

Towards a *de-novo* route to drop-in  
biobased 3-methyl phthalic anhydride,  
a first principles computational study

WJ van Weesep

December 1<sup>st</sup> 2023

This thesis is submitted in partial fulfillment of the requirements for the degree  
of MSc Nanomaterials Science at Utrecht University.

Supervisors: N Artrith, RE Bulo

Daily supervisor: JM Boereboom

Examiner: P Ngene

Debye Institute; Faculty of Science; Department of Chemistry;

Materials Chemistry and Catalysis

## Abstract

The search for functional replacements of fossil-based fuels and chemical building blocks has produced promising sugar-based chemicals. Among these is phthalic acid.

An experimental procedure for the zeolite-catalyzed synthesis of 3-methyl phthalic anhydride, which was performed in our group, serves as the basis for a modeling investigation. Prior reported observations were used to design a computational DFT workflow.

With a systematic application of the model Brønsted reaction a partial reaction network is produced. Inspection of this network yields observations on both thermodynamic and kinetic descriptions of the reaction. The model workflow gains validity by being able to rediscover intermediates that were assigned experimentally.

The findings indicate that a previously 'unexpected' lactone intermediate results directly from a 1,3-substitution pattern on the reactant. This lactone is instrumental in determining the product composition and selectivity towards a phthalic anhydride precursor.

The modular workflow can serve as a platform for future mechanistic analysis in similar reactions as well.

## Samenvatting voor de leek

In de zoektocht naar vervangers voor brandstoffen en chemische bouwstenen voor hoogwaardige materialen zijn suikers een belangrijke grondstof. Vanuit experimenten zijn verschillende kandidaten voortgekomen, waaronder ftaalzuur.

In dit verslag presenteren wij de resultaten van computerberekeningen aan hoe de bio-uitgangsstoffen met elkaar reageren en hoe deze reactie te sturen. De theoretische basis die uit deze resultaten voortkomt biedt aanknopingspunten voor het verbeteren van dit proces door katalyse toe te passen.

De uitkomsten in de vorm van het vergelijken van twee mogelijke reactie-verlopen wijst er op dat ongewenste bijproducten kunnen worden vermeden door zorgvuldig gekozen uitgangsstoffen toe te passen met en de chemische omgeving te beheersen.

List of abbreviations and chemicals:

(ML)FF: (machine learning)  
forcefield

2-MF: 2-methyl furan

3-MPA: 3-methyl phthalic anhydride

B3LYP: Becke 3 parameter Lee Yang  
Parr

BTX: benzenes toluenes and xylenes

CPI: chemical-physical intuition

DA: Diels Alder

DFT: Density Functional Theory

DPE: deprotonation energy

EDG: electron donating group

EWG: electron withdrawing group

FAIR: findable accessible  
interoperable & reusable

FP: fingerprint

GC: gas chromatography

GGA: generalized gradient approach

GO: geometry optimization

H-BEA: hydrogen B zeolite

HOMO: highest occupied molecular  
orbital

H-Y: hydrogen Y zeolite

IR: infrared

LT: linear transit

LUMO: lowest unoccupied molecular  
orbital

MA: maleic anhydride

MFMA[H]: hydrogenated Diels Alder  
adduct of methyl furan with maleic  
anhydride

MM: molecular modeling

MSA: methyl sulfonic acid

PA: proton affinity

PBE: Purdew Burke Ernsterhof

PES: potential energy surface

RDS: rate determining step

RT: room temperature

SCF: self-consistent field

SMILES: simplified molecular input  
line-entry system

TA: toluic acid

TSS: transition state search

VSEPR: valence shell electron pair  
repulsion

VUSY: very ultra stable Y zeolite

i Title page	1
ii Abstract	2
iii Samenvatting voor de leek	3
iv List of abbreviations and chemicals	4
v Table of contents	5
1 Introduction	7
1.1 Backgrounds	7
1.2 Advances in phthalic anhydride from biomass-derived feedstocks	8
1.3 Research goals.	10
1.4 Research questions	10
2.0 Methods	12
2.1. Part A: Exploration	14
Step 1: Structure generation	15
Step 2: DFT geometry optimizations	16
Step 3: Energy-based screening	16
Step 4: 3D- 2D conversion	17
Step 5: Structure-based screening	17
Step 6: Enumeration of reactive acceptor sites (neutral) and donor protons (cations)	18
Step 7: Iteration on 3D structures	18

