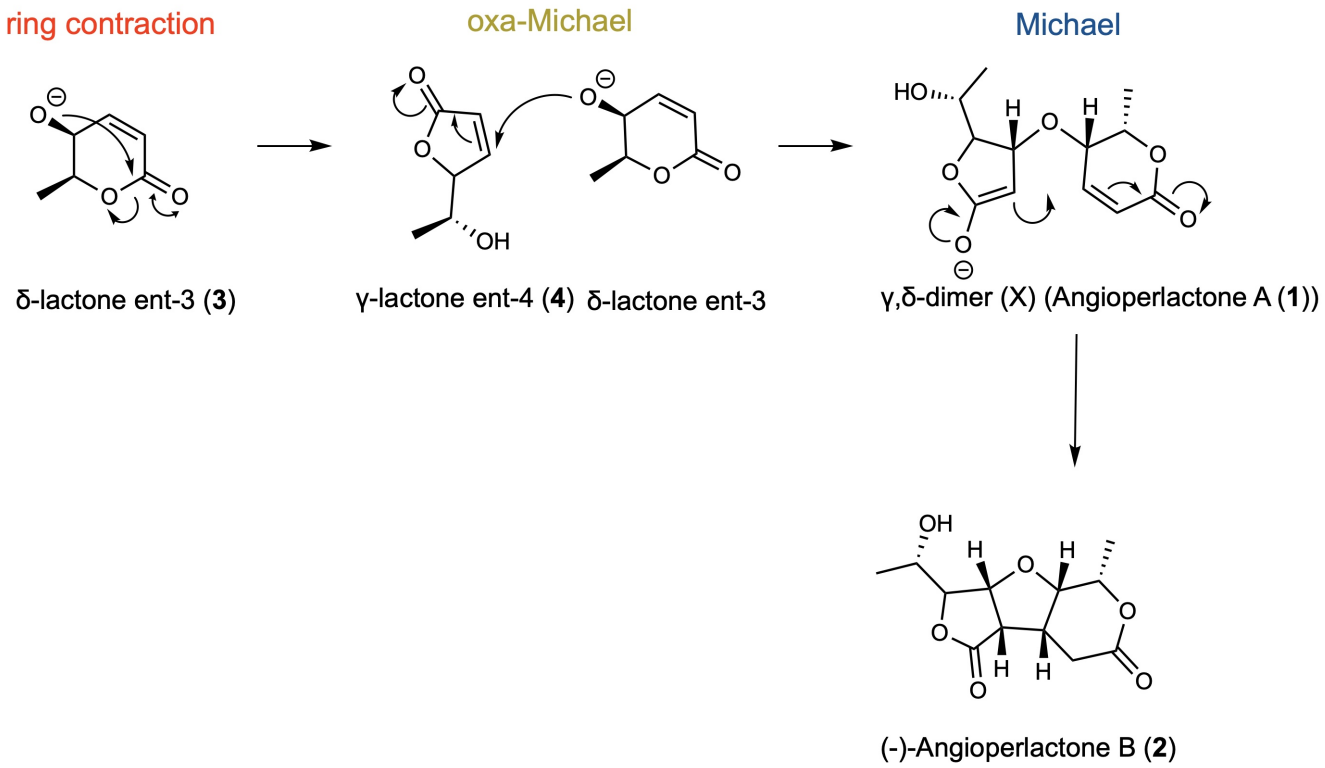
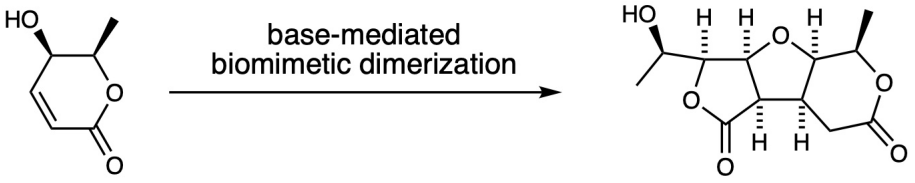


Figure 4 Proposed mechanism, assuming **1** is the precursor of **2**.

is 4-step and enantioselective (7-14% yield, 850 mg scale). In the same year, Bhattacharya *et al.* also reported their synthesis of two diastereoisomers of both the natural and unnatural enantiomers of angiopterlactone B, (+)-angiopterlactone B ([chiral pool]), (-)-angiopterlactone B (**2**) (compound **I**).[11] They started from two different carbohydrate-derived starting materials. The yield of **2** was 12% (in 8 steps), and that of **ent-2** was 4.6% (in 10 steps). By using tetra-n-butylammonium fluoride (TBAF) as a base in tetrahydrofuran (THF), the same type of dimerization as Thompson *et al.* was achieved.

In 2018, Meshram *et al.* obtained **3** from L-lactate in 11 steps (6.8% yield, 76 mg scale), completing a total synthesis of **2** in 12 steps.[12] In 2023, Deska *et al.* applied the enantioselective 3-step approach of **3** from 2-acetylfuran from Lawrence *et al.*, and then used a one-pot multi-enzyme-mediated synthesis of **3** from 2-acetylfuran (49-53% yield, 65 mg scale).[13] Aqueous reaction conditions were then used to access the same type of K_2CO_3-mediated dimerization of **3** to synthesize **2**. This is the shortest reported route of synthesis of **2** so far (2 steps, enantioselective, 11%

yield, 14 mg scale), and only (-)-angiopterlactone **2** was observed. It is noticeable that any case of successful enantioselective total synthesis of **ent-2** has not been reported yet.



3 or **ent-3**

ent-2 or **2**
(+) or (–)-angiopterlactone B

	Conditions	Product	Yield	Scale
Lawrence (2017)	K_2CO_3 (0.2 equiv.), CH_2Cl_2 70 °C, 16 h	**2**	25–35%	850 mg
Bhattacharya (2017)	TBAF (2 equiv.), THF, rt, 2 h	**ent-2** **2**	62% 60%	18 mg 25 mg
Deska (2023)	K_2CO_3 (0.2 equiv.), H_2O, 70 °C, 16 h *then* aq. HCl, 70 °C, 1 h	**2**	20%	14 mg

Figure 5 Summary of previous synthesis of **2**/**ent-2**.

In previous works, the angiopterlactone A (**1**) was not isolated from the final product of the total synthesis.[1,12,13]. Therefore, though the final product (+)/(-)angiopterlactone (**ent-2**/**2**) has been synthesized before, we are still not certain about the synthetic route of **1**. To see if **1** can be formed through the oxa-Michael addition as we proposed, our collaborators mixed **3** and **4** together under the reaction conditions, but the oxa-Michael addition did not happen, so there was no dimerization going on. It is worth checking if the activation barrier of this reaction is too high so that the reaction does not take place, or if there are other factors we have not taken into account in our proposed pathways.

Another point is the enantioselectivity of the reaction. Although these cases reported above all succeeded in synthesizing **2**/**ent-2**, only the (-)-angiopterlactone **2** is preferred. Our collaborator attempted a cross dimerization using the enantiomer of **3** to see if **ent-2** can be synthesized, but failed to do so. Therefore, the enantioselectivity of the mechanistic pathways is also worth checking.

1.2 Contribution of this Thesis

The history and recent progress of the isolation and total synthesis of angiopterlactone natural products have been briefly introduced. To investigate the mechanistic pathways using computational chemistry, the theories of DFT calculations, and NMR (nuclear magnetic resonance) calculations are introduced and discussed later in this chapter. Then a more detailed explanation of specific methods used is included, such as conformational sampling, the energetic span approximation, and the microkinetic model.

Chapter 2 covers the computational details of the methods used in this thesis. In Chapter 3, the benchmarking of geometries is presented first, and then the results of NMR calculations are discussed to confirm the configuration of the final product of the total synthesis.

Chapter 4 answers the question of which pathway is the most favorable by using PK_a calculations to clarify the nucleophile and electrophile in the proposed mechanisms first, and the energy profile for each pathway is computed with DFT. However, we cannot answer our question by analyzing DFT results, so the microkinetic model is applied. Through simulations of different pathways using the microkinetic model, we conclude that pathway **C** is the most favorable pathway.

In summary, this work proposes a possible solution, the microkinetic model, to answer mechanistic questions when DFT calculations fail to do so. We hope that this approach can be further applied in the investigation of complicated mechanistic pathways of natural products in the future, offering a fuller understanding of the reaction of interest.

1.3 Computational structure optimisation: Wavefunction theory and Density functional theory

1.3.1 Wavefunction theory

The computational methods widely used in computational chemistry to model structures, such as DFT, are based on the Hohenberg-Kohn theorem to solve the Schrödinger equation.[14,15] The time-independent Schrödinger equation can be written as:

$$\boldsymbol{H}\psi = E\psi \tag{1}$$

where $\boldsymbol{H}$ is the Hamiltonian operator acting on the wavefunction ψ of a certain system, and E is the total energy of the system.

Based on the Schrödinger equation, the Born-Oppenheimer approximation is made, where the nuclei are considered static. The nuclear kinetic energy is therefore zero, and the internuclear repulsion is constant, as the nuclei are moving much more slowly than the electrons due to their mass. The Born-Oppenheimer Hamiltonian is expressed as (with i, j representing the electron indices, and A, B representing the nuclear indices):

$$\boldsymbol{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i{}^2 + \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{N}\frac{Z_A Z_B}{r_{AB}} \tag{2}$$

where the first term is the expression for the kinetic energy, the second term is the attraction between the electrons and the nuclei, the third term is the repulsion between the electrons, and the last term is the repulsion between nuclei. Only the system with a single electron can be solved exactly using this equation, but the Hartree-Fock (HF) approximation can be applied to solve this problem computationally.[16]

In the HF approximation, the many-electron wavefunction is transformed into a one-electron problem by using the Slater determinant to describe the spin orbitals:

$$
\begin{aligned}
|\Phi\rangle &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \dots & \chi_{\mathrm{M}}(\mathbf{x}_1) \\ \dots & \dots & \dots \\ \chi_1(\mathbf{x}_{\mathrm{N}}) & \dots & \chi_{\mathrm{M}}(\mathbf{x}_{\mathrm{N}}) \end{vmatrix} \\
&= \frac{1}{\sqrt{N!}} \sum_{i=1}^{N!} (-1)^{p_i} \boldsymbol{P}_i (\chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2)\dots\chi_{\mathrm{M}}(\mathbf{x}_{\mathrm{N}}))
\end{aligned}
\tag{3}
$$

where $\chi_1(\mathbf{x}_1)$ means that the electron numbered as 1 occupies the orbital numbered as 1. $\boldsymbol{P}_i$ is the permutation operator that switches the rows of the determinant. For a given permutation operation $\boldsymbol{P}_i$, p_i times of swaps of rows have to be done.

The eigenvalue ε_{i} of the ith orbital χ_i is its energy, and each electron is affected by the repulsion from the mean field of the remaining electrons, υ_{HF} (b indicating a two-electron operator):

$$
\begin{aligned}
\left[-\frac{1}{2} \sum_{i=1}^{N} {\nabla_1}^2 + \sum_{A=1}^{M} \frac{Z_A}{r_{A1}} + \upsilon_{HF} \right] \chi_i &= \varepsilon_i \chi_i \\
\upsilon_{HF} &= \sum_{b}^{N} (\boldsymbol{J_b} - \boldsymbol{K_b})
\end{aligned}
\tag{4}
$$

where $\boldsymbol{J_b}$ is the Coulomb operator, giving the Coulombic repulsion of an electron, and $\boldsymbol{K_b}$ is the exchange operator, the non-classical term for the energy of swapping the electrons of the same spin for different orbitals, due to the indistinguishability of electrons. Following this, the total energy can be written as

$$
E_{HF} = \sum_{i}^{N} \varepsilon_i - \frac{1}{2} \upsilon_{HF} \tag{5}
$$

The total energy is equal to the sum of all the orbital energies subtracting half of

v_{HF}, since the repulsion between pairs of electrons is counted twice.

The HF approximation is limited in the sense that it does not count for the electron correlation, the interaction between electrons, due to which the position of one electron affects another. This is because the repulsion between the electrons is calculated based on the average position of other electrons in the system, instead of the exact position of every electron. Therefore, this repulsion is overestimated. Part of the electron correlation is included in the exchange term in the HF approximation, which is the same-spin (Fermi) correlation, and the opposite-spin (Coulomb) correlation is not included. The HF approximation is thus an upper bound of the exact energy.

1.3.2 Computational methods for retrieving dynamic correlation

As stated above, HF approximation does not include an accurate representation of electron correlation. There are various approaches that introduce electron correlation into the HF wavefunction. For instance, a common way is to incorporate excited configurations. In general, there are two possible approaches: a) using a single Slater determinant, varying the weights of the excited configurations to correct the total energy and recover dynamic correlation (the misattributed repulsion mentioned above) b) using multiple Slater determinants, varying both the weights of the configurations and the orbitals to take into account the so-called static correlation, which comes from the fact that a single determinant sometimes cannot represent the electronic structure properly (for instance, systems with degeneracies).

For our system, we only need correction of dynamic correlation. The most common approaches used are the second-order Møller-Plesset perturbation theory (MP2) and the coupled cluster method with single, double, and perturbative triple excitations [CCSD(T)].[17,18]

For most closed-shell systems of small to medium-sized molecules, such as most organic molecules, the MP2 method is usually effective and performs well.[19] While working well for systems with weak correlations or weak interactions, the MP2 method tends to overestimate dispersion interaction and thus cannot describe the systems with strong correlations or multireference character very well.[20,21]

The perturbation theory assumes that a new problem to be solved differs slightly from a problem that has been solved before so that we can add perturbation terms to solve the new one. For Schrödinger equation, we can define a Hamiltonian operator with a reference ($\boldsymbol{H_0}$) and a perturbation ($\boldsymbol{H'}$), so that it can be solved:

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{H_0} + \lambda \boldsymbol{H'} \\ \boldsymbol{H_0}\boldsymbol{\Phi}_i &= E_i \boldsymbol{\Phi}_i, \qquad i = 0, 1, 2, \cdots, \infty \end{aligned} \tag{6}$$

where λ is the parameter for controlling the strength of perturbation. In other words, the reference is a reasonable solution (a wavefunction) to the Schrödinger equation. This might not be true for some open-shell systems, leading to the method's major limitation.[19]

Therefore, the total energy of the MP2 method can be expressed as:

$$E = E_0 + E_1 + E_2 \tag{7}$$

The zeroth-order wavefunction (W_0) is the same as the one for HF approximation:

$$W_0 = \langle \Phi_0 | \mathbf{H_0} | \Phi_0 \rangle = \left\langle \Phi_0 \middle| \sum_{i=1}^{N_{\text{elec}}} \mathbf{F}_i \middle| \Phi_0 \right\rangle = \sum_{i=1}^{N_{\text{elec}}} \varepsilon_i \tag{8}$$

And the first-order wavefunction takes the average over the perturbation operator the zeroth-order one, and the perturbation term is the exact $\boldsymbol{V_{ee}}$ (the electronic repulsion operator) minus twice the $\langle \mathbf{V}_{\text{ee}} \rangle$ operator.

$$W_1 = \langle \Phi_0 | \mathbf{H'} | \Phi_0 \rangle = \langle \mathbf{V}_{\text{ee}} \rangle - 2\langle \mathbf{V}_{\text{ee}} \rangle = -\langle \mathbf{V}_{\text{ee}} \rangle \tag{9}$$

Therefore, if MPn is the notation of the total energy for the nth order correction, the nth order perturbation $E(MPn)$ can be written as

$$\begin{aligned} \text{MP0} &= E(MP0) = \sum_{\text{i=1}}^{N_{\text{elec}}} \varepsilon_i \\ \text{MP1} &= E(MP1) = E(MP0) + E(MP1) = E(HF) \end{aligned} \tag{10}$$

The total energy for second-order perturbation can still be obtained using the same way as in equation 6 by promoting two electrons from two occupied orbitals (i and j) to two virtual orbitals (a and b), in terms of Slater determinants. After simplification, it can be written as:

$$W_2 = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} \frac{\langle \Phi_0 | \mathbf{H'} | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \mathbf{H'} | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}} \tag{11}$$

The second-order correction is the difference between the total energy of the two Slater determinants, expressed in terms of two-electron integral:

$$E(MP2) = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} \frac{\langle \phi_i \phi_j | \mathbf{H'} | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \mathbf{H'} | \phi_b \phi_a \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{12}$$

The MP2 method usually recovers 80-90% of the correlation energy, acting as an efficient method including electron correlation.[16] Since the MP2 method includes the HF wavefunction and can be viewed as an improved version of HF approximation, it describes the same-spin (SS) and opposite-spin (OS) with different accuracy. The SS and OS components can be corrected empirically by scaling the MP2 correlation energy, thus giving rise to the *Spin Component Scaled* MP2 (SCS-MP2), which is employed as the benchmarking method in this work.[22]

The CCSD(T) method, a truncated coupled cluster method, on the other hand, includes a given type of correction to an infinite order. The excitation operator $\boldsymbol{T}$ can be expressed as the sum of $\boldsymbol{T}_i$ operators, where i stands for the ith excited state.

$$\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \cdots + \boldsymbol{T}_{N_{\text{elec}}} \tag{13}$$

Then the HF reference wavefunction Φ_0 can be expanded to express the Slater determinants of all the excited states:

$$\begin{aligned} \boldsymbol{T}_1\Phi_0 &= \sum_{i}^{\text{occ}} \sum_{a}^{\text{virt}} t_i^a \Phi_i^a \\ \boldsymbol{T}_2\Phi_0 &= \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} t_{ij}^{ab} \Phi_{ij}^{ab} \end{aligned} \tag{14}$$

In terms of Taylor expansion, the exponential operator is expanded as:

$$e^{\boldsymbol{T}} = \sum_{n=0}^{\infty} \frac{\boldsymbol{T}^n}{n!} = 1 + \frac{\boldsymbol{T}}{1!} + \frac{\boldsymbol{T}^2}{2!} + \frac{\boldsymbol{T}^3}{3!} + \cdots \tag{15}$$

We can draw an analogy between the configuration interaction (CI) method and the expansion of the coupled cluster method. The CI method is also a post-Hartree method to solve the Schrödinger equation based on the variational principle. The energy is minimized from the trial wavefunction, a linear combination of the excited Slater determinants. The configuration interaction wavefunction can be expressed by using the excitation operator:

$$\Psi_{CI} = (1 + \boldsymbol{T})\Phi_0 = (1 + \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \cdots)\Phi_0 \tag{16}$$

And the coupled cluster wavefunction can be written as:

$$\Psi_{CC} = e^{\boldsymbol{T}}\Phi_0 \tag{17}$$

Using Equation 16 and 17 above, we can rewrite the Taylor expansion as:

$$e^{\boldsymbol{T}} = 1 + \boldsymbol{T}_1 + \left(\boldsymbol{T}_2 + \frac{\boldsymbol{T}_1{}^2}{2!}\right) + \left(\boldsymbol{T}_3 + \boldsymbol{T}_2\boldsymbol{T}_1 + \frac{\boldsymbol{T}_1{}^3}{3!}\right) + \cdots \tag{18}$$

The Schrödinger equation then becomes:

$$\boldsymbol{H}\mathrm{e}^{\boldsymbol{T}}\boldsymbol{\Phi}_0 = E_{CC}\,\mathrm{e}^{\boldsymbol{T}}\boldsymbol{\Phi}_0 \tag{19}$$

CCSD(T) is a truncated form of the couple cluster method. The approximation of $\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2$ gives the CCSD method:

$$\mathrm{e}^{\boldsymbol{T}_1+\boldsymbol{T}_2} = 1 + \boldsymbol{T}_1 + \left(\boldsymbol{T}_2 + \frac{\boldsymbol{T}_1{}^2}{2!}\right) + \left(\boldsymbol{T}_2\boldsymbol{T}_1 + \frac{\boldsymbol{T}_1{}^3}{3!}\right) + \cdots \tag{20}$$

The triple contribution is given by the Møller-Plesset perturbation theory, resulting in the hybrid method of CCSD(T).[18] A linearly scaled version of this method, the domain-based local pair-neutral orbital (DLPNO) CCSD(T) is applied in this work to calculate the final single point energies.[23] The CCSD(T) method is often described as the "gold standard" method when experimental data is not available, and only possible for small systems with 10s of atoms.[24]

1.3.3 Density functional theory

Another approach to solving the problem of the wavefunction theory discussed above is the density functional theory (DFT). The basis of the DFT was proved by Hohenberg and Kohn, which states that the ground state electronic energy is completely based on the electron density ρ, so there is a one-to-one mapping between the electronic density and the energy of the system, with a unique functional of electron density:[25]

$$E[\rho] = \int \upsilon(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + F[\rho] \tag{21}$$

where the first term describes the attraction between the electrons due to the electron density, and the second term contributes to the electronic kinetic energy and electron-electron repulsion. The orbitals are introduced here to compute the kinetic energy.

E.B. Wilson stated the following arguments to prove this[26]:

- The number of electrons is defined by the integral of the density of electrons.
- The position of the nuclei is defined by the cusps in the density function.
- The corresponding nuclear charges are defined by the heights of the cusps.

The general form of the Hamiltonian operator can be written as:

$$\boldsymbol{H} = \boldsymbol{H}_{\mathbf{K}} + \boldsymbol{H}_{\mathbf{V}} + \boldsymbol{H}_{\mathbf{J}} + \boldsymbol{H}_{\mathbf{XC}} \tag{22}$$

where $\boldsymbol{H}_{\mathbf{K}}$ represents the kinetic term, $\boldsymbol{H}_{\mathbf{V}}$ the electron nuclear interaction term, $\boldsymbol{H}_{\mathbf{J}}$ the Coloumb term (repulsion), $\boldsymbol{H}_{\mathbf{XC}}$ the exchange-correlation term. While the form of the first three terms is known, the true form of the exchange-correlation term is not known.[15]

Kohn-Sham (KS) DFT is one of the most commonly used approaches in modern density functional theory.[27] For the two terms expressing total energy in the DFT Equation 21, a) the electrons are assumed as non-interacting in the first term of attraction, b) the electronic interaction in the second term is to be approximated.[28] In KS-DFT, the total energy can be written as a functional of the electron density:

$$E[\rho] = T_{\mathrm{s}}[\rho] + \int \upsilon(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + E_{\mathrm{J}}[\rho] + E_{\mathrm{XC}}[\rho] \tag{23}$$

where the first term corresponds to the first term that expresses the kinetic energy in Equation 22, and the other three terms are also defined as those in Equation 22. The third term, $E_{\mathrm{J}}[\rho]$ is a term for the Coulomb functional to add to the kinetic term to obtain the inter-electron repulsion, and the fourth term $E_{\mathrm{XC}}[\rho]$ is an exchange-correlation functional to describe correlation interactions.

Several approaches using a variety of different functionals have been developed to solve this problem.[27] According to the hierarchy of functions proposed by Perdew,[29]

going from lower accuracy to higher accuracy, there is LSDA (local spin density approximation, where the density of uniform electron gas is used), GGA (generalized gradient approximation, where both the local electron density and the local gradient are used), meta-GGA (the density, gradient, and the second-order derivative are used). By including a fraction of Hartree-Fock (HF) exchange energy in the functional, the hybrid DFT functionals provide error cancellations resulting from the self-interaction error (SIE) of electrons. To further increase the accuracy, double hybrid DFT functionals include expressions of virtual orbitals in the exchange-correlation functional, in a similar manner to second-order Møller-Plesset perturbation theory (MP2) correction. In general, the accuracy of DFT calculations improves as the level of theory goes up (Figure 6), but since Kohn-Sham (KS) DFT is variational, this is not necessarily true all the time.

1.3.4 Basis set

To optimize geometries using DFT methods, a finite basis set, which is a combination of orbital functions, is required to describe the system of interest. There are two types of basis sets: Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs). The STOs have an exponential dependence the same as the exact orbitals of hydrogen-like atoms, which decay as exp(-αr). However, the integration of STOs might be difficult as the three-centre and four-centre integrals are computationally intensive, and thus cannot be performed analytically.

GTOs, on the other hand, decay in the form of exp(-αr^2). For GTOs, the higher-order orbital functions can be transformed into lower-order orbital functions through the transformation of coordinates, especially in quantum chemistry programs.[16] Although GTOs do not contain cusps and fall off much more quickly than STOs from the nucleus, requiring more GTOs to form the basis set, the computational efficiency compensates for that.[16] The Gaussian-type 'def2' basis sets, as a part of the "Ahlrichs' family", are used for our optimizations and single-point calculations in this work.[31] The 'def2' basis sets are split-valence, which means that the valence orbitals are represented with more than one basis function, improving their flexibility.

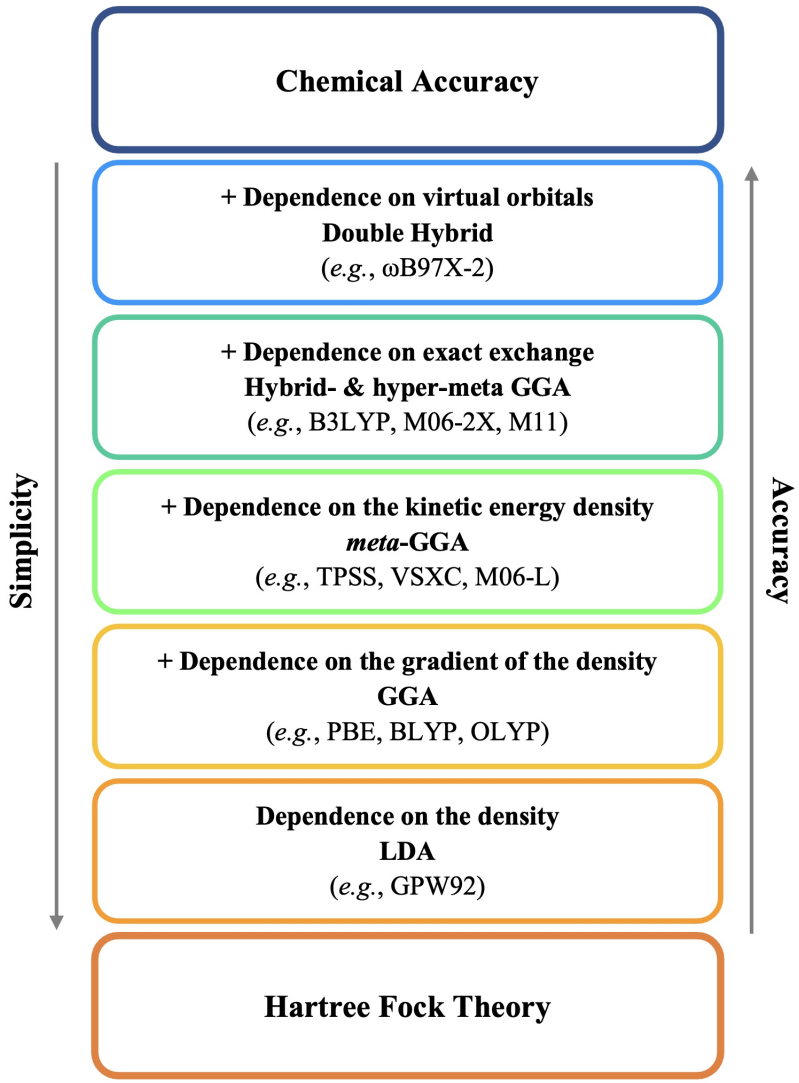


Figure 6 Jacoubs ladder of functionals, accuracy of level of theory vs. simplicity (adapted from *Chem. Soc. Rev.* 2016, **45**, 6093-6107).[30]

1.3.5 Solvent effects

In computational chemistry, there are two major types of solvation models: explicit solvation and implicit solvation. In the explicit solvation model, the solvent molecules are placed as discrete solvent molecules into the solute. The explicit solvation model is capable of describing interactions such as hydrogen bonding between two molecules explicitly. This type of solvent can describe multiple configurations of solvent molecules, taking temperature into account. While potentially leading to an accurate representation of the system of interest, the explicit solvation might also result in an excessively large number of configurations due to a lack of strong

interactions, or errors in entropic calculations. Ho *et al.* proved that the explicit solvent might perform worst, for instance, in the case of a simple Menschutkin reaction (NH_3 + CH_3Cl). It underestimated the reaction barrier by at least 5.98 kcal/mol, worse than the implicit solvent.[32]

The implicit solvent, on the other hand, represents the solvent as a continuum medium, instead of explicitly, with discrete solvent molecules. It speeds up the computations more than the explicit solvent. In DFT calculations, the polarizable continuum model (PCM) is often applied, with a dielectric medium considering the charges at the surface.[33] Here we use the SMD solvation model, which takes the effects of the electrostatic interaction using the Conductor-like Polarizable Continuum Model (CPCM) into account and includes a self-consistent reaction field treatment that facilitates DFT calculations and considers the solvent as a continuum medium with surface tension.[34]

1.4 Computational structure elucidation: Nuclear Magnetic Resonance Spectroscopy (NMR)

1.4.1 Application of computational NMR studies

Nuclear Magnetic Resonance Spectroscopy (NMR) is an experimentally useful tool for chemists to elucidate structures of molecules, especially when they are complicated and have several diastereoisomers.[35] The NMR chemical shifts and coupling constants obtained through experiments are often compared with the computed ones to confirm the configuration of the final product. To achieve this, we use DFT methods to perform NMR calculations computationally.

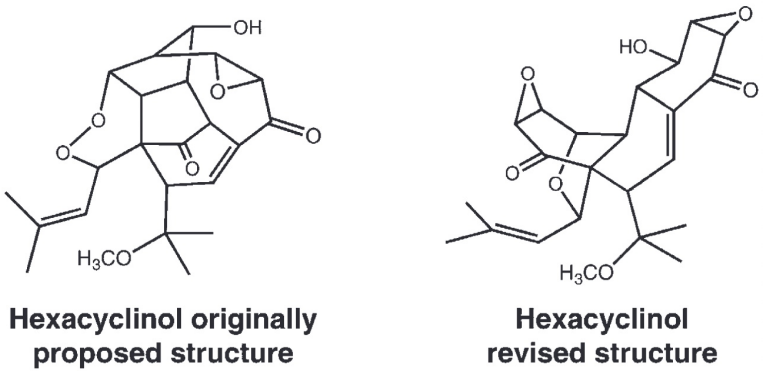


Figure 7 Originally assigned and revised structures of hexacyclinol (adapted from *Chemical Reviews*, 2012, **112**, 1839–1862).

Computational NMR predictions have been widely used to help elucidate the configurations of molecules of complicated structures.[35] There are many examples of revising the assigned structure based on the computed NMR chemical shifts, and a remarkable one is the case of hexacyclinol (Figure 7).[35–38] Following the isolation and total syntheses of this natural product in 2002, there arose this so-called "hexacyclinol disputed" regarding the computed NMR data of this compound.[36,39,40] Rychnovsky proposed a revised structure of hexacyclinol based on the computed ^{13}C chemical shifts, which was soon confirmed by total synthesis and crystallography.[40,41] Given this conclusion, Siaelli and Bagno then raised the question of whether two different structures could have the same NMR spectra, and recomputed the ^{1}H and ^{13}C of the originally assigned structure and the revised structure, concluding that the revised structure was correct.[36]

There are also other examples of assigning structures based on computed NMR chemical shifts or coupling constants, including complex structures of natural products.[31,42,43] When experimental NMR spectra do not match the previously reported ones in the literature, NMR calculations can be used to check the reason for that based on biosynthetic speculations. In 2018, Tantillo *et al.* applied computational NMR spectroscopy to revise the assigned structure of dichrocephone A obtained from a total synthesis.[44] After the first assigning of the structure of dichrocephone A, it was realized that the ^{1}H and ^{13}C data did not match the computed ones, and it was deduced that dichrocephone A might be a diastereoisomer of the proposed structure. They went on to suggest that dichrocephone A is the precursor to dichrocephone B (Figure 8), and the originally assigned structure (**1**) and the revised structure (8-epi-**1**) are the only possible diastereoisomers conducive to the proposed mechanistic pathway. The possibilities were thus narrowed down for synthesis.

More specifically, the computed NMR chemical shifts can be applied to determine the structure of a diastereoisomer conducive to its mechanistic pathway, through which biosynthetic speculation can be addressed and utilized for synthesis.[44] Smith and Goodman proposed the CP3 parameter in 2009, which assigns two different structures to two sets of experimental data based on differences in computed chemical shifts and differences in experimental chemical shifts by applying Bayes' theorem. This approach can be improved because two sets of experimental data and two or more candidate structures are needed to assign possibilities.[45]

In 2010, Smith and Goodman published an improved method of assigning probabilities, DP4, whose advantage is that it requires one set of experimental data, which can be assigned to two or more candidate structures. Based on the assumption that calculated error probabilities follow a statistical Student's t-distribution, this method is then used to convert the product of the individual error probabilities into the overall probability of a certain structure.[46]

In the work of Tantillo *et al.* on dichrocephone mentioned above,[44] after realizing that only two diastereoisomers had to be considered, the DP4 analysis was applied to support their biosynthetic speculation, which resulted in a 99.8% probability that

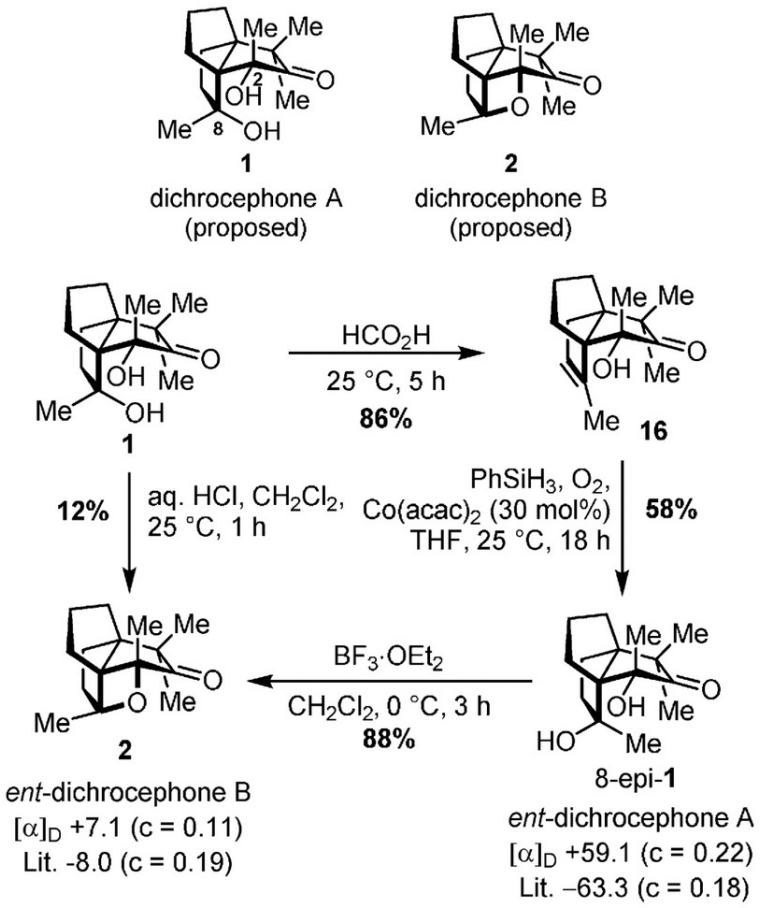


Figure 8 Originally assigned and revised structures of dichrocephone A and B, and the synthesis of the dichrocephone natural products (adapted from *Angewandte Chemie International Edition*, 2012, **57**, 2419–2422).

the revised structure 8-epi-**1** is the correct structure of dichrocephone A.

Therefore, computational NMR tools are a powerful tool to study the total synthesis of natural products and mechanistic pathways, as computed NMR shifts can be compared to experimental values to assign configurations of products.

1.4.2 Chemical Shifts

The NMR shielding tensor (δ) is calculated as the second derivative of the total energy (E) with respect to the external magnetic field (B) and the magnetic moment of the nucleus of interest (μ):

$$\sigma_{\alpha\beta} = \frac{\partial^2 \mathrm{E}}{\partial \mu_\alpha \partial \mathbf{B}_\beta} \tag{24}$$

where the external magnetic field responds to the shielding of electron density around the nucleus. In DFT calculations, the so-called "gauge" problem occurs, as a result of using finite basis sets in expansion of wave function or electron density.[47] The gauge-dependent terms (paramagnetic and diamagnetic) cannot be canceled out completely through the approximation in finite basis sets. For a gauge transformation of the vector potential A, where:

$$\mathbf{A}(\mathbf{r}) \rightarrow \mathbf{A}(\mathbf{r}') = \mathbf{A}(\mathbf{r}) + \Delta\lambda(\mathbf{r}) \tag{25}$$

the exact electronic wave function is

$$\boldsymbol{\psi}(\mathbf{A}') = \mathrm{e}^{-\mathrm{i}\Delta}\boldsymbol{\psi}(\mathbf{A}) \tag{26}$$

with Δ relevant to $\lambda(\mathrm{r_i})$, a continuous function of the electron i's positions

$$\Delta = c^{-1} \sum_i \lambda(\mathbf{r}_i) \tag{27}$$

While complete variational basis sets are gauge-invariant and satisfy Equation 26, the approximate wave functions need to be forced into a gauge-invariant form, the so-called GIAOs:

$$\chi_p(\mathbf{r}, \mathbf{A}) = \chi_p(\mathbf{r}) e^{\left(\frac{-i}{c}\mathbf{A}(\mathbf{R}_p)\cdot\mathbf{r}\right)} \tag{28}$$

where $\chi_p(r)$ are the actual atomic orbital functions (Slater or Gaussian orbital function), and R_P is the center of orbitals. By applying the variational principle in wave functions of this type of expansion, only those that fulfill the requirement of being GIAOs would be considered, and as a result, the energy is also invariant. In general, the use of GIAOs leads to better results for smaller basis sets, because it allows more flexibility in the presence of an external magnetic field.

To better convert the NMR tensors into chemical shifts and compare with experimental results, we use tetramethylsilane, $SiMe_4$, as the reference compound, so that we can calculate the chemical shift by subtracting the shielding tensor, σ_i, of the compound of interest from that of the reference compound, σ_{ref}:[48]

$$\delta_i = \sigma_{ref} - \sigma_i \tag{29}$$

1.4.3 Coupling constants

The coupling constants are defined as the difference between two adjacent sub-peaks in a split signal in NMR spectra, in units of Hertz, and depend on multiple factors such as bond distances and dihedral angles.[47] One of the first works on the coupling constants was completed by Karplus,[49] where the Karplus equation was proposed to calculate coupling constants for approximation, $^3J_{H\text{-}H}$:

$$J_{H\text{-}H} = A + B\cos\phi + C\cos 2\phi \tag{30}$$

where A, B, and C are empirically derived constants for a nonionic six-electron,

six-orbital fragment system of H-C-C'-H', and ϕ is the dihedral angle between H-C-C'-H'.[47,49]

The spin-spin coupling constants are calculated using DFT differently. There are four contribution terms of the spin-spin coupling constants: the spin dipole (SD), the Fermi contact (FC), diamagnetic spin-orbit (DSO) and paramagnetic spin-orbit (PSO) terms.[50] For a system with n bonds between the two nuclei:

$$^{n}\boldsymbol{J} = {}^{n}\boldsymbol{J}_{FC} + {}^{n}\boldsymbol{J}_{SD} + {}^{n}\boldsymbol{J}_{PSO} + {}^{n}\boldsymbol{J}_{DSO} \tag{31}$$

The SD and FC terms depend on the interaction between the nuclear magnetic field and the spin of the electrons, and the PSO and DSO terms result from the interactions between the nuclear magnetic moments and those created by the movement of electrons.[47]

1.5 Methods

1.5.1 Conformational sampling

The open-source Python package autodE is applied to locate Transition States (TSs) TSs (saddle points), reactants, and products (minima) for each step in each pathway, and produce a reaction energy profile based on the input of the structure (in this case, .xyz files of reactants and product) through double-ended methods.[51] The general workflow is to find TS based on the structures of reactants and products input by the user, and then the lowest energy TS conformer are kept from the generated ones, and finally single point calculations (level of theory defined by the user) are performed to obtain the reaction profile (Figure 1). Two levels of theory are set in using the package: a lower-level method (lmethod) and a higher-level method (hmethod). For reactants and products, conformational searches are done using the ETKDGv2 algorithm implemented in RDkit.[52,53] All the conformers are the first optimized with lmethod, excluding potentially the same structures by an RMSD threshold (default 0.3 Å) and an energy threshold (default 1 kJ/mol), and then optimized with hmethod, keeping the one of the lowest energy. For reactive reactant complex in bimolecular reactions ($b_P = b_R$), to find an initial reactant complex on the PES, an energy function U_{RR} is minimized with respect to rigid-body rotation and translation of one reactive component

$$U_{RR}(X) = \sum_{\{ac\}} c_1(r_{ac} - c_5 r_{ac}^{avg})^2 + \sum_{\{acx\}} c_2(1 - \cos\vartheta)^2 + \sum_{\{acx\}} c_3(1 - \cos\varphi)^2 + \sum_{ij} \frac{c_4}{r_{ij}^4} \tag{32}$$

where empirical parameters c_1=c_2=c_3=1, c_4=10 and c_5=2.5 (1.5) for charged (uncharged) complexes. All the four terms (harmonic, angle, and repulsion) are set here to maximize the colinearity of v_{an}, v_{ac} and v_{cx} while keeping steric repulsion low as the distance r_{ac} approaching $|v_{ac}|$.

To locate TS, an associated molecular graph for each species, where atoms corre-

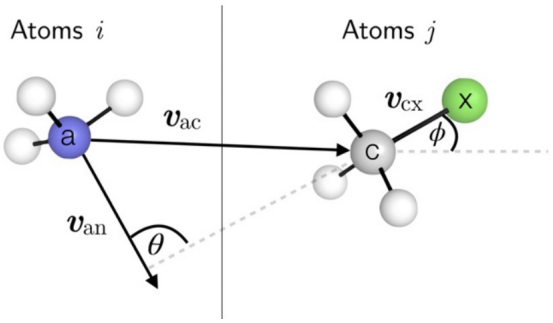


Figure 9 Illustration of the geometric parameters used in Equation 32: using the S_N2 reaction $NH_3 + CH_3Cl$ as an example: v_{an} is the average of vectors v_{am} where m is a nearest neighbor to a (adapted from Angewandte Chemie, 2021, 133, 4312–4320).

spond to vertices (V) and bonds correspond to edges (E), is used to find a bijective mapping from the reactant graph (R) to the product graph (P), representing all the possible bond rearrangements [g: E(R) → E(R')], using NetworkX.[54] A set of transformations of R, [R'], are then generated by adding and/or removing edges. To reduce the search space for g, the transformations are limited to a maximum of four "active" bonds, where two bonds are formed and two bonds dissociated. In a graph X with b bonds ($b_X = |E(X)|$), the change in the number of bonds Δb is defined as $\Delta b = b_R - b_P$, and if $b_R < b_P$, the principle of microscopic reversibility would be used to swap reactants and products. To further speed up, Δb_k is only calculated when k is a known bond type such as "C-C", and an atom cannot exceed its maximal valence. For instance, for a carbon atom, $d(a) \leq 4$. Finally, a TS is generated for each element in the set of valid g found above, and after conformational sampling, the TS with the lowest energy is picked as the final guess. The program would at first attempt to obtain a TS guess through constrained geometry optimization along a linear path between reactants and products. If it fails to do so, a 2D PES scan would be performed through constrained optimizations and fitting a polynomial up to the order of 3, or through a nudge elastic band, and then the lowest-energy saddle point is located where the first derivative vanishes, and then with Dijkstra's algorithm and implemented in NetworkX, followed by a constrained optimization to generate the TS guess.[54–56] The TS guess is then optimized by the user-defined electronic structure theory package and level of theory, where the "Goodness" of a TS is measured by the vibrational frequency of a significant imaginary mode ($|\nu_{imag}|$>45 cm^{-1}) and

those of significant active bonds. This is based on the method of quick reaction coordinate calculation.[57] In the non-isomorphic case, where the reactant or product(s) are not isomorphic to R or P, isomorphism would be checked for forward and backward displacements of the rigid-body conformers generated and optimized. To reduce the computational cost of TS generation, C-X bonds where the carbon atoms are fully saturated and at least two bonds away from the active atoms are truncated, and replaced by hydrogen atoms. The truncated complex is used to locate the TS with TS templates. The reactant molecule or complex must match the template in charge, multiplicity, and solvent parameters, and the active bonds in terms of number and atom type, and their connectivity with nearest neighbors. After successfully locating the TS, the template can be saved into a local library, and the length of the active bonds is fixed at the values in the TS saved in the template. TS conformers are generated based on the equation

$$U_{RB'}(X) = \sum_{i,j \in bonds} k_1 (r_{ij} - r_{ij}^{avg})^2 + \sum_{i,j \in bonds} k_1' (r_{ij} - r_{ij}^{avg})^2 + \sum_{i>j} \frac{k_2}{r_{ij}^8} \qquad (33)$$

where the first represents the harmonic oscillator manner of bonds (r_{avg} is used as the ideal bond length here, obtained as averages from the Cambridge Structural Database, and $k_1' = 10k_1$), and the second term represents the non-physical repulsion in a notion of steric repulsion.[58] 3D conformer structures are generated by minimizing the function U_{RB} with respect to all the coordinates within a cubic box. Same as for locating TS, the lowest energy conformer is found through constrained optimization with the active bonds and is set as a model for the TS, transition state analog (TSA), and its "goodness" and energy are checked to ensure that it is both a "good" TS and is lower in energy than the previous TS.

1.5.2 Energetic span approximation

The energy span approximation can be applied here since our reaction can be viewed potentially as a catalytic cycle, where the anion of the final product, (-)-angiopterlactone is protonated by the neutral starting material, δ-lactone **ent-**

3.[59] In this model, we assume there is one transition state and one intermediate which determine the turnover frequency (TOF), the TOF-determining transition state (TDTS), and the TOF-determining intermediate (TDI). We choose TDI and TDTS that maximize the energy span, ΔG_r (the difference in free energy between the reactants and products). TOF can be expressed as:[59]

$$\mathrm{TOF} = \left(\frac{k_B T}{h}\right) e^{-\delta E/RT} \tag{34}$$

if TDTS appears after TDI

$$\delta E = T_{TDTS} - I_{TDI} \tag{35}$$

if TDTS appears before TDI

$$\delta E = T_{TDTS} - I_{TDI} + \Delta G_r \tag{36}$$

1.5.3 Microkinetic model

The basic idea of the microkinetic model is that by using the rate law expression, the rate of a reaction can be expressed in a differential equation, and then by inputting the initial concentration of reactants and rate constants, the differential equation can be solved at each time step. Using the activation barriers of DFT calculations and the Eyring–Polanyi equation, the rate constant of both forward and backward reactions can be obtained.[60]

For a reaction $n_a A + n_b B \rightleftharpoons n_c C + n_d D$, the rate can be expressed as:

$$\mathrm{rate} = \frac{1}{n_c}\frac{d[C]}{dt} = \frac{1}{n_d}\frac{d[D]}{dt} = -\frac{1}{n_a}\frac{d[A]}{dt} = -\frac{1}{n_b}\frac{d[B]}{dt} = k_{\mathrm{forward}}[A]^x[B]^y \tag{37}$$

where x and y are the orders of the reactants. Taking the reaction's reversibility into account, where we have a forward reaction and a backward reaction, and ΔG = $\Delta G^{\ddagger}_{\text{forward}}$ - $\Delta G^{\ddagger}_{\text{backward}}$, the equilibrium constant K_{eq} is:

$$K_{\text{eq}} = \frac{[C]^{n_c}[D]^{n_d}}{[A]^{n_a}[B]^{n_b}} = \frac{k_{\text{forward}}}{k_{\text{backward}}} = \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \tag{38}$$

And the rate constant k is related to the free energy barrier by the Transition State Theory (TST), and is defined by the Eyring–Polanyi equation,

$$k = \frac{\kappa k_{\text{B}} T}{h} \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \tag{39}$$

where k is the reaction rate constant, $\varkappa$ is the transmission factor (usually assumed to be 1 in homogenous catalysis), k_B is the Boltzmann's constant, T is the temperature, h is the Planck's constant, $\Delta G^{\ddagger}$ is the activation free energy, and R is the ideal gas constant. The rate constant k is calculated from the activation barriers for each step obtained in DFT calculations, using the Erying equation. Depending on the order of reaction, the dimensions of the rate constant would vary based on concentration per time unit.