2.2. Part B: Intramolecular transformations and pathways	19
2.2.1 Constructing reaction coordinate diagrams	20
3.0 Results	22
3.1 Part A	22
3.1.1 Generation 1	23
3.1.2 Generation 2	25
3.1.3 Generation 3	26
3.2 Part B: Connections and elementary steps into pathways	27
3.2.1 Competition between fragmentation and tautomerization	29
3.2.2 Native reference barriers	30
3.2.3 Motivation for decarbonylation at uniquely the 3-position	31
3.2.4 Motivation for decarboxylation at uniquely the 2-position	32
3.3 Partial reaction network	32
3.4 Tracing Pathways	33
3.4.1 The direct dehydration	33
3.4.2 Via Lactone	35
3.4.3 <i>m</i> -toluic acid	38
3.4.4 <i>o</i> -toluic acid	40
3.5 Global results and outlook	41
Acknowledgements	43
Bibliography	44
Presentation slides	45

## 1. Introduction

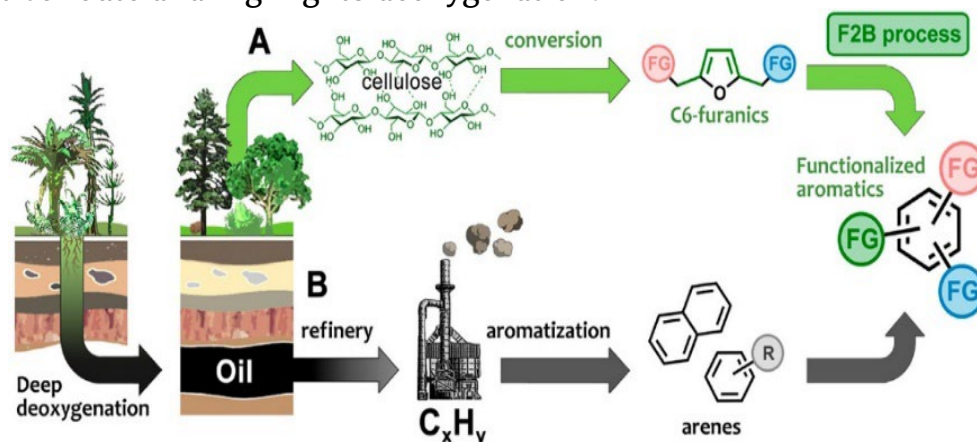
### 1.1 Backgrounds

Aromatic chemicals are used as fuels and as building blocks in diverse industries **Ennaert2016**. Traditionally, the lightweight aromatics, such as benzene, toluene, and xylene, are recovered fractions of the naphtha reformat and require laborious separation to obtain the compound in marketable form. One of the principles of green chemistry is to decrease our dependence on nonrenewable fossil feedstocks in favor of bio-sourced alternatives.

This study investigates a novel catalytic route towards high-demand bulk chemicals in an effort to replace the petrochemical sources with identical structures derived from renewable feedstocks. Such compounds that can be processed without requiring any change in downstream infrastructure or protocol are so called 'drop-in' chemicals **Belgacem2008**.

Biomass-derived feedstocks are high in oxygen content relative to petrochemical sources. The selective removal of oxygen from biomass-derived precursors is an essential step for their catalytic transformation into platform chemicals **Gupta2016**. Dehydration, decarboxylation and decarbonylation of biomass-derived chemicals allows for selective elimination of oxygen-containing functional groups to produce various substituted benzenes.

**Kucherov2021** illustrates the biobased furans to benzenes (F2B process) as an alternative route to functionalized aromatics. Scheme 1 illustrates the desired, alternative route and highlights deoxygenation.



Scheme 1. Schematic representation of the F2B process. From: **Kucherov2021**

## 1.2 Advances in phthalic anhydride from biomass-derived feedstocks

Table 1 lists six key reports on the catalytic conversion of biomass-sourced 3-methyl phthalic anhydride **Mahmoud2014**, **Genuino2014**, **Thiyagarajan2014**, **Thiyagarajan2015**; and the more primitive xylenes **Williams2014**, **Chang2014**.