2 Computational Details

2.1 Optimisation and benchmarking methods

All calculations were carried out using the ORCA suite of programs (version 5.0.3).[61] Both optimizations and single-point energy calculations were carried out using "Tight" convergence criteria, corresponding to tolerances of an energy change of 10^{-8} between two cycles of optimization. All optimizations were conducted at the level of ωB97X-D3/def2-TZVP, with a grid scheme "DEFGRID3", with the coulomb fitting basis set of def2/J, with the resolution-of-identity chain-of-spheres exchange (RIJCOSX) approximation.[62–64]

Benchmarking was done at the level of SCS-MP2/def2-TZVPP and RI-SCS-MP2/def2-TZVPP.[65] For RI-SCS-MP2, the RIJCOSX approximation was employed, and for correlation-exchange, "AngularGrid 6" was employed (corresponding to a Lebedev-590 angular grid), and with a radial grid scheme (IntAcc) of 5.0. For the Hartree-Fock exchange, "AngularGridX 2,3,4" was employed for the SCF grid, and "AngularGridX 4" was employed for the MP2 gradient, "IntAcc 4.5" for MP2 second derivatives.[66–69] The auxiliary basis of def2-TZVPP/C set was used.[68] Weigend's "universal" Coulomb fitting basis, the def2/J auxiliary basis set, and the basis set fitting Coulomb and exchange, def2/JK were applied.[68]

Numerical frequencies were calculated for the SMD solvation model, using dichloroethane (DCE), in ORCA.[70] For all optimized structures, frequency calculations were checked to ensure that this is either a minima, for reactants, intermediates, and products, or a saddle point, for transition states (TSs). Grimme's quasiRRHO approach was used to calculate free energies at 343 K.[71] The python-script *otherm.py* was used for energy correction.[72] Single point calculation was done at the level of DLPNO-CCSD(T)/def2-TZVPP//ωB97X-D3/def2-TZVP, with an auxiliary basis set of def2-TZVPP/C applied.[62,63,73] DLPNO-CCSD(T) calculations were run with the default "NormalPNO" cut-offs, corresponding to $T_{CutPairs} = 10^{-4}$ Hartrees, $T_{CutPNO} = 3.33 \times 10^{-7}$ and $T_{CutMKN} = 10^{-3}$, where $T_{CutPairs}$ is the cut-off for estimated pair correlation energies, T_{CutPNO} is the threshold to decide if a virtual orbital should be

neglected based on PNO occupation numbers, and T_{CutMKN} is the threshold based on orbital Mulliken populations.[74]

For Pathway **A**, preliminary geometries of all structures were obtained via autodE using three levels of theories, including MP2/cc-PVQZ//PBE0-D3BJ/def2-SVP, MP2/cc-PVQZ//ωB97X-D3/def2-TZVP, MP2/cc-PVQZ//SCS-MP2/def2-TZVPP, using an epsilon of 10.36 CPCM solvent for dichloroethane, and GFN2-XTB for initial conformational sampling.[51,62,63,65,75–78]

2.2 NBO analysis

Natural bond orbital analysis (NBO) was done at the level of ωB97X-D3/def2-TZVP using NBO 7.0 in ORCA.[?] With the natural polyelectron population analysis (NPEPA) method numerically implemented in the NBO 7.0 program, an updated version of polyelectron population analysis (PEPA) that evaluates conventional orbital population, the NBO analysis transforms a given form of the wavefunction into a localized form, such as 1-center atomic orbital (NAO), 1-center hybrid orbital (NHO), or 2-center bond orbital (NBO) type, and specifies occupied and vacant spin-orbital.[79–81]

2.3 NMR

All nuclear magnetic resonance (NMR) calculations were done using ORCA 5.0.3. NMR chemical shifts were calculated at the level of ωB97X-D3/def2-TZVP, with several test runs at the level of DSD-PBEP86-D3BJ/def2-QZVP, RI-B2PLYP/def2-QZVP, and mPW1PW91/def2-TZVP, under the "NMR" module of ORCA, utilizing Gauge Including Atomic Orbitals (GIAOs).[35,47,82–87]

2.4 pK_a

The pK_a calculations for a molecule (HA) to its deprotonated form G(A^-) were done using Ho's approach, following this equation to compute solution phase dissociation free energies:[88]

$$\Delta G* = G(A^-) - G(HA) + G*(H^+) \quad (40)$$

G(A^-) and G(HA) were computed at the same level of theory, DLPNO-CCSD(T))/def2-TZVPP//ωB97X-D3/def2-TZVP at 343K as all the other geometries here, and in three different solvents, dichloroethane (DCE), dimethylsulfoxide (DMSO), and water. Values for the proton free energy, G*(H^+), were cited from reported values in previous literature.[89] There were no reported values for DCE, and the free proton energy of methanol, -263.5 kcal/mol, was used. For dimethylsulfoxide (DMSO), the reported value of -273.3 kcal/mol was used, and for water, -265.9 kcal/mol was used. Since our purpose here is to compare relative acidity rather than computing absolute acidity, the pK_a values were compared with each other without empirical corrections.

2.5 Microkinetic model

The software Tenua is used to simulate reaction.[90] The initial concentration of the only starting material, δ-latone **ent-3** was input for 0.124 mol/L, same as the experimental condition.The rate constants for backward and forward reactions for each step are calculated by applying the Erying equation to DFT-calculated barriers. For deprotonation of δ-lactone **ent-3**, the activation barrier obtained through the method in 1.5 tends to be overestimated, and for both forward and backward reaction, the rate constant of 1 s^{-1} is used, corresponding to a barrier of 20.2 kcal/mol, 5-10 kcal/mol lower than the computed barrier for deprotonation, as a correction for the overestimated barriers.

3 Mechanistic studies

3.1 PK_a

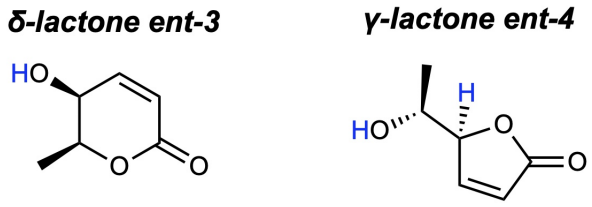


Table 1 Free energy (ΔG) and pK_a values for lactones ent-3 and ent-4 in DCE, DMSO, and water at 343 K.

Solvent	property	δ-lactone ent-3 $O^- \rightarrow OH$	γ-lactone ent-4 $O^- \rightarrow OH$	γ-lactone ent-4 $C^- \rightarrow CH$
DCE	Δ G (kcal/mol)	46.3	58.3	40.3
	pK_a	29.5	37.1	25.7
DMSO	ΔG (kcal/mol)	44.3	56.7	37.9
	pK_a	28.3	36.1	24.1
H_2O	ΔG (kcal/mol)	37.3	48.9	38.0
	pK_a	23.8	31.1	24.2

From the structure of the product (Figure 1) it is deduced that a dimerization happens between the two lactone molecules, but the nucleophile and the electrophile are to be determined by pK_a calculation. To propose the mechanistic pathways, pK_a calculations are performed for the alcohol group and the C-H group on **ent-3** and **ent-4** (Table 1). The activation barriers of deprotonation tend to be overestimated due to the approach used here, and though there were empirical corrections proposed in previous literature, the computed values are not corrected because we only need to compare the relative pK_a values to deduce the possible mechanisms in this work.[91]

The pK_a of the alcohol moiety of **ent-4** is more than seven pK_a units higher than that of **ent-3** in all three solvents. Therefore, under the reaction condition, a higher concentration of **ent-3** will be present than that of **ent-4**. The pK_a of the CH proton on **ent-4** is more than ten pK_a units lower than that of the OH proton on **ent-4** in dichloroethane (DCE) and DMSO, and more than six pK_a units in water. This

result explains the experimentally observed epimerisation at this position under the reaction condition. From the difference between pK_a values of the three groups, we conclude that it is more likely that **ent-3** acts as the nucleophile, attacking **ent-4** in our reaction.

3.2 Pathways and energy diagrams

3.2.1 Mechanistic pathways

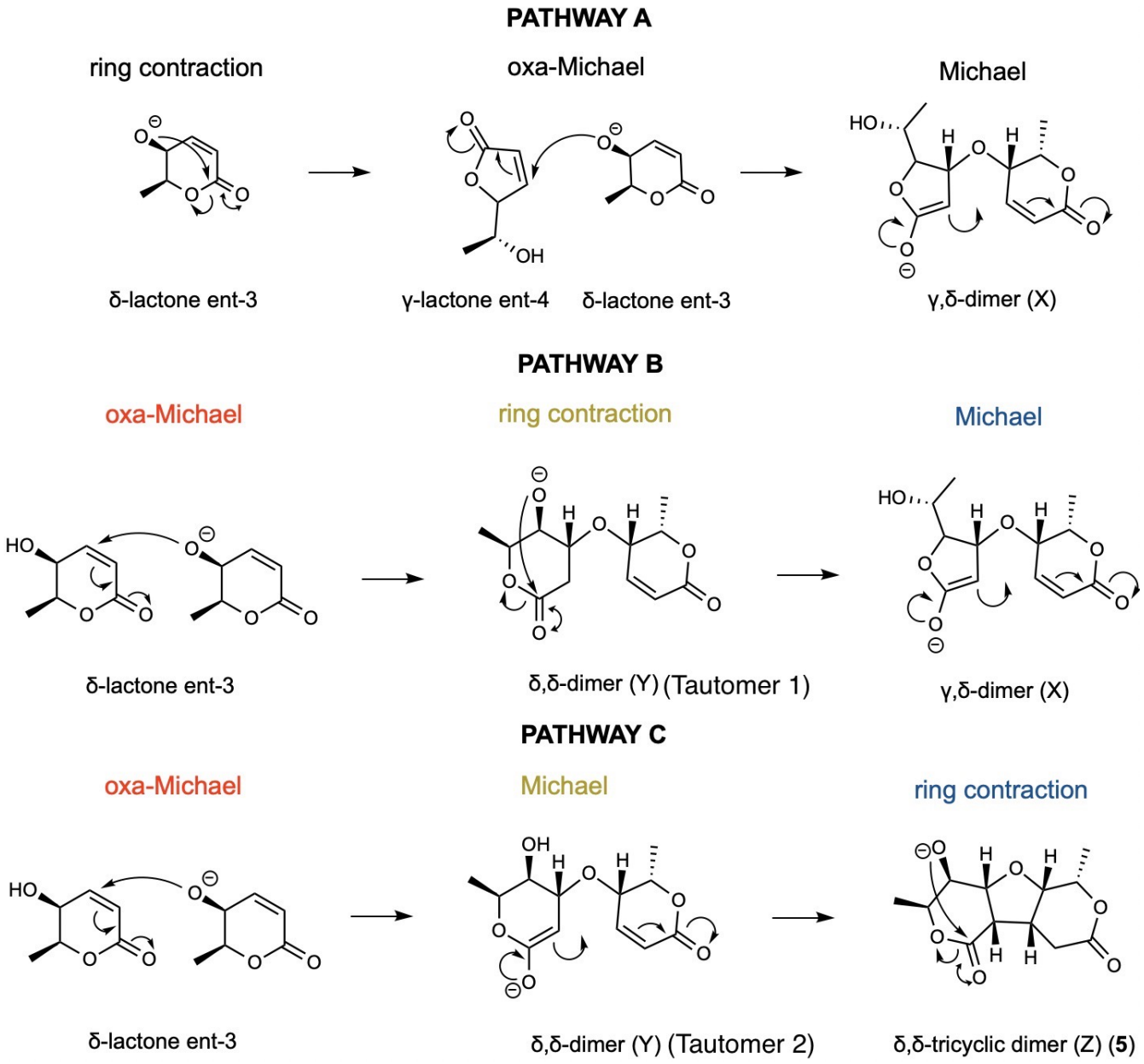


Figure 10 All three mechanistic pathways proposed, labeled from **A** to **C**. The intermediate formed in the ring contraction/Michael step in pathway **B** and **C** are the same type of dimer in terms of dimerization (δ-lactone to δ-lactone), but different tautomers.

A domino sequence of a ring-contraction/oxa-Michael/Michael reaction is proposed, with the anion of **ent-3** attacking **ent-4** followed by a ring contraction (pathway **A**).

The order of these steps might be different, so pathway **B** and **C** are proposed. In the oxa-Michael step, the double bond on the eletrophile can be attacked from both sides of the plane it belongs to, giving rise to a chiral center in a single step, which is defined as prochirality. The face attacked can be labeled as Re or Si. The Re face is where the substituents of the attacked carbon are arranged in increasing Cahn-Ingold-Prelog priority order clockwise, and the Si face is in anti-clockwise order.

Pathway A: ring contraction(A1)/oxa-Michael(A2)/Michael sequence(A3). The first step in pathway **A** involves a ring contraction from **ent-3** to **ent-4**. The ring contraction process has two steps: the first step is a 5-exo-trig cyclization (RC1), and the second step is a retro-5-exo-trig ring opening (RC2). Then an intermolecular oxa-Michael addition (OM) happens between **ent-3** and **ent-4**, forming the anion of angiopterlactone A (**1**), followed by a Michael addition (M) which forms the final product angiopterlactone B (**2**).

Pathway B: oxa-Michael(B1)/ring contraction(B2)/Michael sequence(B3). Pathway **B** starts with an oxa-Michael addition between neutral **ent-3** and anion **ent-3**. The resulting δ,δ-dimer(Y) then undergoes a two-step ring contraction, followed by a Michael addition of the γ,δ-dimer (X), the same Michael addition as the one in pathway **A**.

Pathway C: oxa-Michael(C1)/Michael sequence(C2)/ring contraction(C3) sequence. Pathway **C** initiates with the same oxa-Michael addition between two molecules of **ent-3**. An intramolecular Michael addition then takes place to form **5**, an intermediate observed in the experiment by us. The last step is a two-step ring contraction.

3.2.2 Pathway A

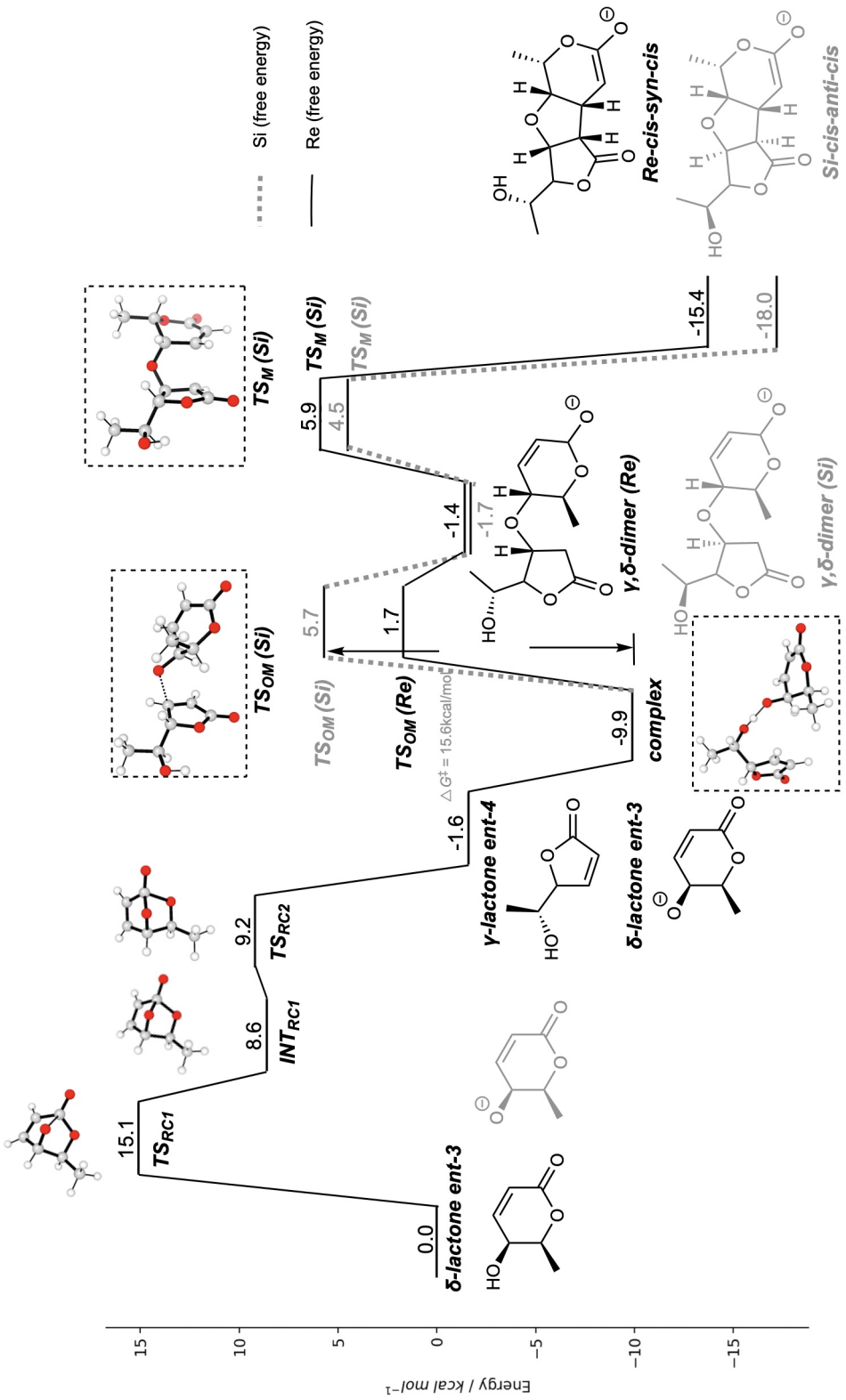


Figure 11 Free energy profile for pathway **A**, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

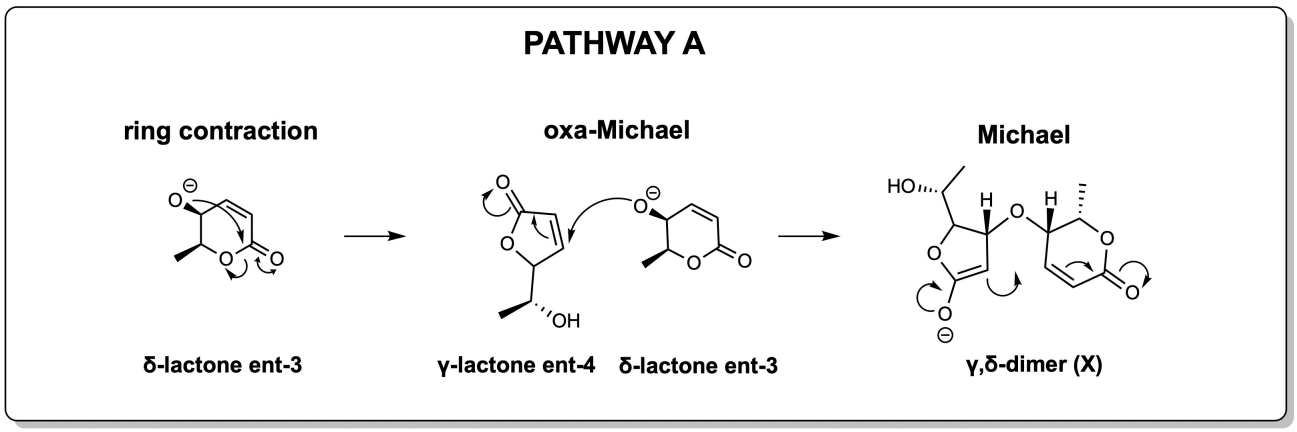


Figure 12 Mechanistic pathway **A**.

By applying the energy span approximation to the energy profile (Figure 11), we obtain δE = 15.8 kcal for **A**-Re (TDTS being TS_M, TDI being the complex), and δE = 15.6 kcal/mol for **A**-Si (TDTS being TS_{OM}, TDI being the complex).

The two-step ring contraction in Pathway **A** starts with an RC1 step with a high activation barrier, 15.1 kcal/mol, close to both δE. Since there are two steps close in the activation barrier, the energy span approximation is not very effective here. The RC2 step, though, has a low activation barrier of 0.6 kcal/mol. TS_{RC1} has an activation barrier much higher than TS_{RC2}. The energy cost for the six-member ring **3** to overcome the strain that stabilizes it to form the bridge-like INT_{RC1} is relatively high. The five-member ring of **ent-4** formed in this step is 1.6 kcal/mol lower than **ent-4**, so **ent-4** is thermodynamically more stable than **ent-3**. The intermediates formed in this step, INT_{OM} (Re) and (Si), are close to zero (-1.4 kcal/mol and -1.7 kcal/mol, respectively), suggesting that the oxa-Michael step could be reversible.

The Re/Si prochirality starts to diverge in the oxa-Michael step, where the double carbon bond on **ent-3** can be attacked from both sides, as we can see in the TS structures (Figure 13). The activation barrier of the Re attack is 4 kcal/mol lower than that of the Si attack, due to the hydrogen bonding forming between the hydrogen atom of the OH group on **ent-4** and the axial oxygen atom on **ent-3**, which facilitates the reaction. The hydrogen-bonded complex (-9.9 kcal/mol) formed between the oxa-Michael step contains a similar type of hydrogen bonding, and the TS_{OM} (Si) loses this hydrogen bonding, resulting in a higher activation barrier than the TS_{OM}



Figure 13 TS structures for the Re and the Si face for the oxa-Michael step (TS_{OM}) in pathway **A**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the labeled hydrogen bonding for Re is 1.65Å. The length of the forming C-O bond Re is 1.88Å, and that for Si is 2.09Å.

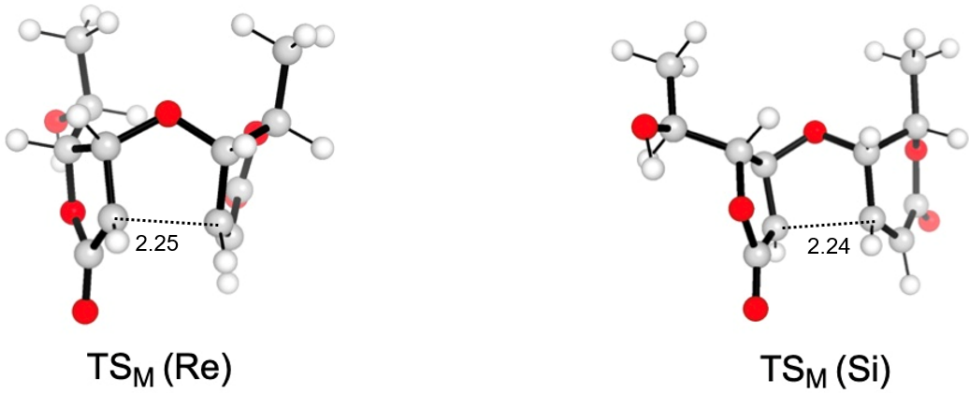


Figure 14 TS structures for the Re and the Si face for the Michael step (TS_{M}) in pathway **A**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the forming C-C bond for Re is 2.25Å, and that for Si is 2.24Å.

(Re).

In the last step of Michael addition, the difference between the activation barriers of TS_M (Re) and TS_M (Si) is 1.4 kcal/mol. The TS_M (Re) has the C-H bonds at the four stereocenters in the middle of the molecule pointing out, making the steric hindrance larger than that of the TS_M (Si) (Figure 14). Therefore, TS_M (Re) is higher in activation barrier than TS_M (Si). The final Re-cis-syn-cis product, **I** anion, is 2.6 kcal/mol higher in energy than the Si-cis-anti-cis product, **II** anion (Figure 15). Comparing the structure of **I** and **II**, we can see that **I** undergoes a strain, similar to TS_M (Re), due to the configuration of C-H bonds in the middle point out, forcing the

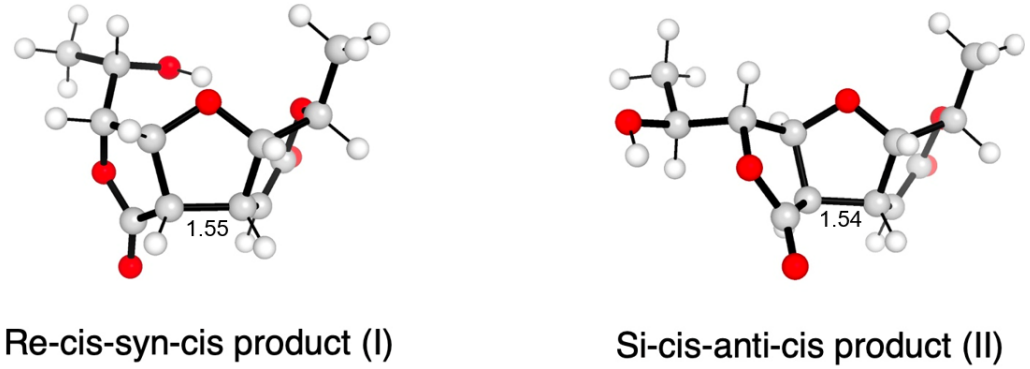


Figure 15 Final product structures for the Re and the Si face for the Michael step (Re-cis-syn-cis (**I**) and Si-cis-anti-cis (**II**)) in pathway **A**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the formed C-C bond for Re is 1.55Å, and that for Si is 1.54Å.

two rings on the left and right side to align on the same side of the molecule. While the two rings on the left and right side align on opposite sides, lowering the strain, and resulting in lower energy. The Si product is lower in energy than the Re product, but only the Re product is observed, so the reaction cannot be simply recognized as under thermodynamic control.

3.2.3 Pathway B

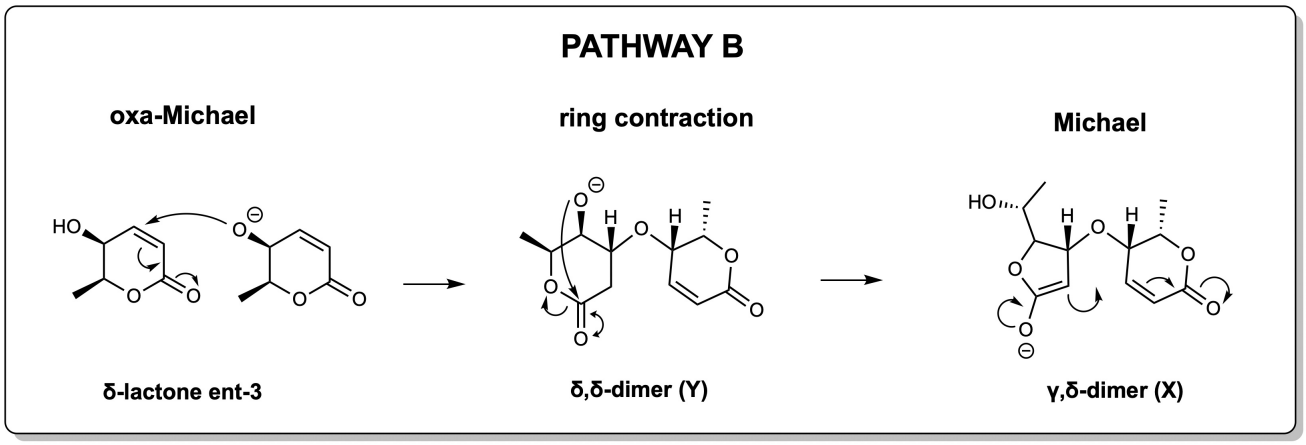


Figure 16 Mechanistic pathway **B**.

Applying the energy span approximation to pathway **B** (Figure 17), for **2**-Re, δE = 16.2 kcal/mol (TDTS being TS_{RC1}, TDI being the complex), and for **2**-Si, δE = 15.2 kcal/mol (TDTS being TS_{OM}, TDI being the complex).

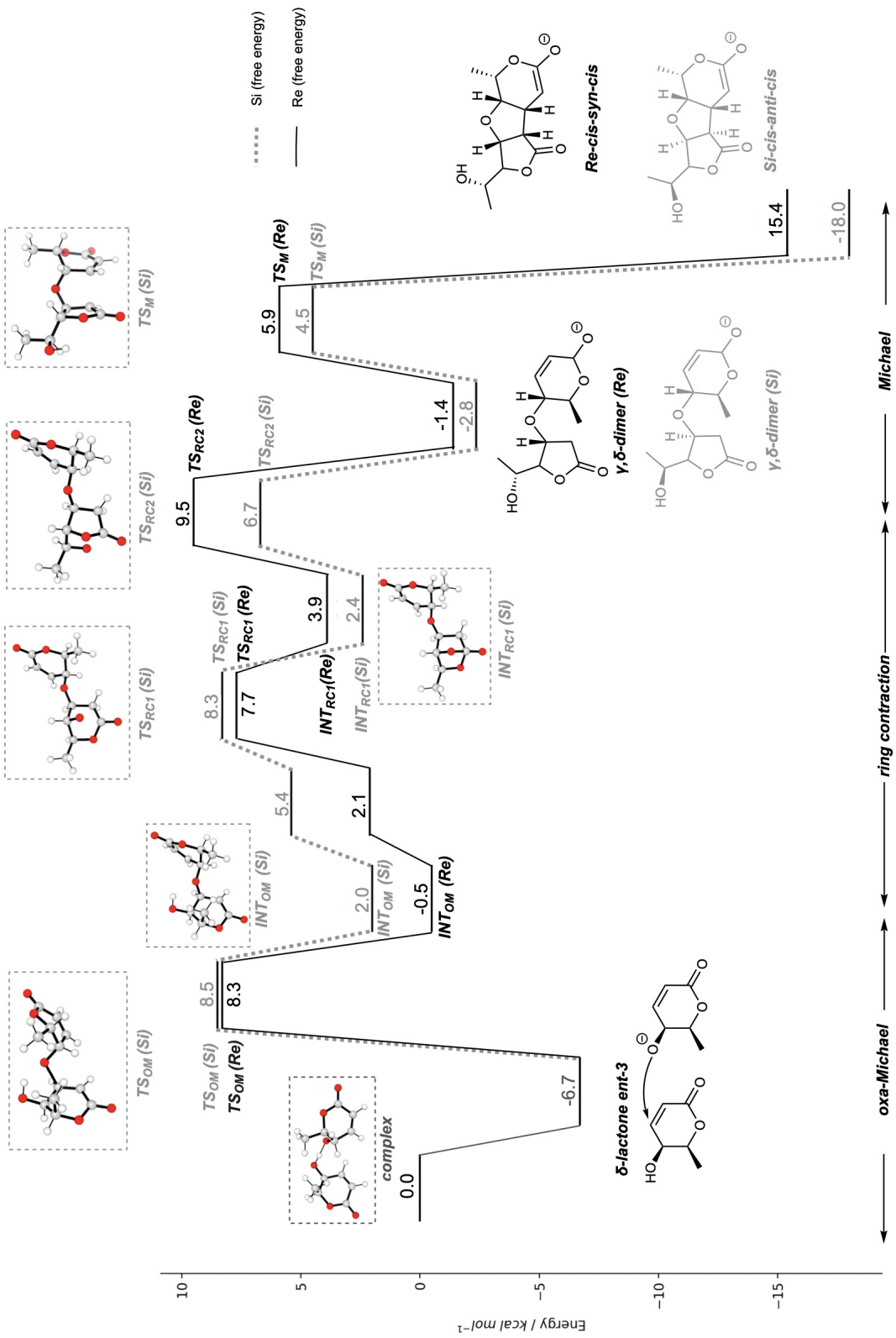


Figure 17 Free energy profile for pathway **B**, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

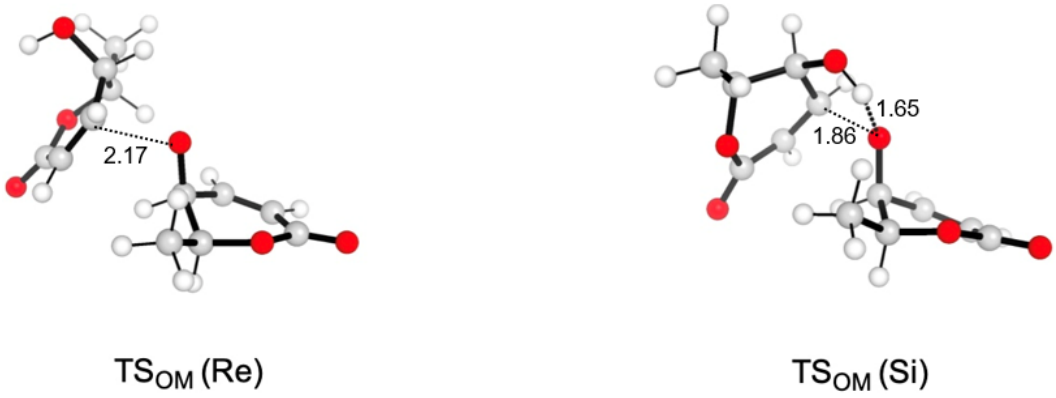


Figure 18 TS structures for the Re and the Si face for the oxa-Michael step (TS$_{OM}$) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The labeled hydrogen bondidng for Si is 1.65Å. The length of the forming C-O bond Re is 2.17Å, and that for Si is 1.86Å.

The first step is an oxa-Michael addition between **ent-3** anion and neutral **ent-3**. The double carbon bond on **ent-3** can be attacked from both sides, giving Re/Si products. The two molecules of **ent-3** first form a low-energy complex (-6.7 kcal/mol) with a hydrogen bonding between the hydrogen atom on the OH group of neutral **ent-3** and the axial oxygen atom with negative charge on anion **ent-3**. The TS$_{OM}$ (Si) keeps the same intermolecular hydrogen bonding, while the TS$_{OM}$ (Re) does not (Figure 18). However, in this oxa-Michael addition between the two six-member rings, unlike the one in pathway **A**, the Re and Si activation barrier is close in energy, with a difference of 0.2 kcal/mol, and the TS$_{OM}$ (Re), instead of the TS$_{OM}$ (Si) structure with a hydrogen bonding, is lower in energy. The INT$_{OM}$ (Si) (2.0 kcal/mol) is higher in energy than the INT$_{OM}$ (Re) (-0.5 kcal/mol), and the energy of INT$_{OM}$ (Re) are close to zero, suggesting that the oxa-Michael step is reversible for the Re attack.

In the second step, ring contraction, INT$_{OM}$ from the last step first undergoes a tautomerization to become an enolate with higher energy than INT$_{OM}$, so that the neutral axial OH group on neutral **ent-3** in INT$_{OM}$ (Figure 19) becomes negatively charged and forms an enolate (Figure 20) to be prepared for the ring contraction step next. Both of the enolates are slightly higher in energy than the INT$_{OM}$, with the Re enolate 2.6 kcal/mol higher in energy than INT$_{OM}$ (Re), and the Si enolate 3.4 kcal/mol higher than INT$_{OM}$ (Si). To form these enolates, the INT$_{OM}$ has to change

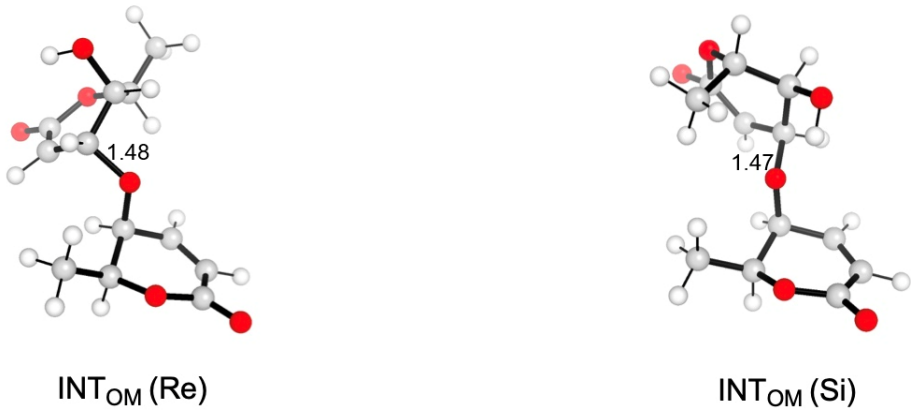


Figure 19 Intermediate structures for the Re and the Si face for the oxa-Michael step (INT_{OM}) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the formed C-O for Re is 1.48Å, and that for Si is 1.47Å.

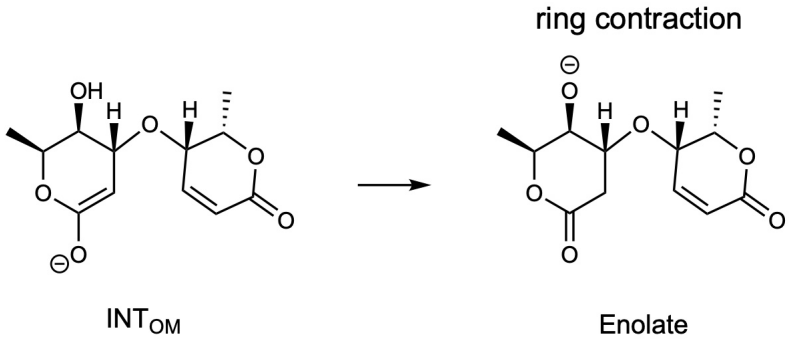


Figure 20 Tautomerization of INT_{OM} in pathway **B**.

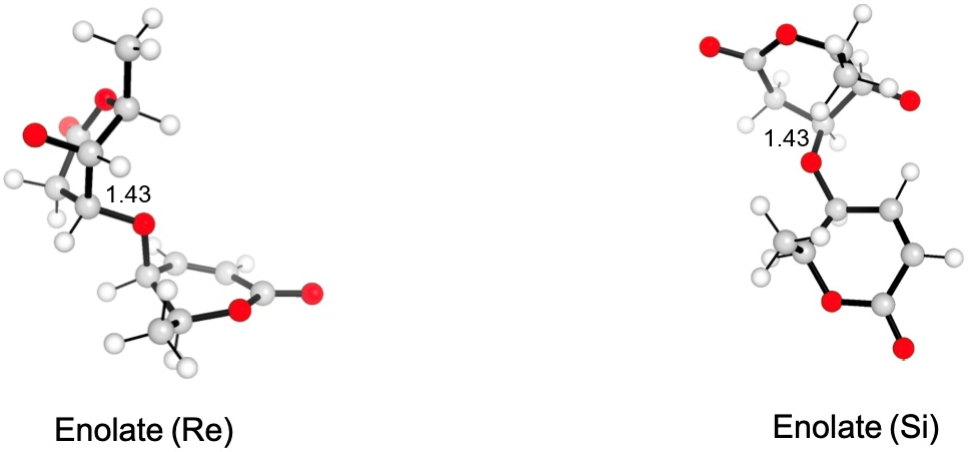
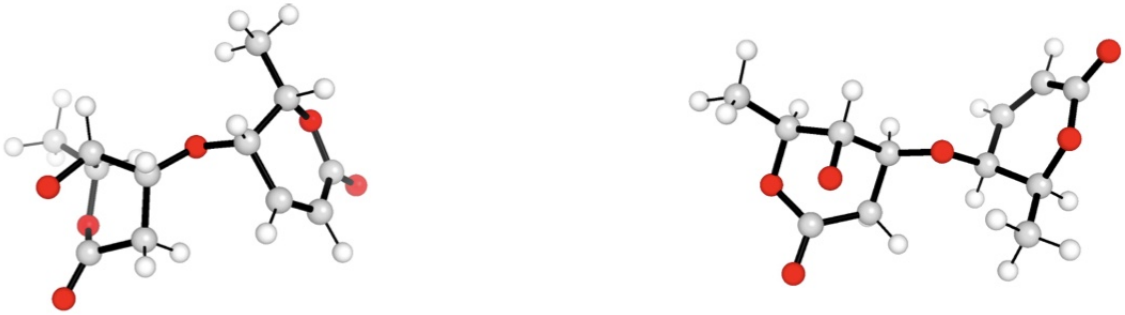
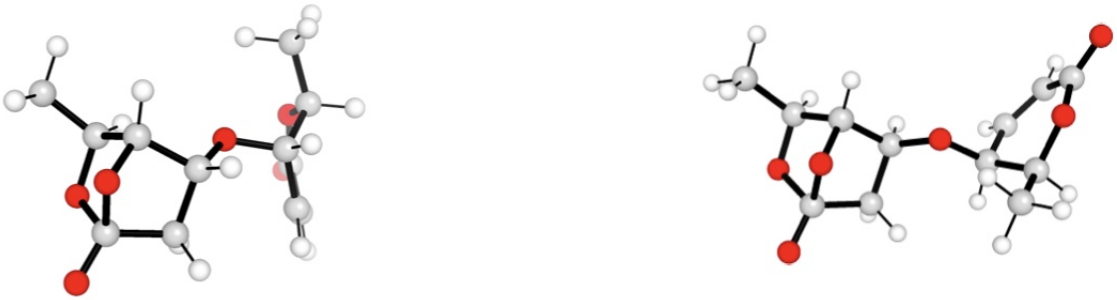


Figure 21 Enolate structures for the Re and the Si face before the ring contraction (RC) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the formed C-O for Re is 1.43Å, and that for Si is 1.43Å, different from those of INT_{OM}.



TS_{RC1} (Re) TS_{RC1} (Si)

Figure 22 TS structures for the Re and the Si face for the first ring contraction step (RC1) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.



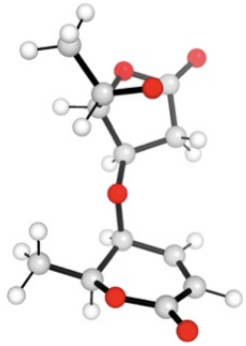
INT_{RC1} (Re) INT_{RC1} (Si)

Figure 23 Intermediate structures for the Re and the Si face for the first ring contraction step (RC1) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

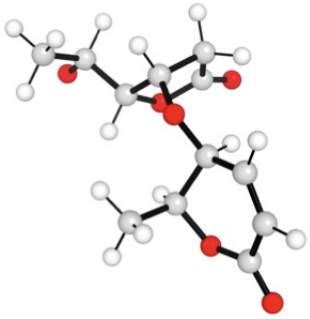


TS_{RC2} (Re) TS_{RC2} (Si)

Figure 24 TS structures for the Re and the Si face for the second ring contraction step (RC2) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.



INT_{RC2} (Re)



INT_{RC2} (Si)

Figure 25 Intermediate structures for the Re and the Si face for the second ring contraction step (RC2) in pathway **B**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

its conformation, with a molecule of **ent-3** rotated to another direction, or slightly changing the conformation of **ent-3** (Figure 19, 20 and 21), increasing steric effects, making the enolates higher in energy than the INT_{OM}.

Unlike the RC1 step in pathway **A** where the activation barrier is high compared to other steps ($\Delta G^{\ddagger}$ = 15.1 kcal/mol), the activation barrier for TS_{RC1} is less than 10 kcal/mol (for TS_{RC1} (Re), $\Delta G^{\ddagger}$ = 8.2 kcal/mol, for TS_{RC1} (Si), $\Delta G^{\ddagger}$ = 6.3 kcal/mol). The equatorial negatively charged oxygen in enolate (Si) becomes axial in TS_{RC1} (Si), while the oxygen in the same position on enolate (Re) remains axial. To make the 5-exo-trig cyclization (the first step of ring contraction, RC1) possible, this oxygen atom has to be in the axial position in TS_{RC1} to form a bridge-like shape (Figure 22). Since the equatorial oxygen in INT_{OM} (Si) becomes axial in TS_{OM} (Si), the activation barrier of TS_{OM} (Si) is 2.1 kcal/mol higher than that of TS_{OM} (Re). The conformations will be discussed in more detail in this chapter later.

TS_{RC1} then forms INT_{RC1}.The INT_{RC1} (Re) (3.9 kcal) is 1.5 kcal/mol higher in energy than INT_{RC1} (Si) (2.4 kcal/mol) (Figure 23). The INT_{RC1} (Si) undergoes smaller steric hindrance due to the alignment of the two six-member rings, as the two six-member rings of INT_{RC1} (Re) are aligning in the same direction, while INT_{RC1} (Si) has them on different sides.

The second step of ring contraction is a retro-5-exo-trig ring opening happening

on the bridge-shaped six-member ring. The activation barrier of the TS_{RC2} (Si) ($\Delta G^{\ddagger}$ = 4.3 kcal/mol) is 1.3 kcal/mol lower than that of the TS_{RC1} (Re) ($\Delta G^{\ddagger}$ = 5.6 kcal/mol) (Figure 24). Similar to the case of TS_{RC1} (Re) and INT_{RC1} (Re), the steric hindrance caused by the aligning of the two rings leads to the higher activation barrier of TS_{RC2} (Re).

The INT_{RC2} following TS_{RC1} is the same as the INT_{OM} in the pathway **A**. The last step in pathway **B**, the Michael addition, is the same as the last step in pathway **A**.

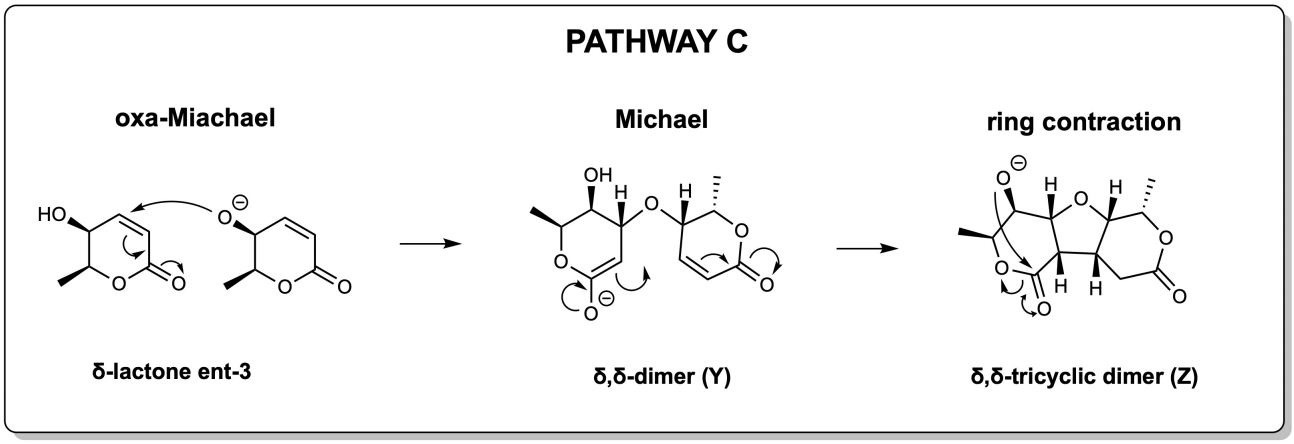


Figure 26 Mechanistic pathway **C**.

3.2.4 Pathway C

According the computed energy profile of pathway **C** (Figure 27, the energy span of pathway **C** is 15.0 kcal/mol for the Re face (TDTS being TS_{OM} (Re), and TDI being the complex), and 17.2 kcal/mol for the Si face (TS_M (Si), and TDI being the complex). The fist step in pathway **C**, is the same as the fist step in pathway **B**.

The second step in pathway **C** is a Michael addition happening on the dimer formed from two molecules of **3** the oxa-Michael step, INT_{OM}. The activation barrier of the TS_M (Si) ($\Delta G^{\ddagger}$ = 8.9 kcal/mol) is 3.1 kcal/mol higher than that of the TS_M (Re) ($\Delta G^{\ddagger}$ = 5.8 kcal/mol). INT_{OM} (Re) and INT_{OM} (Si) (Figure 28) differ in the alignment of the directions of the two six-member rings, resulting in INT_{OM} (Re) lower in energy than INT_{OM} (Si). Similarly, TS_M (Re) is lower in energy than TS_M (Si) because the two rings are aligned in the same direction, or close to being parallel with each other, making the C-C bond easier to form, whereas the two rings in TS_M (Si) are in opposite directions.

The INT_{OM} (Re) is 4.5 kcal/mol lower in energy than the INT_{OM} (Si) (Figure 29. Unlike the case of the final products **I** (Re-syn-cis-syn) and **II** (Si-cis-anti-cis), where the Re product after the Michael step is higher in energy, here the INT_{Re} is lower in energy. The difference is that the final product is formed from a Michael addition following an oxa-Michael addition of **ent-3** (five-member ring) and **ent-4** (six-member ring) in pathway **A**, while the INT_M in pathway **C** is from the two

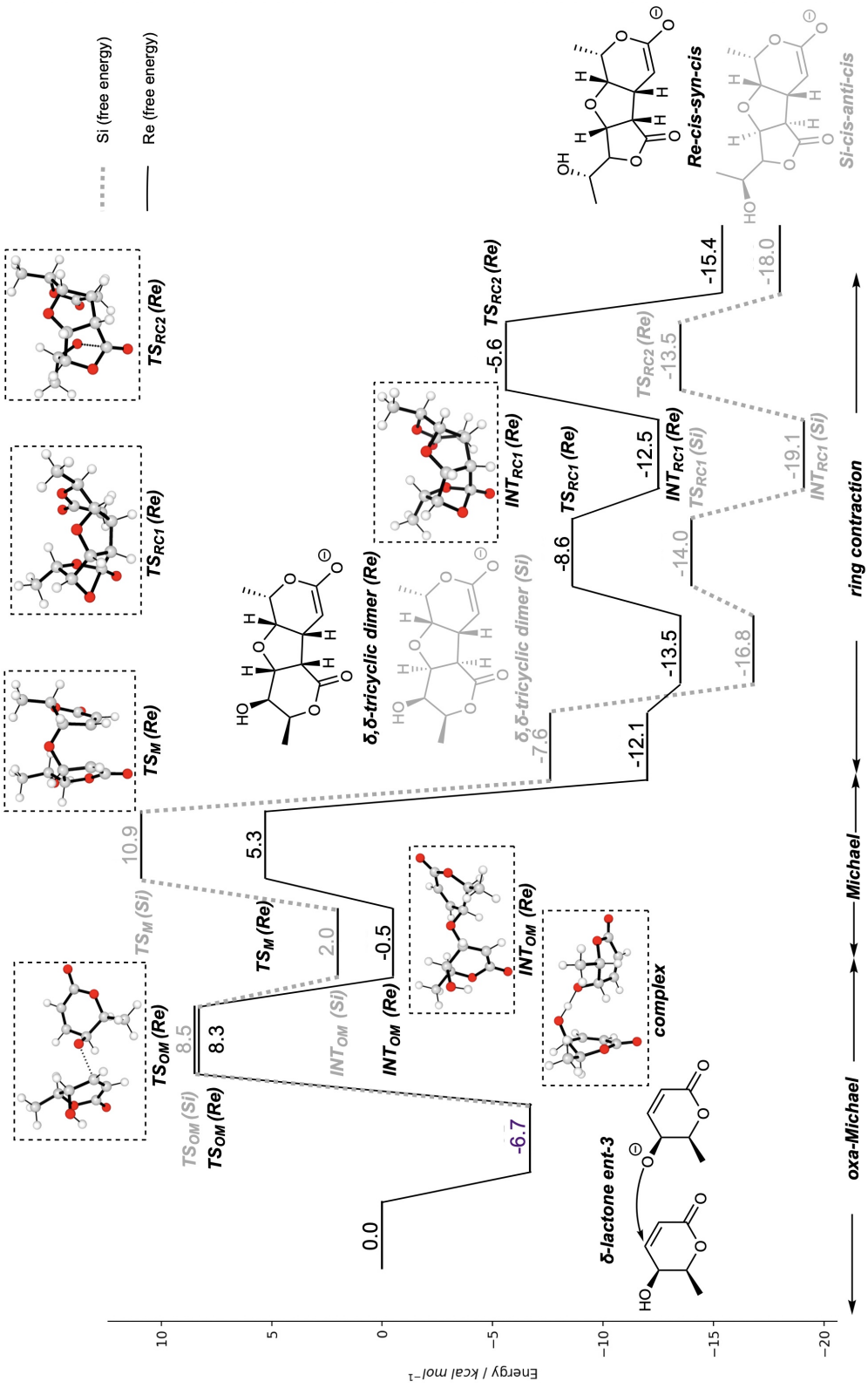


Figure 27 Free energy profile for pathway **C**, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

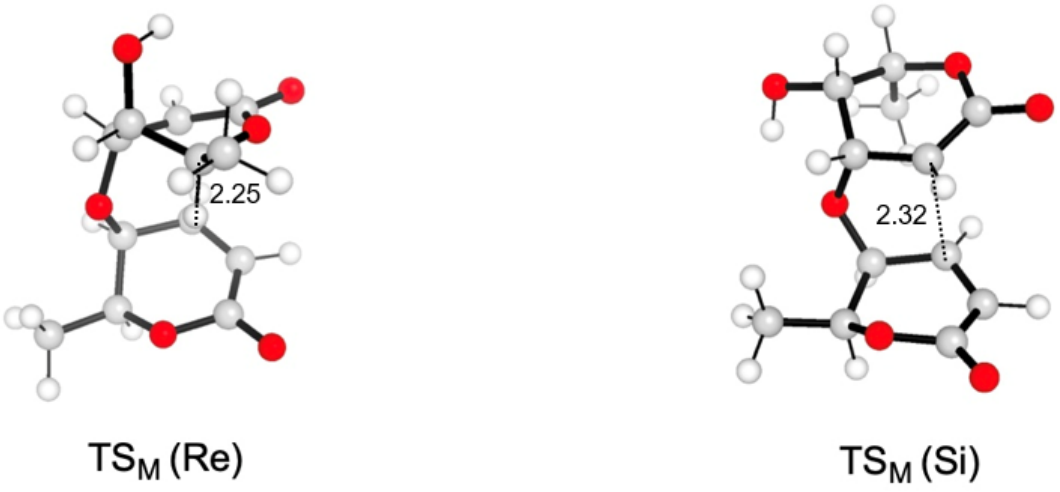


Figure 28 TS structures for the Re and the Si face for the Michael step (TS_M) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.The length of the forming C-C bond for Re is 2.25Å, and that for Si is 2.24Å.

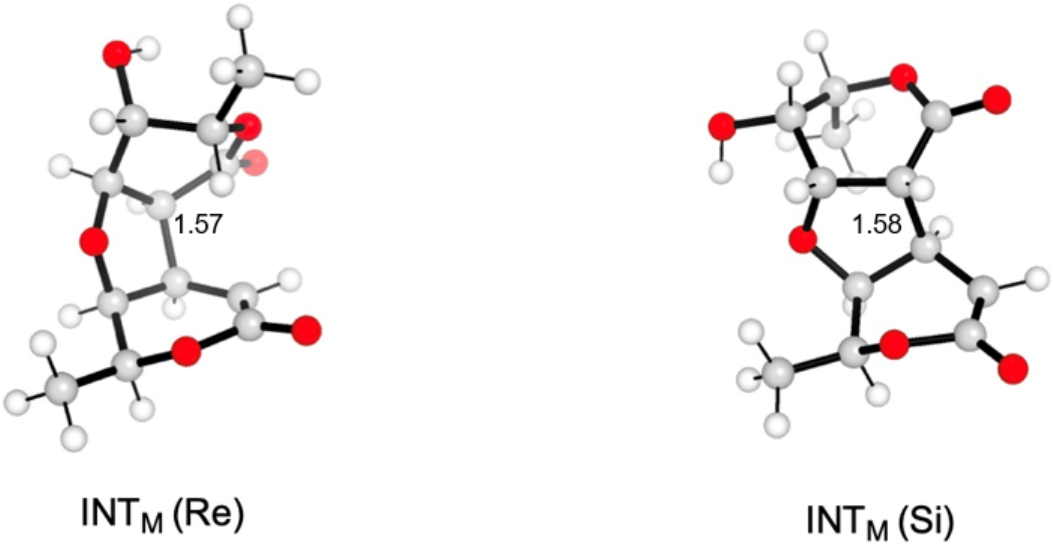


Figure 29 Intermediate structures for the Re and the Si face for the Michael step (TS_M) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. The length of the formed C-C bond for Re is 1.57Å, and that for Si is 1.58Å.

six-member rings. This difference leads to the relative difference in energy of Re and Si intermediates/products.

The last step in pathway **C** is the two-step ring contraction. The INT_M would form an enolate to trigger the first step of ring contraction (Figure 30). The enolate (Si) (-16.8 kcal/mol) is 3.3 kcal/mol lower in energy than the enolate (Re) (-13.5 kcal/mol). The

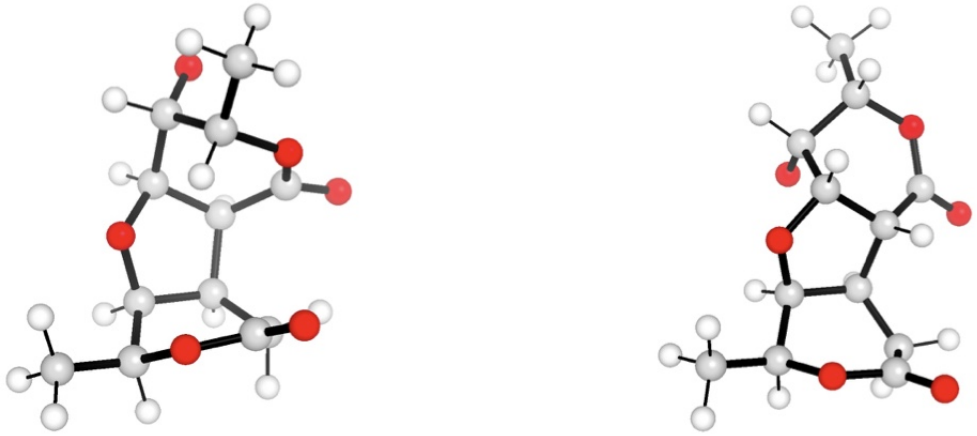


Figure 30 Enolate structures for the Re and the Si face for the first ring contraction step (RC1) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

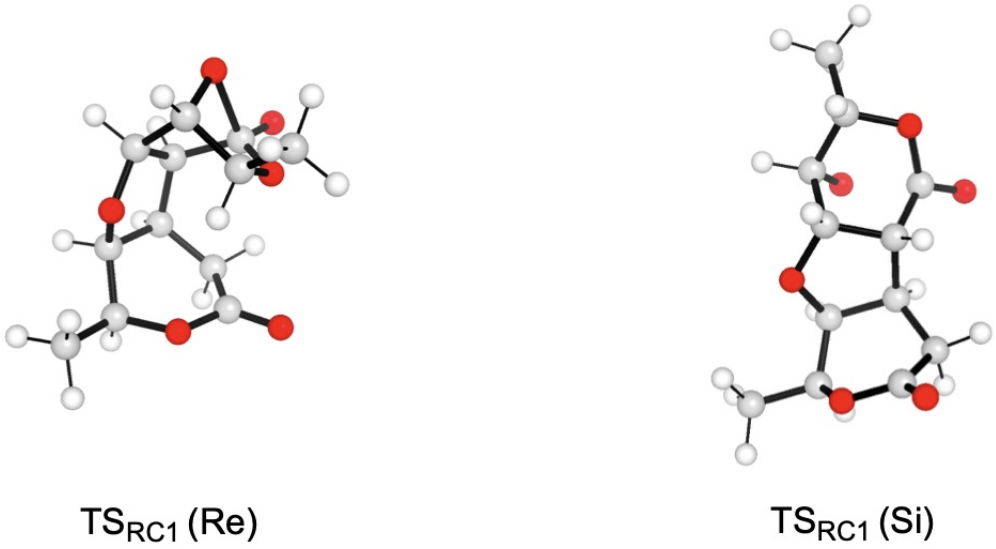
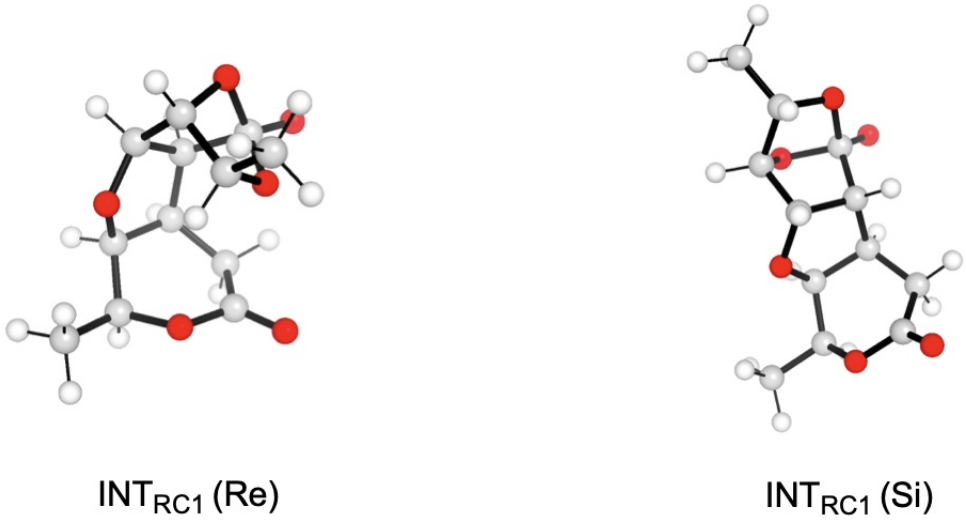


Figure 31 TS structures for the Re and the Si face for the first ring contraction step (RC1) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

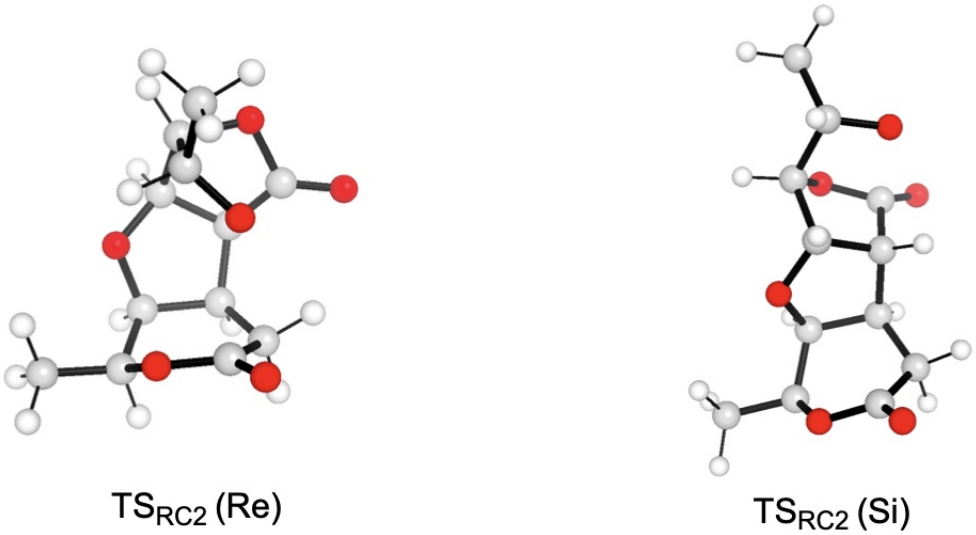
structure of the enolate (Si) is more twisted from INT_M (Si) compared to enolate (Re) from INT_M (Re), since the oxygen on the OH group of INT_M (Si) becomes axial in the enolate (Si). This change in conformation leads to lower energy.

The activation barrier of the TS_{RC1} (Re) ($\Delta G^{\ddagger}$ = 4.9 kcal/mol) is 2.1 kcal/mol higher in energy than that of the TS_{RC2} (Si) ($\Delta G^{\ddagger}$ = 2.8 kcal/mol) (Figure 31).



INT_{RC1} (Re) INT_{RC1} (Si)

Figure 32 Intermediate structures for the Re and the Si face for the first ring contraction step (RC1) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.



TS_{RC2} (Re) TS_{RC2} (Si)

Figure 33 TS structures for the Re and the Si face for the second ring contraction step (RC2) in pathway **C**, optimized at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

In the structure of TS_{RC1} (Re), the six-member ring where the first step of ring contraction happens is in a much more twisted shape than that in the TS_{RC1} (Si), also due to the alignment of the two rings.

The INT_{RC1} (Re) (-12.5 kcal/mol) is 6.6 kcal/mol higher in energy than the INT_{RC1} (Si) (-19.1 cal/mol) (Figure 29). Similarly, the structure of INT_{RC1} (Re) is much more twisted than INT_{RC1} (Si).

The INT_{RC1} continues with the ring opening step (Figure 33). The activation barrier of the TS_{RC2} (Re) ($\Delta G^{\ddagger}$ = 6.9 kcal/mol) is 1.3 kcal/mol higher than the TS_{RC2} (Si) ($\Delta G^{\ddagger}$ = 5.6 kcal/mol). Similar to the RC2 step in pathway **B**, TS_{RC2} has a higher barrier, which results in more steric hindrance and distortion. Then the final product is formed, with the Re product in higher energy.

Table 2 Energy for each pathway, in Re or Si. For pathway **A** and **B**, Si attack is preferred, while Re attack is preferred in pathway **C**.

Pathway/face	energy span(kcal/mol)
A (Re)	15.8
A (Si)	15.6
B (Re)	15.6
B (Si)	15.2
C (Re)	15.0
C (Si)	17.6

3.2.5 All 3 pathways

Within each pathway, there are two possible faces to attack during the oxa-Michael step, Re and Si. For ease of analysis, and only the face with the lower barrier in the rate-determining step is shown here (Figure 34). From Figure 34, we can see that both pathway **A** and **B** prefer the Si face attack; pathway **C** prefers the Re face attack, which leads to compound **I**. The activation barriers in the rate-determining steps are close to each other: pathway **A**, 15.6 kcal/mol (oxa-Michael step); pathway **B**, 15.2 kcal/mol (oxa-Michael step); pathway **C**, 15.0 kcal/mol (oxa-Michael step). For all three pathways, the oxa-Michael step is rate-determining, and they are very close in energy (difference within 0.6 kcal/mol). Since our reaction is multi-step and happens in a domino sequence, it is necessary to analyze the energy profile using an approach that accounts for this.

Instead of only considering activation barriers in the rate-determining step for each pathway, the energetic span approximation is applied to the DFT-calculated profiles because a) our reaction can be viewed potentially as a catalytic cycle, where the anion of the final product, **1** is deprotonated by the neutral starting material, **ent-3**, b) the oxa-Michael and Michael addition reactions are highly reversible, and the energy flow of multiple steps can be addressed with this approximation.

In pathway **A**, for the Si face, the energy span is 15.6 kcal/mol (Table 2). In pathway **B**, the energy span of the Si face is 15.2 kcal/mol. Pathway **C**-Re has the lowest energy span, 15.0 kcal/mol. In pathway **A**, for both faces, the ring contraction has an activation barrier of 15.1 kcal/mol, close to the energy span of both Re and Si.

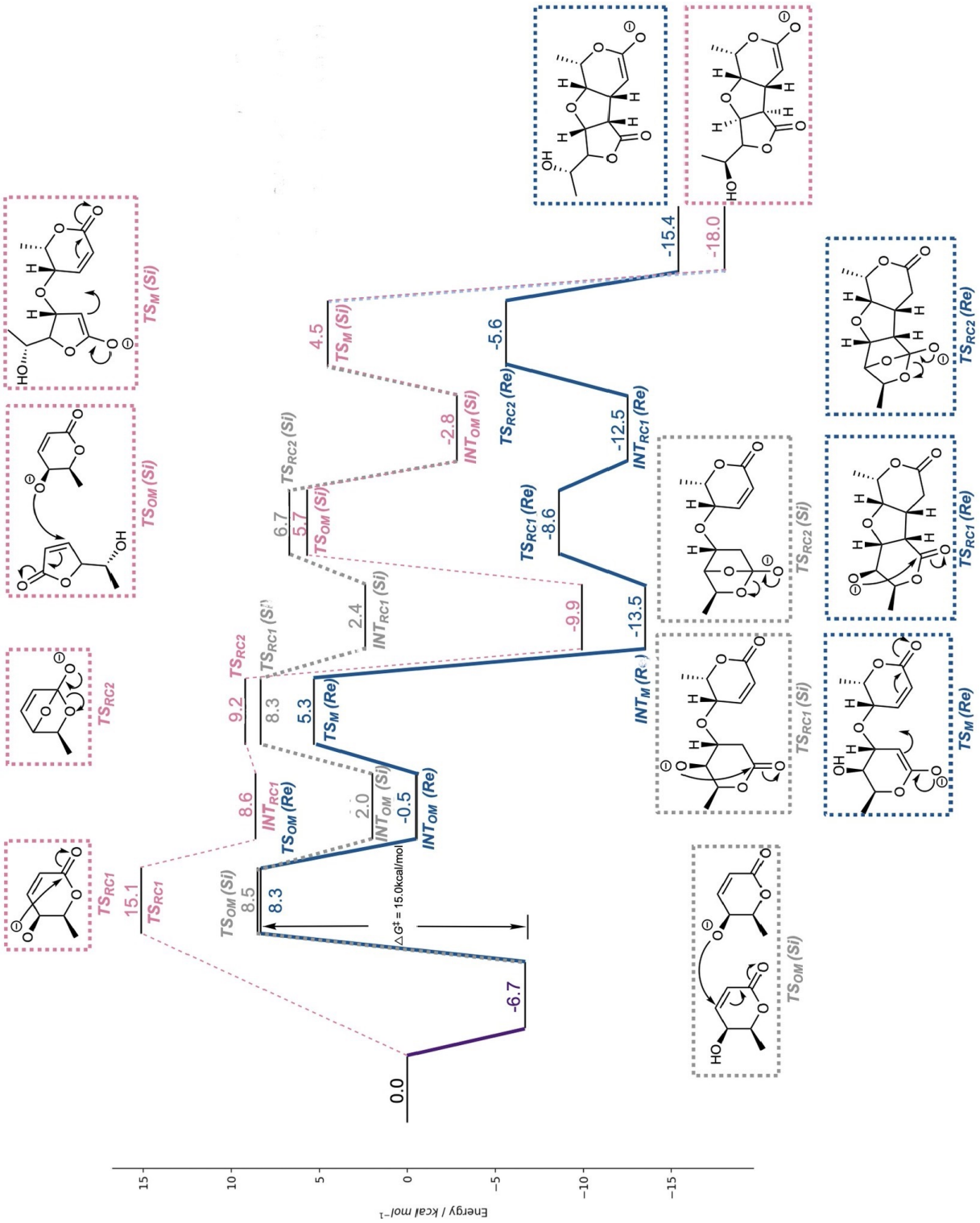


Figure 34 Free energy profile for the pathway with lower energy span in pathway **A**, **B**, and **C**, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP. Pathway **A** is represented by pink dashes, pathway **B** by gray dashes, and pathway **C** by blue lines.

δE, which only considers intermediates and TSs in the oxa-Michael and Michael step, does not involve states in this ring contraction step. If the activation barrier of another step, which does not contain the summit or trough of the cycle, is close to δE, the energy approximation would be much less effective. An analogy would be a pathway with two rate-determining steps of high barriers compared to that with only one. This would make pathway **A** much less preferable than **B** and **C**. However, the difference of δE between **B**-Si and **C**-Re is still too small (0.2 kcal/mol) for us to differentiate them.

3.3 Analysis on electronic structures and noncovalent interactions

3.3.1 Comparison of structures and electronic energies of different conformations

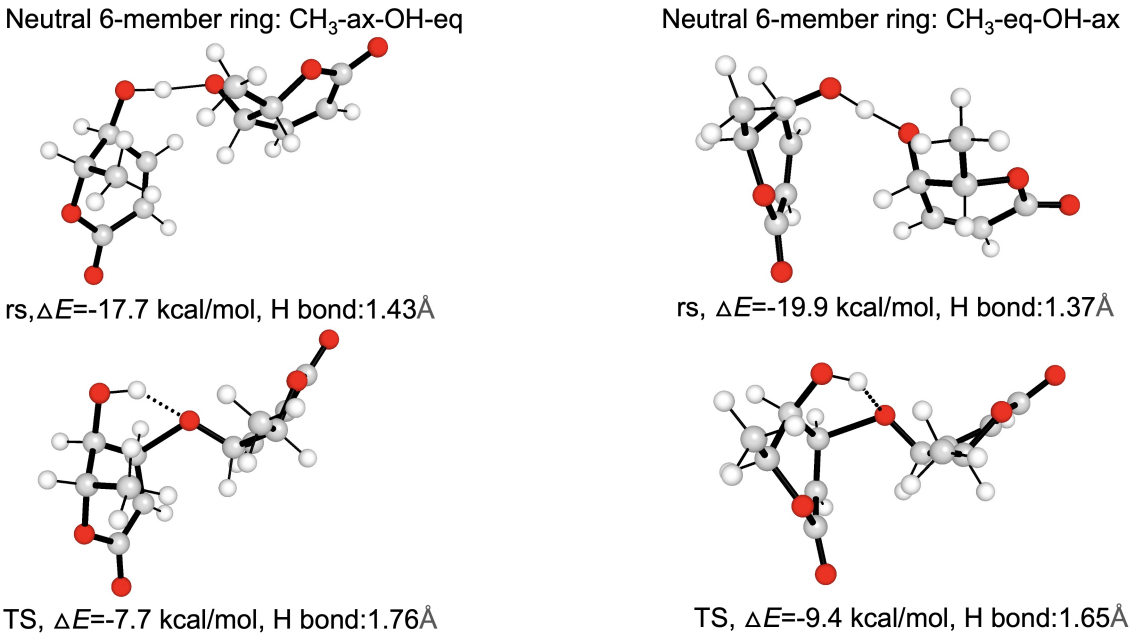


Figure 35 Geometries of reactant complex and TS of pathway **B**/**C**, Si with OH group equatorial and OH group axial for the neutral 6-member ring of δ-lactone, with electronic energy with respect to the two separate reactants, and the length of hydrogen bonding.

We are dealing with compounds in the solvent under the reaction conditions, and their conformations can freely change in the solution. This would lead to differences up to several kcal/mols in the energy profiles, especially for structures containing six-member rings of lactones with axial and equatorial groups, so it is necessary to check the energies and structures of different conformations to find the conformer with the lowest energy. We did a conformational search for these structures first, and the ones with the lowest energy were kept and optimized by DFT.

Conformational analysis was conducted based on DFT-optimised structures. The alcohol group on the neutral δ-lactone **ent-3** can be equatorial and axial, both forming hydrogen bonding with the anion (Figure 35). For both reactants and TS, the one

with the axial alcohol is lower in terms of electronic energy (by 2.1 kcal/mol and 1.7 kcal/mol), and the length of hydrogen bonding is shorter. In terms of free energy, the difference between the two configurations of reactants is 3.1 kcal/mol, and for TS it is 0.2 kcal/mol. The axial configuration is always lower in energy.

We conclude that the conformation with the methyl group being equatorial and the alcohol group being axial on the neutral six-member ring has a lower energy than the other conformation. By comparing the two different conformations of the reactant complex and TS, the conformation with lower energy was kept to compute the whole mechanistic pathway.

3.3.2 NBO analysis

Table 3 Second order perturbation analysis from NBO of geometric structures with Hydrogen bonding, at the level of ωB97X-D3/def2-TZVP.

Step/Structure	Hydrogen bonding contribution(kcal/mol)
Pathway **A** rs (Re) (complex)	76.99
Pathway **A**, TS_{OM} (Re)	34.06
Pathway **B/C**, rs (Si) (complex,OH-ax)	113.21
Pathway **B/C**, TS_{OM} (Si) (OH-ax)	28.13
Pathway **B/C**, rs (Si) (complex, OH-eq)	89.67
Pathway **B/C**, TS_{OM} (Si) (OH-eq)	14.51

After discussing the hydrogen bonding in different conformations for the oxa-Michael step in pathway **B/C**, we continue to check the strength of the hydrogen bonding and if it facilitates the reaction. For each pair of reactants and TS where noncovalent interaction exists, the strength of hydrogen bonding is quantified by the second-order perturbation analysis (Table 3). For each pair, the hydrogen bonding contribution is higher for reactants than for TS. So the hydrogen bonding results in facilitating the reaction.

3.3.3 Distortion Interaction analysis (DIA)

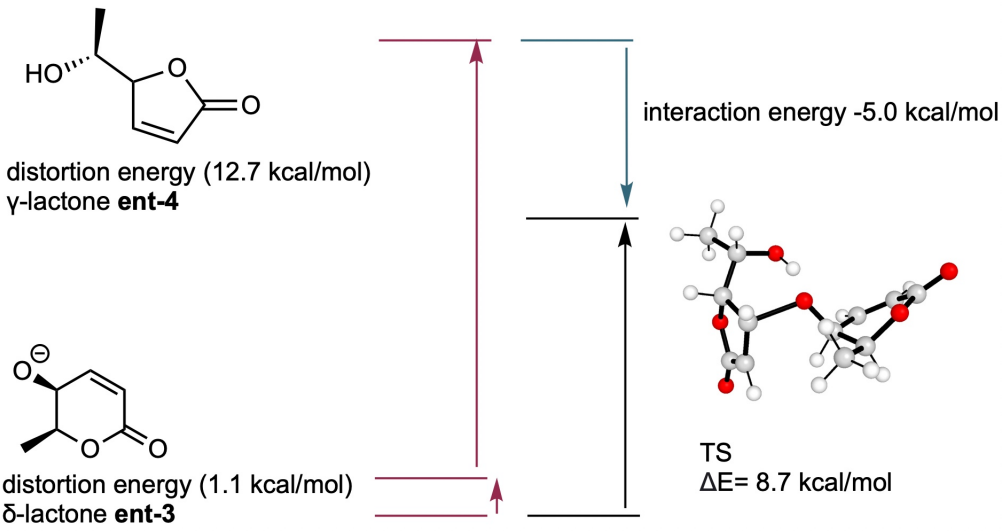


Figure 36 Distortion interaction analysis of pathway **A**, TS_{OM} (Re), at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

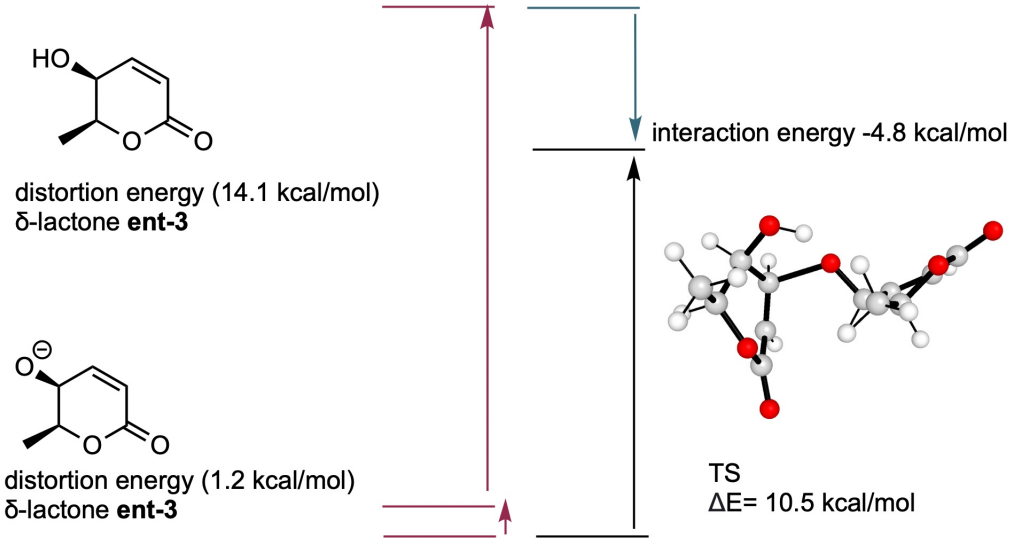


Figure 37 Distortion interaction analysis of pathway **B**/**C**, TS_{OM} (Si), with OH group axial for the neutral 6-member ring of δ-lactone at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

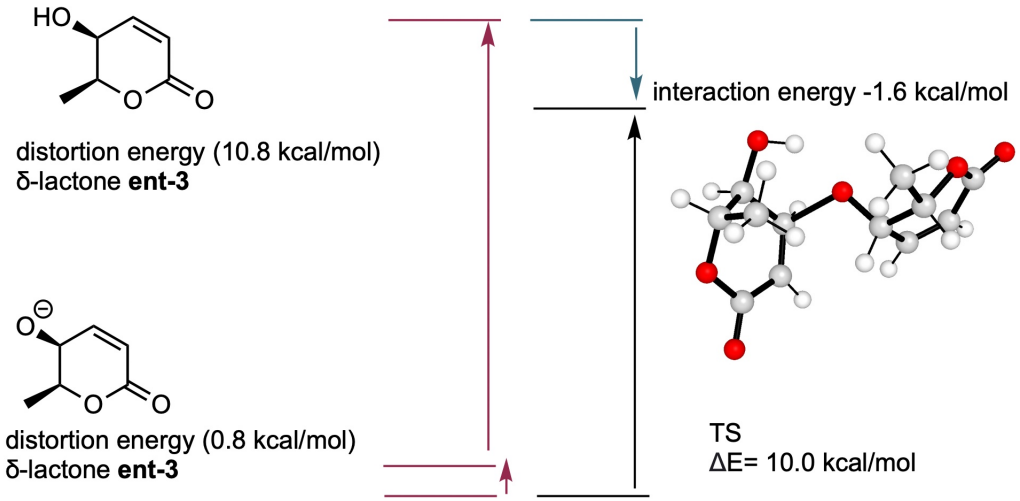


Figure 38 Distortion interaction analysis of pathway **B**/**C**, TS_{OM} (Si), with OH group equatorial for the neutral 6-member ring of δ-lactone at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Following the NBO analysis, distortion-interaction analysis was done for all the oxa-Michael reactions where hydrogen bondings are involved (Figure 36-38). For the oxa-Michael between the γ-lactone **ent-4** and δ-lactone **ent-3** (Figure 36), and between the neutral **ent-3** (axial) and **ent-3** anion (Figure 37), the interaction energy is more than 30 percent of the electronic activation barrier. For the oxa-Michael between δ-lactone **ent-3** (equatorial) and its anion (38), this is less than 14 percent. In terms of hydrogen bonding contribution (Table 1), the former two are also both around or more than two times higher than the latter, indicating that the interaction comes from hydrogen bonding.

3.4 The microkinetic model

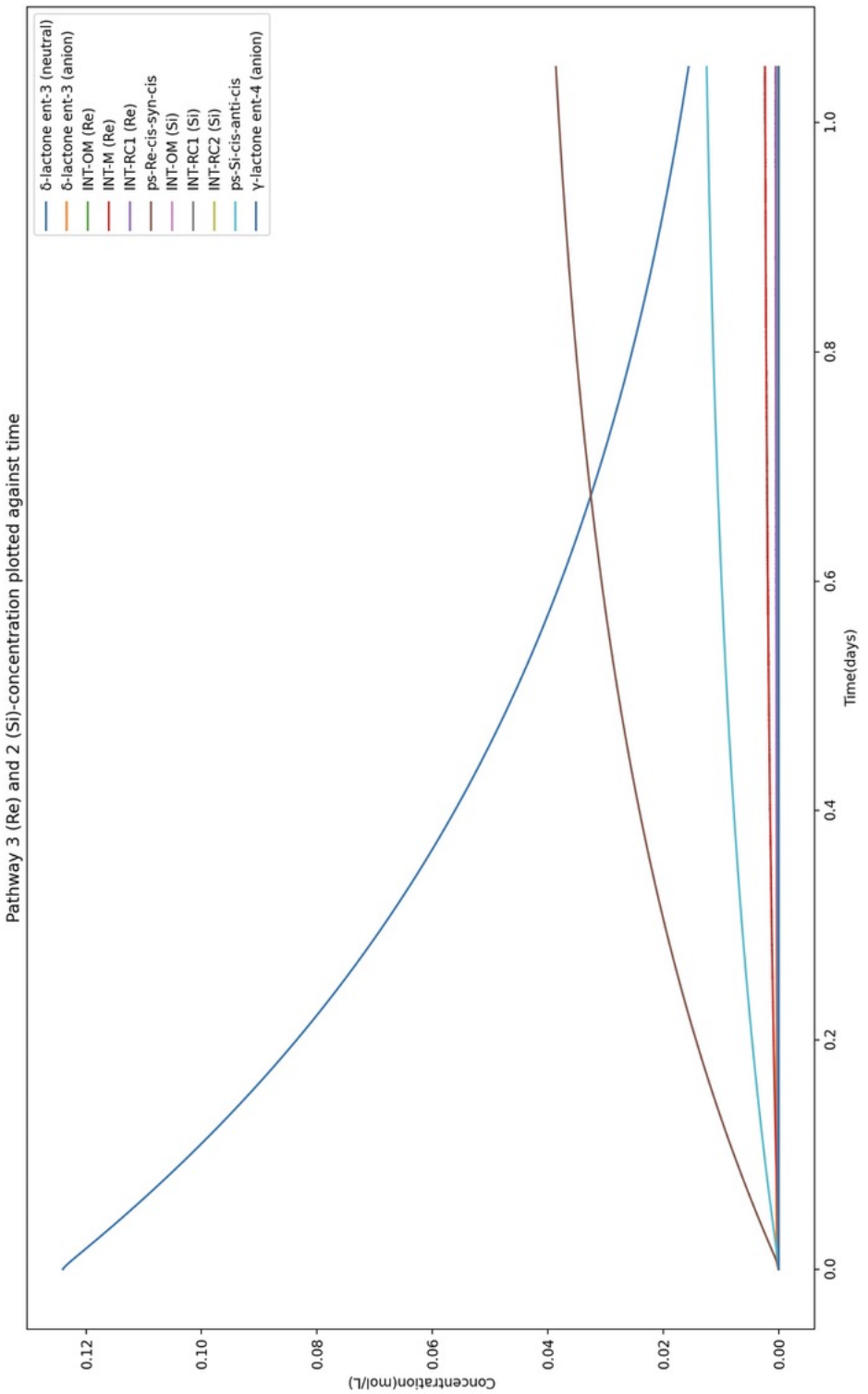


Figure 39 The microkinetic model applied for the two pathways (**B**-Si, **C**-Re) with the lowest energy span approximation, also including the ring contraction step of pathway **A**

Since the most favorable pathway cannot be determined by analyzing DFT-calculated energy diagrams, other types of analysis are required to answer our question. To take into account concentration effects, a microkinetic model is used to study the domino reaction sequence (Figure 39). Here, we attempt to simulate the formation of intermediates/products in all three pathways, but the last two steps (the oxa-Michael (OM) step and the Michael (M) step) in pathway **A** are not simulated because a) the experiments of our collaborators showed that these two steps never took place, with the two reactants (**3** and **4**) mixed resulting in no intermediates/products b) including these two steps would significantly increase the computational cost. Therefore, for pathway **B** and **C**, all three steps for the faces attacked with lower energetic span in the other two pathways (pathway **B**-Si and **C**-Re) are included in the microkinetic model, while for pathway **A**, only the first step (ring contraction) is included.

The domino sequence starts with only one molecule, δ-lactone **ent-3**. The microkinetic model shows that the concentration of **3** keeps decreasing during the reaction, while those of the intermediates/final products keep increasing. Among all the intermediates/final products formed in the reaction, the concentration of the final product, ps-Re-cis-syn-cis, is predominantly high, matching the results of our NMR calculations. This is the product conducive to pathway **C**-Re, thus leading to the conclusion that pathway **C** is the most favorable pathway. Under another experimental condition (tetra-n-butylammonium fluoride (TBAF), tetrahydrofuran (THF), room temperature), INT_M (Re) from pathway C was observed. The microkinetic model shows that a small amount of INT_M is formed in the reaction, with a concentration less than 0.01 mol/L, also matching the experimental result. The amount of ps-Si-cis-anti-cis formed is around a third of that of ps-Re-cis-syn-cis, but from the NMR calculations, we can only conclude that the Re-cis-syn-cis product is most likely to be the one formed in the experiment.

In the experiment, a 5% yield of γ-lactone **ent-4** was observed, while the concentration of **4** keeps being close to zero, with a magnitude of 10^{-9}. The experimental result shows the ring contraction step in pathway **A** happens, while the microkinetic model fails to do so.

Other intermediates were not observed in the experiment. The microkinetic model also does not show an observable concentration of other intermediates or final products, corresponding to the experimental results. Therefore, by running the microkinetic model for our reaction, we conclude that the Re face attack of pathway **C** is the most favorable pathway.

4 Structure benchmarking and structure elucidation using NMR

4.1 Benchmarking of geometries

To validate the DFT level of theory we used the mechanistic pathways proposed above. Benchmarking of geometries was performed for the structures (reactants, intermediates, products, and TSs) in one of the proposed pathways (pathway **A**) using RI-SCS-MP2/def2-TZVPP. There are three steps in pathway **A**: ring contractions, oxa-Michael, and Michael.

The differences in geometries (in terms of bond length and root mean square deviation (RMSD)) of the intermediates, TSs, and the final product in these steps are compared with lower-level theories. We compared two levels of theory/basis sets that were chosen: PBE0/def2-SVP and ωB97X-D3/def2-TZVP. For the ωB97X-D3 method, larger basis sets such as def2-QZVP were also tested, but the difference between the two is small enough to be ignored. For PBE0/def-SVP, a larger basis set, def2-TZVP was also tested, but still, the difference is small enough to be ignored.

The geometry optimization for the TS in the cyclization in the ring contraction (TS_{RC2}) failed at the level of PBE0/def2-SVP, probably because the barrier of this step is particularly low (0.5 kcal/mol). Therefore, we continued to validate the level of computation at ωB97X-D3/def2-TZVP.

Table 4 RMSD of hydrogen-removed geometries computed at the level of SMD(1,2-DICHLOROETHANE)-RI-SCS-MP2/def2-TZVPP and SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP in Pathway **A**.

Step/Structure	RMSD(Å)
RC1,TS_{RC1}	0.01
RC1, INT_{RC1}	0.01
RC2, TS_{RC2}	0.02
OM, TS_{OM} (Re)	0.03
OM, TS_{OM} (Si)	0.19
M, TS_{M} (Re)	0.06
M, TS_{M} (Si)	0.07
M, ps-Re-cis-syn-cis	0.20
M, ps-Si-cis-anti-cis	0.07

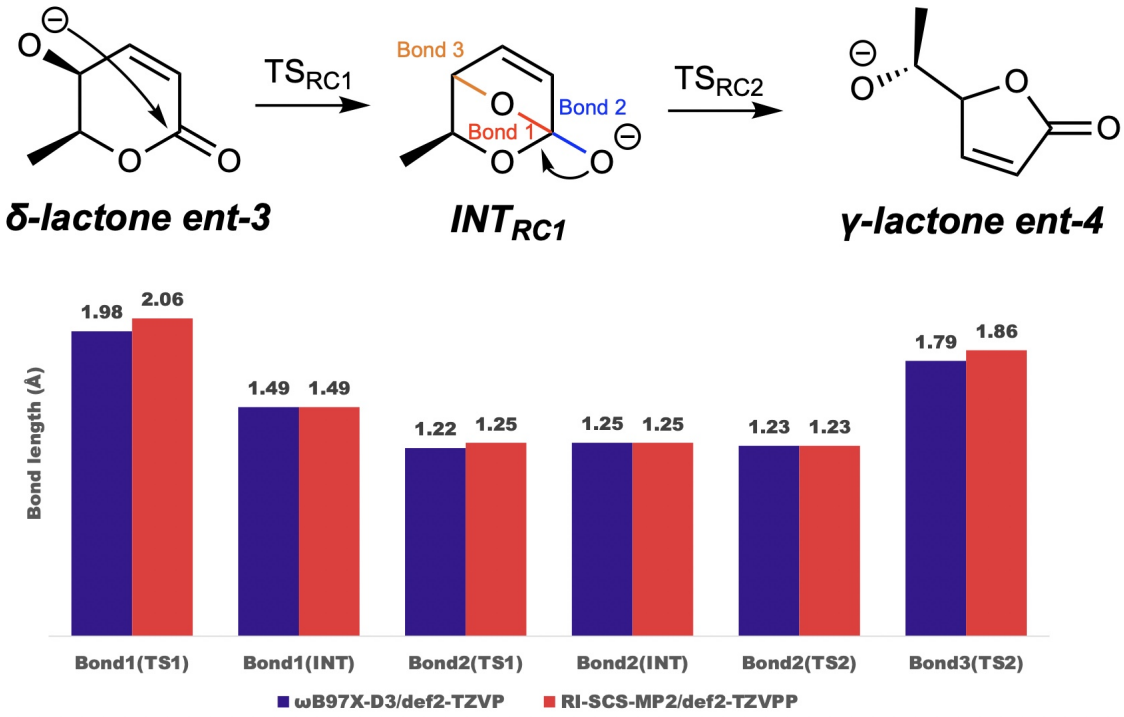
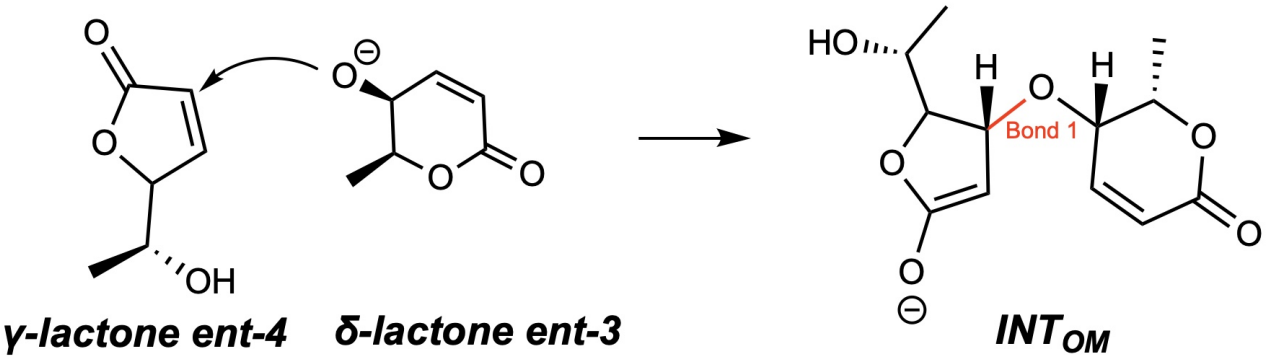


Figure 40 Detailed mechanism of **A1** and bond length comparison at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP and RI-SCS-MP2/def2-TZVPP.

Comparing the root mean square deviation (RMSD) for the different geometries in pathway **A** (Table 4), we can see that the largest RMSD is 0.20 Å(ps-Re-cis-syn-cis) and 0.19 Å(TS_{OM} (Si)), and for all the other geometries, RMSD is less than or equal to 0.07 Å. The differences between the bond lengths of the geometries at these two computational levels are all within 0.1Å.

Geometries. The most important bonds in our mechanistic pathways in all steps are the forming bonds of that specific step. For all steps, the length of forming bonds and other bonds relevant bonds or distances (for instance, Bond 2 is involved in ring contraction in pathway **A**, and Bond H, a hydrogen bonding, in the oxa-Michael step, facilitates the reaction) calculated at the level of ωB97X-D3/def2-TZVP and RI-SCS-MP2/def2-TZVPP are compared together.

Pathway A, step 1 (A1): ring contraction. (Figure 40, Table 4) The differences in bond length are small for all the bonds we compare (within 0.07 Å) and RMSD, so the two levels of theory agree well with each other.



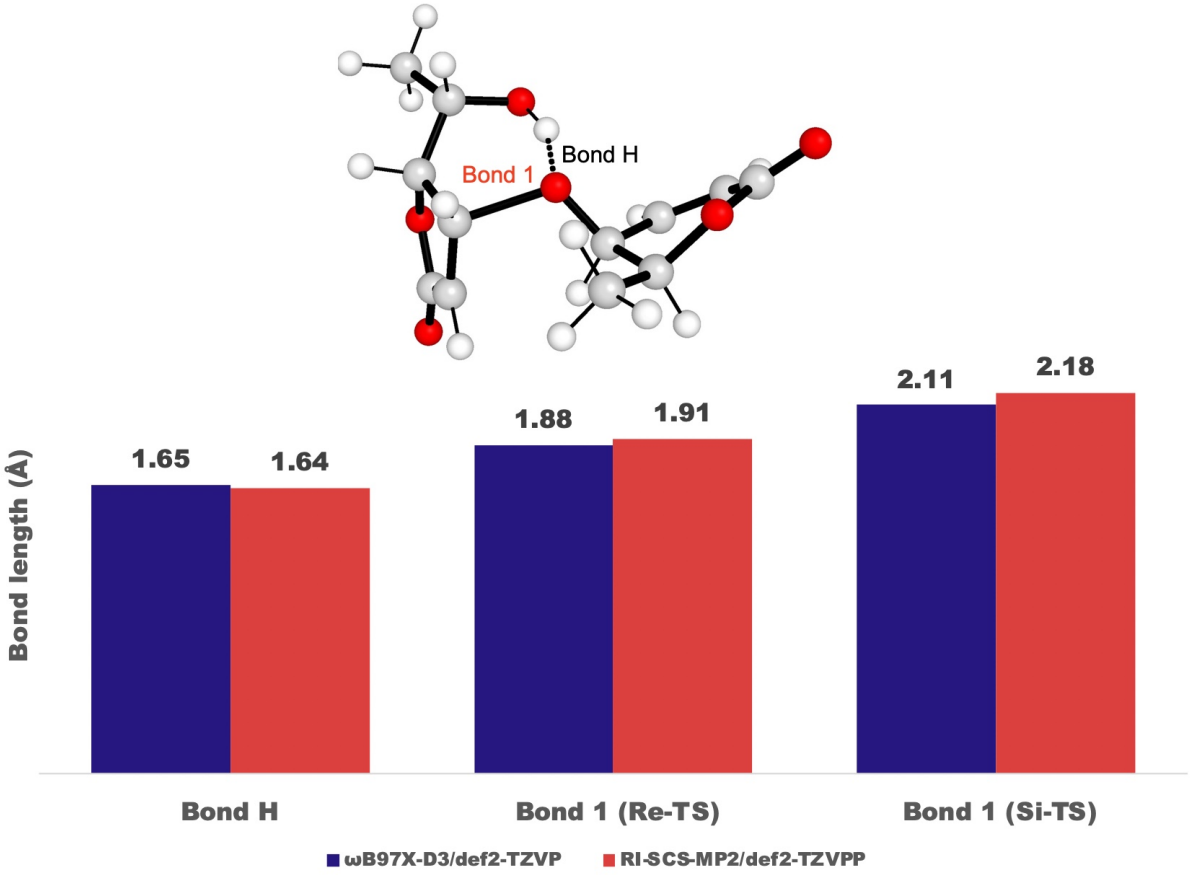


Figure 41 Detailed mechanism of **A2** and bond length comparison at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP and SMD(1,2-DICHLOROETHANE)-RI-SCS-MP2/def2-TZVPP.

Pathway A, step 2 (A2): oxa-Michael. (Figure 41, Table 4) Same as in step 1, the difference of bond length is with 0.07 Å, but the RMSD of TS_{OM} (Si) is larger than 0.1 Å. This is because the C-O next to the double bond on the five-member is freely rotatable, and the overlapping of this bond is not very good.

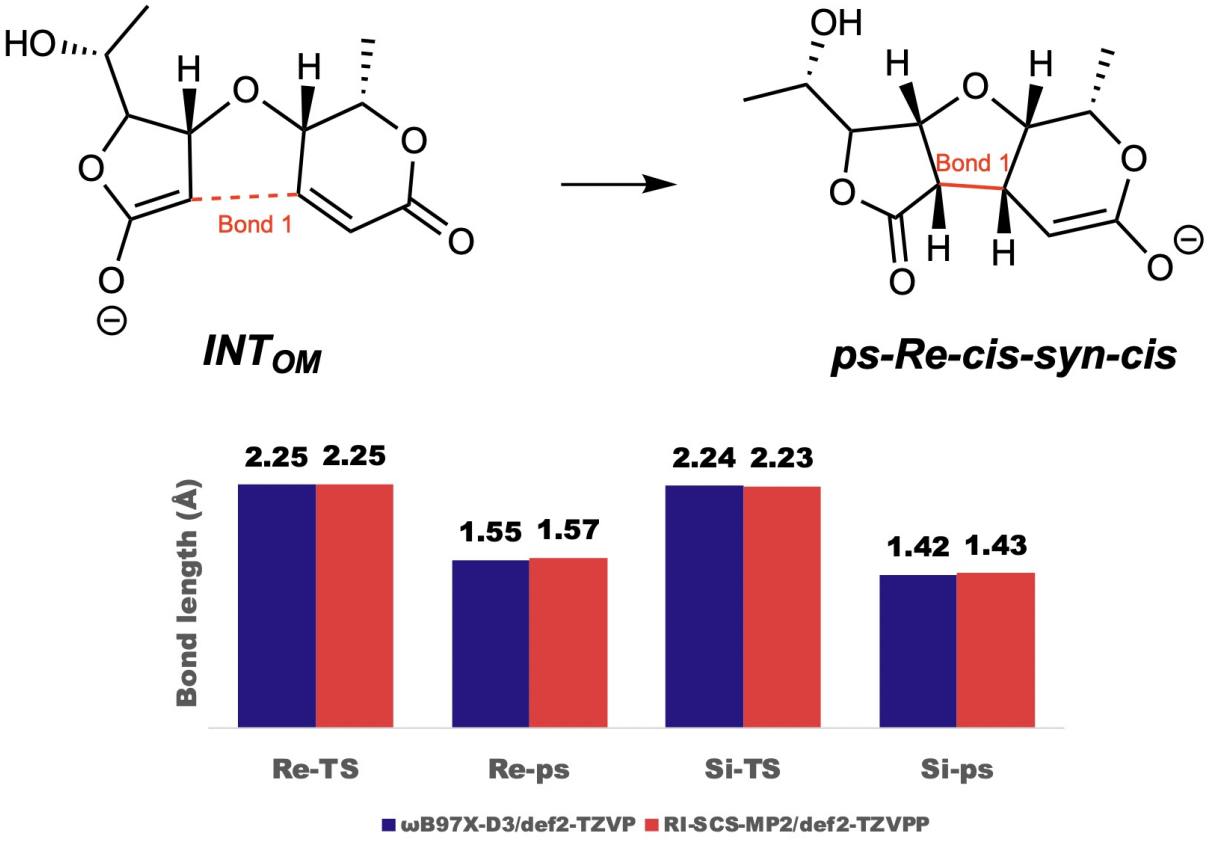


Figure 42 Detailed mechanism of **A3** and bond length comparison at the level of SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP and SMD(1,2-DICHLOROETHANE)-RI-SCS-MP2/def2-TZVPP.

Pathway A, step 3 (A3): Michael. (Figure 42, Table 4) The bond length difference in this step is particularly small (with 0.02 Å), and only for the Re-cis-syn-cis product the RMSD is 0.02 Å, for the same reason of the relatively large RMSD in step 2.

Electron density difference analysis. (Figure 43, Figure 44) Alongside the geometries, we also compared the electron densities obtained by these computational methods, using the same approach as in previous works using DFT.[92] Although most of the RMSD is small (less or equal to 0.07Å), there are still geometries computed at the two different levels of theory (ωB97X-D3/def2-TZVP and RI-SCS-MP2/def2-TZVPP) for which the overlapping is not perfect. The density differences of benchmarked structures in Pathway **A** at the level of ωB97X-D3/def2-TZVP are compared to linear and unrelaxed densities computed at the level of DLPNO-CCSD/def2-TZVPP since the same type of density is not available for CCSD(T). We also compare the DLNPO-CCSD density with that of RI-SCS-MP2/def2-TZVPP, the method of our benchmarking reference.

We first look at the density difference between ωB97X-D3/def2-TZVP and CCSD/def2-TZVPP. From the lobes of in Figure 43 and 44, we can see that the magnitude of the density difference in different structures is close to each other, except the relatively small difference of TS-Re in step 2 and step 3. For all these structures, the difference in densities is good enough for benchmarking (within 0.30). The RI-SCS-MP2 density appeared to have very little difference with respect to DLPNO-CCSD (within 0.001), for which the isovalue cutoff was set at 0.0001 to make the density difference visible. In fact, they are identical to each other, and different due to numerical errors.

Overall, ωB97X-D3/def2-TZVP agrees with RI-SCS-MP2/def2-TZVPP reasonably well: the differences in geometry (bond length and RMSD) are mostly small (within 0.1 Å), and the electron density analysis further illustrates that the difference in density is acceptable (within 0.30), therefore validating ωB97X-D3/def2-TZVP.

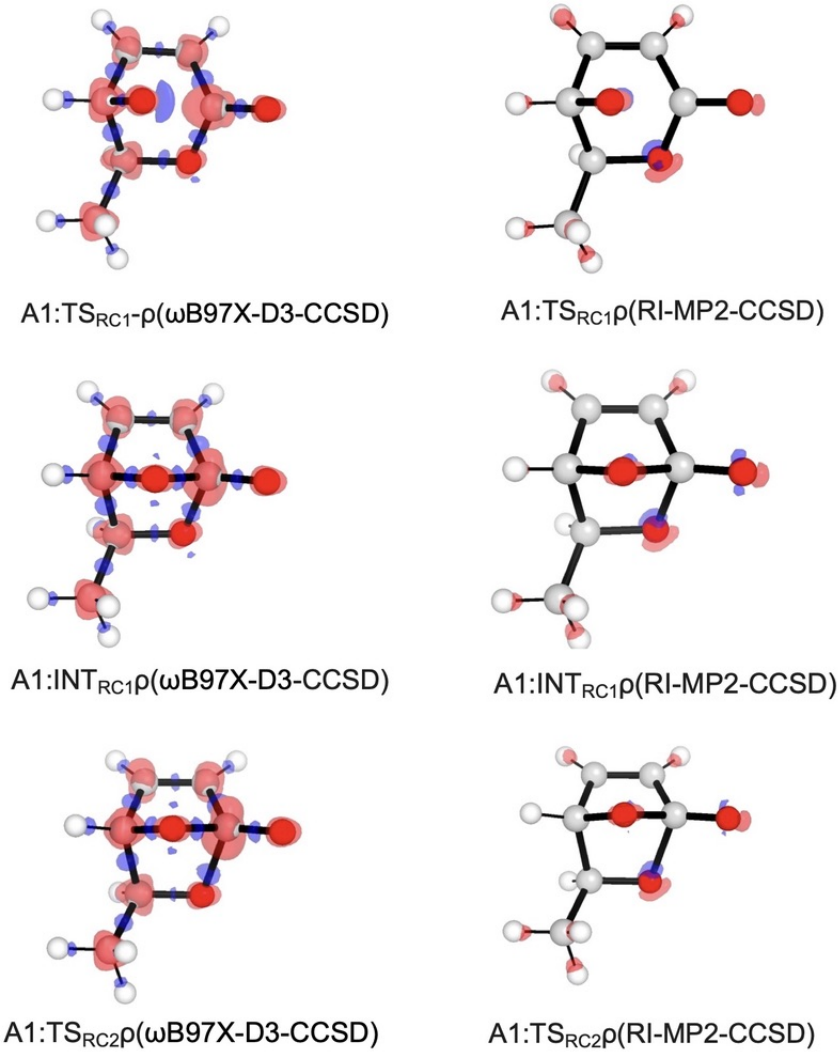


Figure 43 Visualized electronic density difference between linearized and DLPNO-CCSD/def2-TZVPP densities for the ring contraction step, with optimised geometry at the level of ωB97X-D3/def2-TZVP, with an integration grid of 120. For the density difference between ωB97X-D3/def2-TZVP and DLPNO-CCSD/def2-TZVPP, the isovalue cutoff is ±0.01 electrons/Å^3. For density difference between RI-SCS-MP2/def2-TZVPP and DLPNO-CCSD/def2-TZVPP, the isovalue cutoff is ±0.0001 electrons/Å^3.

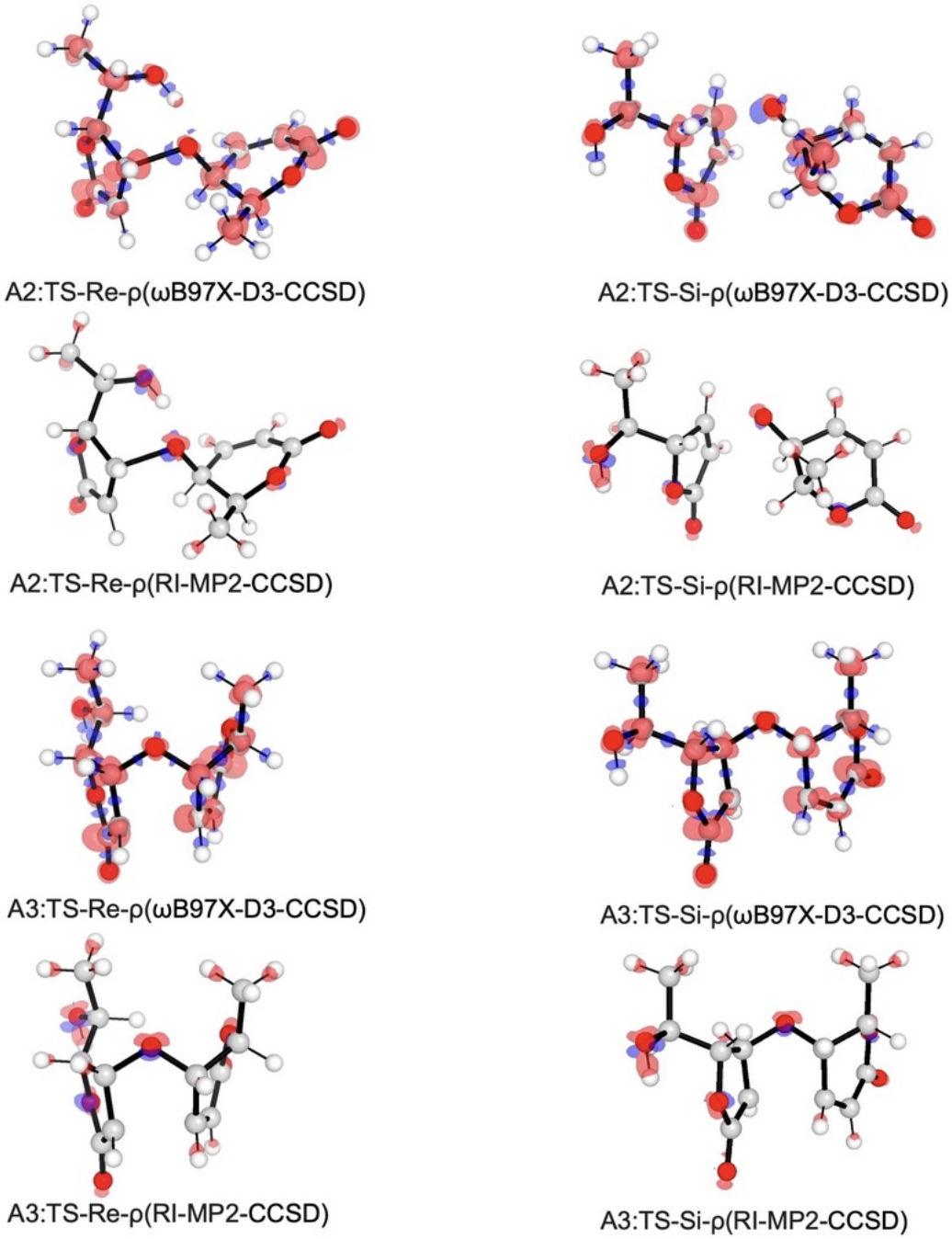


Figure 44 Visualized electronic density difference between linearized and DLPNO-CCSD/def2-TZVPP densities for the oxa-Michael and Michael step, with optimised geometry at the level of ωB97X-D3/def2-TZVP, with an integration grid of 120. For the density difference between ωB97X-D3/def2-TZVP and DLPNO-CCSD/def2-TZVPP, the isovalue cutoff is ±0.01 electrons/Å^3. For density difference between RI-SCS-MP2/def2-TZVPP and DLPNO-CCSD/def2-TZVPP, the isovalue cutoff is ±0.0001 electrons/Å^3.

4.2 NMR

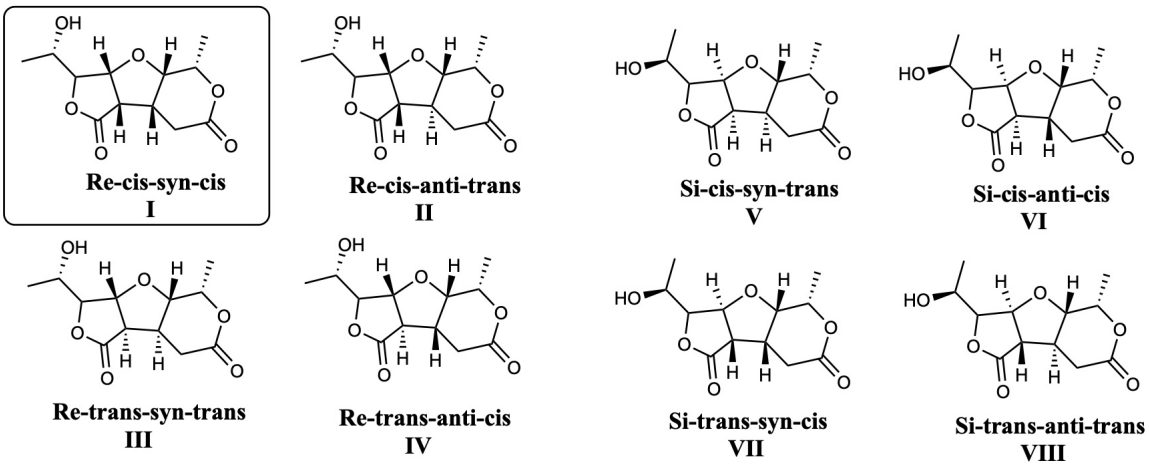


Figure 45 Eight possible diastereoisomers of (-)-Angiopterlactone B (**2**).

Table 5 ^{1}H and ^{13}C MAE values of NMR chemical shifts (-)-angiopterlactone B (**2**) (compound **I**). MAE values are calculated by difference between the computed chemical shift of **I** relevant to the values of the previously benchmarked DSD-PBEP86 method (DSD-PBEP86-D3BJ/def2-QZVP).[85–87]

Level of theory	^{13}C MAE / ppm	^{1}H MAE / ppm
ωB97X-D3/def-TZVP	2.1	0.17
RI-B2PLYP/def2-QZVP	6.0	0.09
mPW1PW91/def2-TZVP	6.1	0.16

Table 6 ^{1}H, ^{13}C, and $^{3}J_{H\text{-}H}$ MAE values of NMR chemical shifts (-)-angiopterlactone B (**2**) (compound **I**) and seven of its diastereoisomers (**II-VIII**). MAE values are calculated by difference between the computed chemical shift of the eight possible isomers (**I**-**VIII**) relevant to the experimental shift of the product in chloroform.

	I	II	III	IV	V	VI	VII	VIII
^{1}H MAE / ppm	0.10	0.25	0.22	0.24	0.27	0.24	0.28	0.34
^{13}C MAE / ppm	5.0	6.4	5.6	5.5	5.8	7.3	6.3	5.7
$^{3}J_{H\text{-}H}$ MAE / Hz	1.0	1.8	5.8	2.1	1.9	6.4	3.1	2.0

The optimized geometries of the eight possible diastereoisomers are used to calculate NMR chemical shifts and coupling constants and then compared with the experimental data. The double-hybrid method is widely considered accurate, and the DSD-PBEP86 method was the most accurate one according to benchmarking

done previously.[85] Using this method as a reference, different levels of theory were tested. For the ^{13}C chemical shifts, ωB97X-D3/def-TZVP has the smallest MAE (2.1 ppm). For the ^{1}H chemical shift, RI-B2PLYP/def2-QZVP has the smallest MAE (0.1 ppm) (Table 5). However, since ωB97X-D3/def-TZVP is computationally much more efficient and the error of 0.17 is still acceptable, ωB97X-D3/def-TZVP was chosen to perform NMR calculations.

To confirm the configuration of the final product, mean absolute errors (MAE) are calculated for the NMR chemical shifts of all the eight possible diastereoisomers of the final product (Figure 45, Table 6), (-)-angiopterlactone B (**2**), comparing the differences between computed values and experimental values. Compound **I** has the smallest MAE in all three categories: 0.010 ppm, 5.0 ppm for ^{13}C and 1.0 Hz for $^{3}J_{H\text{-}H}$. Compound **VIII** has the largest MAE for ^{13}C (0.34 ppm), compound **VI** 7.3 ppm for ^{13}C and 6.4 Hz for $^{3}J_{H\text{-}H}$. For the MAEs of ^{1}H, ^{13}C, and $^{3}J_{H\text{-}H}$, compound **I**, has the smallest values among all the diastereoisomers, therefore confirming that the final experimental product is diastereoisomer **I** from the NMR calculations.

5 Conclusions

5.1 Computational results and analysis

In previous works, a total synthesis of (-)-angiopterlactone B (**2**) was carried out by our collaborator, but the mechanistic pathway of this reaction was still unclear merely through the experimental observations. The experimentalists were only able to isolate and identify the final product, the angiopterlactone, from NMR spectra, without any intermediates that could potentially help us deduce a pathway.

Originally, a domino sequence of ring-contraction/oxa-Michael/Michael reactions was proposed, with the biosynthetic speculation that angiopterlactone A (**1**) is the precursor of angiopterlactone B (**2**) in the mechanistic pathway. Following this speculation, another two pathways were proposed, changing the order of the domino sequence. Altogether we have three different pathways: pathway **A** (ring-contraction/oxa-Michael/Michael sequence), pathway **B** (oxa-Michael/ring-contraction/Michael sequence), pathway **C** (oxa-Michael/Michael/ring-contraction sequence).

Since the experimental results alone could not provide sufficient evidence on which pathway took place, computational chemistry was utilized to answer this question. For our system, we choose to use DFT methods to optimize the geometries and compute NMR data.

The experimental NMR chemical shifts and coupling constants were first compared with the computed values of the eight possible diastereoisomers of angiopterlactone B (**2**). We find out that the Re-cis-syn-cis product (compound **I** in Figure 44) has the smallest MAE among the eight, so we conclude that the final product has such a configuration.

After optimizing all the geometries of the structures (reactants, intermediates, TSs, and products) in each pathway, an energy profile was depicted for each of them. For each pathway, the prochirality is considered in the oxa-Michael step, so in each pathway, there are two possibilities of face attack, Re/Si. Comparing the face

attacked with a lower energy span for each pathway, we can see that pathway **A** and **B** lead to the Si-cis-anti-cis product (compound **VI**), and only pathway **C** leads to the Re-cis-syn-cis product (compound **1**).

By comparing the energy span for each, we observe no decisive difference in ΔE that can tell us which pathway is the most favorable. In pathway **A**, for the Si face, the energy span is 15.6 kcal/mol (Figure 5). In pathway **B**, the energy span of the Si face is 15.2 kcal/mol. Pathway **C**-Re has the lowest energy span, 15.0 kcal/mol. Although pathway **C**-Re has the lowest energy span, the difference between it and the other two pathways is within 1 kcal/mol. Especially for pathway **B**-Si, the difference is only 0.2 kcal/mol.

In pathway **A**, for both faces, the ring contraction has an activation barrier of 15.1 kcal/mol, close to the energy span of both Re and Si. The energy span does not apply very well here, as the ring-contraction step ($\Delta G^{\ddagger}$ = 15.1 kcal/mol) has an activation barrier close to that of the oxa-Michael step ($\Delta G^{\ddagger}$ = 15.6 kcal/mol). So pathway **A** would be much less likely to happen under the reaction conditions, and this is proved by the experiment, where the two reactants of the oxa-Michael step were placed under the reaction conditions, and no products or INT_{OM} was observed.

Since the DFT calculations do not consider concentrations of the molecules involved in the pathways explicitly, the microkinetic model was applied. The initial conditions were input, and the reaction was simulated based on the rate constants obtained from the activation barriers calculated through DFT. Given that the pathway **A** does not take place under the reaction conditions, pathway **B**-Si and pathway **C**-Re were simulated. The result shows that pathway **C**-Re is predominant, which gives the Re-cis-syn-cis product, matching the results of our NMR calculations.

5.2 Outlook

For NMR prediction, though we use MAEs to compare the experimental NMR shifts and the computed ones, ideally the use of the DP4 analysis to identify the correct diastereoisomer would be preferable.[46,93] This method is currently not applicable to our computational method. Future work will explore the use of assigning probabilities to each diastereoisomer.

We also noticed that the cross-dimerization of **3** and its enantiomer failed, and so far there is no synthesis with enantioselectivity preferring the natural product, (+)-angiopterlactone B. Since our collaborator successfully synthesized (-)-angiopterlactone B using **3** as the starting material, we expect the enantiomer of **3** to undergo the same reaction to form (+)-angiopterlactone B, but the reaction does not happen experimentally. The three different pathways proposed by us before can be explored by DFT calculations, using the enantiomer of **3** as the starting material.

The rate constants in the microkinetic model are calculated based on the DFT energy profiles. Although the level of theory we use has been benchmarked, the energy might still have an error within 1 kcal/mol. We can vary the rate constants for the microkinetic model to see if the change of activation barriers has an effect on the proportion or yield of the final product, or the observed intermediates in the experiment.

Appendices

A Appendix: Absolute energies

Table 7 Absolute energy for pathway **A**, step 1, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step1 (Ring contraction)	E(hartree)	H(hartree)	G(hartree)
6-member anion	-457.89388	-457.75957	-457.80592
TS1	-457.87066	-457.73836	-457.78180
INT	-457.88120	-457.74782	-457.79214
TS2	-457.88018	-457.74789	-457.79126
5-member anion	-457.89553	-457.76178	-457.80842

Table 8 Absolute energy for pathway **A**, step 2, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step2 (oxa-Michael)	E(hartree)	H(hartree)	G(hartree)
3 anion	-457.89388	-457.75957	-457.80592
4 neutral	-458.42397	-458.27535	-458.32296
Re-rs	-916.34727	-916.06245	-916.13697
Si-rs(ch3-eq)	-916.32374	-916.04358	-916.10482
Re-TS(ch3-eq-hb)	-916.33343	-916.04854	-916.11859
Si-TS(ch3-eq)	-916.32541	-916.04063	-916.11201
Si-TS(ch3-ax)	-916.32253	-916.03867	-916.11221
Re-ps	-916.33877	-916.05207	-916.12344
Si-ps	-916.33915	-916.05237	-916.12394

Table 9 Absolute energy for pathway **A**, step 3, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step3 (Michael)	E(hartree)	H(hartree)	G(hartree)
Re-rs	-916.33877	-916.05207	-916.12344
Si-rs	-916.33915	-916.05237	-916.12394
Re-cis-anti-trans-TS	-916.31045	-916.02490	-916.09432
Re-cis-syn-cis-TS	-916.32918	-916.04357	-916.11184
Re-trans-anti-cis-TS	-916.28726	-916.00205	-916.07043
Re-trans-syn-trans-TS	-916.26509	-915.97957	-916.04789
Si-cis-anti-cis-TS	-916.33042	-916.04483	-916.11398
Si-cis-syn-trans-TS	-916.31365	-916.02812	-916.09640
Si-trans-anti-trans-TS	-916.27361	-915.98824	-916.05697
Si-trans-syn-cis-TS	-916.29406	-916.00944	-916.07864
Re-cis-anti-trans-ps	-916.35057	-916.06342	-916.13010
Re-cis-syn-cis-ps	-916.36695	-916.07848	-916.14580
Re-trans-anti-cis-ps	-916.33336	-916.04479	-916.11321
Re-trans-syn-trans-ps	-916.30700	-916.01842	-916.08720
Si-cis-anti-cis-ps	-916.36691	-916.07905	-916.14995
Si-cis-syn-trans-ps	-916.35194	-916.06376	-916.13292
Si-trans-anti-trans-ps	-916.32250	-916.03421	-916.10290
Si-trans-syn-cis-ps	-916.34228	-916.05448	-916.12339

Table 10 Absolute energy for pathway **B**, step 1, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step1(oxa-Michael)	E(hatree)	H(hartree)	G(hartree)
6-member anion	-457.89388	-457.75957	-457.80592
5-member neutral	-458.42397	-458.27535	-458.32296
Re-rs	-916.34727	-916.06245	-916.13697
Si-rs(ch3-eq)	-916.32374	-916.04358	-916.10482
Re-TS(ch3-eq-hb)	-916.33343	-916.04854	-916.11859
Si-TS(ch3-eq)	-916.32541	-916.04063	-916.11201
Si-TS(ch3-ax)	-916.32253	-916.03867	-916.11221
Re-ps	-916.33877	-916.05207	-916.12344
Si-ps	-916.33915	-916.05237	-916.12394

Table 11 Absolute energy for pathway **B**, step 2, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step2(ring-contraction)	E(hatree)	H(hartree)	G(hartree)
Re-rs	-916.33324	-916.04658	-916.11792
Si-rs	-916.32823	-916.04169	-916.11268
Re-INT	-916.33351	-916.04656	-916.11510
Si-INT	-916.33452	-916.04795	-916.11748
Re-TS1	-916.32647	-916.04068	-916.10905
Si-TS1	-916.32488	-916.03935	-916.10807
Re-TS2	-916.32352	-916.03804	-916.10612
Si-TS2	-916.32727	-916.04187	-916.11063
Re-ps	-916.33877	-916.05207	-916.12344
Si-ps	-916.33915	-916.05237	-916.12394
Si-ps(oh-ax)	-916.33181	-916.04375	-916.11414
Re-cis-syn-cis-ps	-916.36695	-916.07848	-916.14580
Re-trans-anti-cis-ps	-916.33336	-916.04479	-916.11321
Re-trans-syn-trans-ps	-916.30700	-916.01842	-916.08720
Si-cis-anti-cis-ps	-916.36691	-916.07905	-916.14995
Si-cis-syn-trans-ps	-916.35194	-916.06376	-916.13292
Si-trans-anti-trans-ps	-916.32250	-916.03421	-916.10290
Si-trans-syn-cis-ps	-916.34228	-916.05448	-916.12339

Table 12 Absolute energy for pathway **C**, step 2, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step2 (Michael)	E(hartree)	H(hartree)	G(hartree)
Re-rs	-916.33818	-916.05118	-916.12217
Si-rs	-916.33542	-916.04759	-916.11808
Re-cis-anti-trans-TS	-916.30496	-916.01941	-916.08815
Re-cis-syn-cis-TS	-916.33057	-916.04464	-916.11282
Re-trans-anti-cis-TS	-916.29634	-916.01097	-916.07927
Re-trans-syn-trans-TS	-916.27402	-915.98864	-916.05717
Si-cis-anti-cis-TS(ch3-eq-oh-eq)	-916.32248	-916.03623	-916.10389
Si-cis-syn-trans-TS	-916.30346	-916.01709	-916.08474
Si-trans-anti-trans-TS	-916.28500	-915.99970	-916.06809
Si-trans-syn-cis-TS	-916.30302	-916.01793	-916.08702
Re-cis-anti-trans-ps	-916.34733	-916.05886	-916.12863
Re-cis-syn-cis-ps	-916.36120	-916.07274	-916.14051
Re-trans-anti-cis-ps	-916.34895	-916.06049	-916.12827
Re-trans-syn-trans-ps	-916.32687	-916.03830	-916.10673
Si-cis-anti-cis-ps(ch3-ax-oh-eq)	-916.35390	-916.06487	-916.13250
Si-cis-anti-cis-ps(ch3-eq-oh-eq)	-916.35536	-916.06613	-916.13346
Si-cis-syn-trans-ps	-916.34414	-916.05567	-916.12331
Si-trans-anti-trans-ps	-916.33993	-916.05172	-916.11979
Si-trans-syn-cis-ps	-916.35554	-916.06759	-916.13599

Table 13 Absolute energy for pathway **C**, step 3, at the level of DLPNO-CCSD(T)/def2-TZVPP//SMD(1,2-DICHLOROETHANE)-ωB97X-D3/def2-TZVP.

Step3 (Ring contraction)	E(hartree)	H(hartree)	G(hartree)
Re-cis-syn-cis-rs	-916.36468	-916.07591	-916.14285
Si-cis-anti-cis-rs	-916.36853	-916.08030	-916.14814
Re-cis-syn-cis-INT	-916.36533	-916.07678	-916.14122
Si-cis-anti-cis-INT	-916.37457	-916.08621	-916.15173
Re-cis-syn-cis-TS1	-916.35848	-916.07108	-916.13507
Si-cis-anti-cis-TS1	-916.36520	-916.07813	-916.14354
Re-cis-syn-cis-TS2	-916.35303	-916.06617	-916.13022
Si-cis-anti-cis-TS2	-916.36423	-916.07732	-916.14284
Re-cis-syn-cis-ps	-916.36695	-916.07848	-916.14580
Si-cis-anti-cis-ps	-916.36691	-916.07906	-916.14996
Re-cis-anti-trans-ps	-916.34733	-916.05886	-916.12863
Re-cis-syn-cis-ps	-916.36120	-916.07274	-916.14051
Re-trans-anti-cis-ps	-916.34895	-916.06049	-916.12827
Re-trans-syn-trans-ps	-916.32687	-916.03830	-916.10673
Si-cis-anti-cis-ps(ch3-ax-oh-eq)	-916.35390	-916.06487	-916.13250
Si-cis-anti-cis-ps(ch3-eq-oh-eq)	-916.35536	-916.06613	-916.13346
Si-cis-syn-trans-ps	-916.34414	-916.05567	-916.12331
Si-trans-anti-trans-ps	-916.33993	-916.05172	-916.11979
Si-trans-syn-cis-ps	-916.35554	-916.06759	-916.13599

B Bibliography

[1] M. I. Thomson, G. S. Nichol and A. L. Lawrence, *Organic letters*, 2017, **19**, 2199–2201.

[2] M. I. Thomson, *Doctoral dissertation, The University of Edinburgh*, 2019, **Edinburgh Research Archive**, https://era.ed.ac.uk/handle/1842/35536.

[3] Y. Yu, J. Yang, C. Peng, V. Caer, P. Cong, Z. Zou, Y. Lu, S. Yang and Y. Gu, *Journal of natural products*, 2009, **72**, 921–924.

[4] S. Taveepanich, N. Kamthong, N. Sawasdipuksa and S. Roengsumran, *Journal of scientific research, Chulalongkorn University*, 2005, **30**, 187–192.

[5] K. Hollenbeak and M. Kuehne, *Tetrahedron*, 1974, **30**, 2307–2316.

[6] S. Efimov, I. Khodov, E. Ratkova, M. Kiselev, S. Berger and V. Klochkov, *Journal of molecular structure*, 2016, **1104**, 63–69.

[7] G. Pescitelli, *Chirality*, 2022, **34**, 333–363.

[8] M. S. Buchanan, T. Hashimoto, S. Takaoka and Y. Asakawa, *Phytochemistry*, 1995, **40**, 1251–1257.

[9] M. Ono, X. Y. Zhao, Y. Shida and H. Akita, *Tetrahedron*, 2007, **63**, 10140–10148.

[10] J. A. Marco, M. Carda, S. Rodrıguez, E. Castillo and M. N. Kneeteman, *Tetrahedron*, 2003, **59**, 4085–4101.

[11] T. K. Kotammagari, R. G. Gonnade and A. K. Bhattacharya, *Organic letters*, 2017, **19**, 3564–3567.

[12] N. B. Khomane, J. S. Patel, P. K. Shirsat, P. R. Mali and H. M. Meshram, *ChemistrySelect*, 2018, **3**, 1517–1520.

[13] A. F. Kiefer, Y.-C. Liu, R. Gummerer, C. Jäger and J. Deska, *Angewandte chemie international edition*, 2023, **62**, e202301178.

[14] M. Orio, D. A. Pantazis and F. Neese, *Photosynthesis research*, 2009, **102**, 443–453.

[15] D. S. Sholl and J. A. Steckel, *Density functional theory: a practical introduction*, John Wiley & Sons, 2022, pp. 7–9.

[16] F. Jensen, *Introduction to computational chemistry*, John wiley & sons, 2017, pp. 165–201.

[17] C. Møller and M. S. Plesset, *Physical review*, 1934, **46**, 618.

[18] K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chemical physics letters*, 1989, **157**, 479–483.

[19] C. J. Cramer, *Essentials of computational chemistry: theories and models*, John Wiley & Sons, 2013.

[20] M. Loipersberger, L. W. Bertels, J. Lee and M. Head-Gordon, *Journal of chemical theory and computation*, 2021, **17**, 5582–5599.

[21] E. Soydas and U. Bozkaya, *Journal of chemical theory and computation*, 2015, **11**, 1564–1573.

[22] S. Grimme, *The Journal of chemical physics*, 2003, **118**, 9095–9102.

[23] C. Riplinger and F. Neese, *The Journal of chemical physics*, 2013, **138**, year.

[24] R. O. Ramabhadran and K. Raghavachari, *Journal of chemical theory and computation*, 2013, **9**, 3986–3994.

[25] P. Hohenberg and W. Kohn, *Physical review*, 1964, **136**, B864.

[26] P.-O. Löwdin, *Advances in quantum chemistry*, Elsevier, 1985, vol. 17, pp. 285–334.

[27] N. Mardirossian and M. Head-Gordon, *Molecular physics*, 2017, **115**, 2315–2372.

[28] S. F. Sousa, P. A. Fernandes and M. J. Ramos, *The journal of physical chemistry A*, 2007, **111**, 10439–10452.

[29] J. P. Perdew and K. Schmidt, AIP conference proceedings, 2001, pp. 1–20.

[30] Q. Peng, F. Duarte and R. S. Paton, *Chemical society reviews*, 2016, **45**, 6093–6107.

[31] F. Weigend and R. Ahlrichs, *Physical chemistry chemical physics*, 2005, **7**, 3297–3305.

[32] J. Chen, Y. Shao and J. Ho, *The journal of physical chemistry A*, 2019, **123**, 5580–5589.

[33] J. Tomasi, B. Mennucci and R. Cammi, *Chemical reviews*, 2005, **105**, 2999–3094.

[34] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of physical chemistry B*, 2009, **113**, 6378–6396.

[35] M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chemical reviews*, 2012, **112**, 1839–1862.

[36] G. Saielli and A. Bagno, *Organic letters*, 2009, **11**, 1409–1412.

[37] I. Stappen, G. Buchbauer, W. Robien and P. Wolschann, *Magnetic resonance in chemistry*, 2009, **47**, 720–726.

[38] J. A. Mendoza-Espinoza, F. López-Vallejo, M. Fragoso-Serrano, R. Pereda-Miranda and C. M. Cerda-Garcia-Rojas, *Journal of natural products*, 2009, **72**, 700–708.

[39] A. Williams, M. Elyashberg, K. Blinov, D. Lankin, G. Martin, W. Reynolds, J. Porco Jr, C. Singleton and S. Su, *Journal of natural products*, 2008, **71**, 581–588.

[40] S. D. Rychnovsky, *Organic letters*, 2006, **8**, 2895–2898.

[41] J. A. Porco Jr, S. Su, X. Lei, S. Bardhan and S. D. Rychnovsky, *Angewandte chemie international edition*, 2006, **45**, 5790–5792.

[42] Y. J. Hong and D. J. Tantillo, *Organometallics*, 2011, **30**, 5825–5831.

[43] G. Bifulco, P. Dambruoso, L. Gomez-Paloma and R. Riccio, *Chemical reviews*, 2007, **107**, 3744–3779.

[44] V. M. Schmiedel, Y. J. Hong, D. Lentz, D. J. Tantillo and M. Christmann, *Angewandte chemie international edition*, 2018, **57**, 2419–2422.

[45] S. G. Smith and J. M. Goodman, *The Journal of organic chemistry*, 2009, **74**, 4597–4607.

[46] S. G. Smith and J. M. Goodman, *Journal of the American Chemical Society*, 2010, **132**, 12946–12959.

[47] D. J. Tantillo, *Applied theoretical organic chemistry*, World Scientific, 2018, pp. 165–191.

[48] P. R. Rablen, S. A. Pearlman and J. Finkbiner, *The journal of physical chemistry A*, 1999, **103**, 7357–7363.

[49] M. Karplus, *Journal of the American Chemical Society*, 1963, **85**, 2870–2871.

[50] J. S. Fabian, J. García De La Vega and E. San Fabián, *Journal of chemical theory and computation*, 2014, **10**, 4938–4949.

[51] T. A. Young, J. J. Silcock, A. J. Sterling and F. Duarte, *Angewandte chemie*, 2021, **133**, 4312–4320.

[52] S. Riniker and G. A. Landrum, *Journal of chemical information and modeling*, 2015, **55**, 2562–2574.

[53] G. Landrum, Abstracts of papers of the American Chemical Society, 2019.

[54] A. Hagberg, P. J. Swart and D. A. Schult, *Exploring network structure, dynamics, and function using NetworkX*, Los alamos national laboratory (lanl) technical report, 2008.

[55] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser and J. Bright, *Nature methods*, 2020, **17**, 261–272.

[56] E. W. Dijkstra, *Edsger Wybe Dijkstra: his life, work, and legacy*, 2022, pp. 287–290.

[57] J. M. Goodman and M. A. Silva, *Tetrahedron letters*, 2003, **44**, 8233–8236.

[58] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Structural science*, 2016, **72**, 171–179.

[59] S. Kozuch and S. Shaik, *Accounts of chemical research*, 2011, **44**, 101–110.

[60] M. Besora and F. Maseras, *Wiley interdisciplinary reviews: computational molecular science*, 2018, **8**, e1372.

[61] F. Neese, *Wiley interdisciplinary reviews: computational molecular science*, 2022, **12**, e1606.

[62] J.-D. Chai and M. Head-Gordon, *Physical chemistry chemical physics*, 2008, **10**, 6615–6620.

[63] F. Weigend and R. Ahlrichs, *Physical chemistry chemical physics*, 2005, **7**, 3297–3305.

[64] F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chemical physics*, 2009, **356**, 98–109.

[65] S. Grimme, *The Journal of chemical physics*, 2003, **118**, 9095–9102.

[66] G. L. Stoychev, A. A. Auer and F. Neese, *Journal of chemical theory and computation*, 2017, **13**, 554–562.

[67] K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theoretical chemistry accounts*, 1997, **97**, 119–124.

[68] F. Weigend, *Physical chemistry chemical physics*, 2006, **8**, 1057–1065.

[69] F. Weigend, *Physical chemistry chemical physics*, 2002, **4**, 4285–4291.

[70] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The journal of physical chemistry B*, 2009, **113**, 6378–6396.

[71] S. Grimme, *Chemistry–A European journal*, 2012, **18**, 9955–9964.

[72] T. Young, *otherm*, 2020, **Github**, https://github.com/duartegroup/otherm.

[73] D. G. Liakos and F. Neese, *Journal of chemical theory and computation*, 2015, **11**, 4054–4063.

[74] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. Martin and F. Neese, *Journal of chemical theory and computation*, 2015, **11**, 1525–1539.

[75] C. Bannwarth, S. Ehlert and S. Grimme, *Journal of chemical theory and computation*, 2019, **15**, 1652–1671.

[76] A. Chandrasekaran, *Product Information as taken from the Honeywell Burdick and Jackson, Muskegon, (MI) – USA*, 2003, **Solvent Physical Properties**, https://people.chem.umass.edu/xray/solvent.html.

[77] C. Adamo and V. Barone, *The journal of chemical physics*, 1999, **110**, 6158–6170.

[78] S. Riniker and G. A. Landrum, *Journal of chemical information and modeling*, 2015, **55**, 2562–2574.

[79] P. Karafiloglou and G. Ohanessian, *Journal of chemical education*, 1991, **68**, 583.

[80] R. M. Parrondo, P. Karafiloglou and E. S. Marcos, *International journal of quantum chemistry*, 1994, **52**, 1127–1144.

[81] P. Karafiloglou, *Journal of computational chemistry*, 2001, **22**, 306–315.

[82] F. London, *Journal of physical Radium*, 1937, **8**, 397–409.

[83] R. Ditchfield, *The journal of chemical physics*, 1972, **56**, 5688–5691.

[84] T. Helgaker, M. Jaszunski and K. Ruud, *Chemical reviews*, 1999, **99**, 293–352.

[85] G. L. Stoychev, A. A. Auer and F. Neese, *Journal of chemical theory and computation*, 2018, **14**, 4756–4771.

[86] S. Grimme, *The Journal of chemical physics*, 2006, **124**, year.

[87] C. Adamo and V. Barone, *The Journal of chemical physics*, 1998, **108**, 664–675.

[88] J. Ho, *Physical chemistry chemical physics*, 2015, **17**, 2859–2868.

[89] C. P. Kelly, C. J. Cramer and D. G. Truhlar, *The journal of physical chemistry B*, 2007, **111**, 408–422.

[90] D. Wachsstock, *Tenua: the kinetics simulator for Java*, 2007.

[91] J. J. Klicić, R. A. Friesner, S. Liu and W. C. Guida, *The journal of physical chemistry A*, 2002, **106**, 1327–1335.

[92] A. J. Sterling, A. B. Dürr, R. C. Smith, E. A. Anderson and F. Duarte, *Doctoral dissetation, University of Oxford*, 2021, **Oxford University Archives**, scholar.archive.org.

[93] A. Howarth, K. Ermanis and J. M. Goodman, *Chemical science*, 2020, **11**, 4351–4359.