With the exception of the Bruijninx group reports, all include a modeling component. We thereby identify a knowledge gap in the body of literature for a computational study into the heterogenous catalysis of 3-MPA.

We are aware of ongoing research by the Bruijninx group since the practical work on this project has been concluded. This concerns chiefly the DA adducts of MF with other suitable dienophiles such as maleimides, maleic acids, furfuryl alcohol and similar structures as reviewed in **Li2022**. Notably the advances reported in **Ciok2021** and **Lancefield2020**. These citations however fall outside the direct scope of the current report.



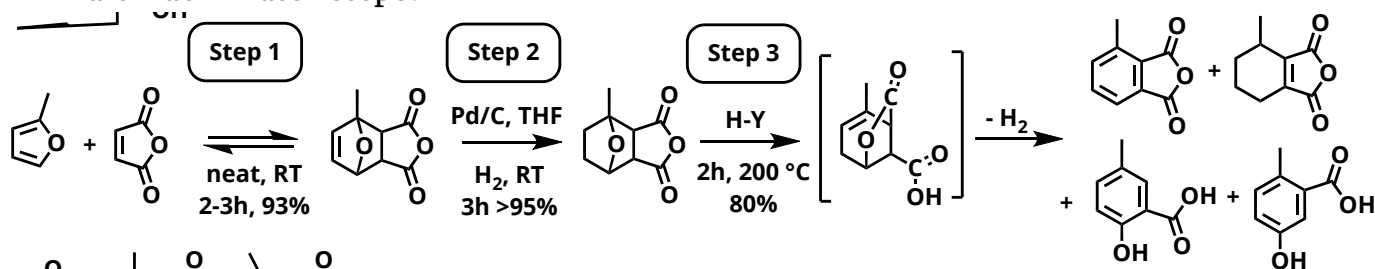
Table 1: Literature overview of reported catalytic conversion of biobased furans.

Reference	Group	Product	Catalyst	Selectivity
Mahmoud2014	Vlachos (UDelaware)	3-MPA	Homogeneous MSA	80%
Williams2014	Dauenhauer (UMinnesota)	<i>p</i> -xylene	H-Y & H-BEA	75%
Chang2014	Dumesic (UWisconsin)	BTX	WO <sub>x</sub> -ZrO <sub>2</sub>	77%
Genuino2014 Thiyagarajan2014 Thiyagarajan2015	Bruijnincx (Utrecht)	3-MPA	H-Y & H-VUSY	80%

### 1.3 Research goals.

This study aims to increase understanding of the underlying mechanism of the H-Y zeolite catalytic transformation of biomass-derived MFMA[H] (partially hydrogenated Diels–Alder adduct of 2-methyl furan (2-MF) with maleic anhydride (MA)) into value-added aromatics, as displayed schematically in Scheme 2. The methyl group on the diene and cyclic anhydride system of the dienophile appear on corresponding positions in the reaction product. The *cis* configuration of the oxygen-bridge to the anhydride system, which produces all oxygen atoms on the same face of the product, is favorable for selective elimination of water and carbon monoxide, carbon

dioxide in later steps.



Scheme 2: Redrawn from **Thiyagarajan2015**.

Schematic representation of the experimental protocol. (step 1) Diels–Alder reaction that creates the substituted 7-oxanorbornene framework;  
(step 2) partial hydrogenation to produce the exocyclic 7-oxanorbornane;  
(step 3) aromatization.

### 1.4 Research questions

- 1) Can a pathway for the overall transformation into each of four final products (toluic acids and phthalic anhydride) be proposed?
- 2) What is the mechanistic role of the prior identified lactone?

The results of identification of reaction intermediates and quantification of the pathways that connect them aid in understanding the factors that influence product selectivity and overall reaction yield and can be used in future investigations into how catalysis affects these. The primary goal is to identify the lowest-energy pathway towards the platform aromatics under non-catalyzed conditions to provide upper bounds in barrier height. The results from the exploratory analyses on pathways are then used to determine relative effects on the equilibrium concentrations of the proposed intermediates and reaction sequence.

The quantitative energetic results of DFT calculations are used to compare the energies of equilibrium for the individual components of a collection of tautomers and elimination products that can exist under acid conditions. Proposed intermediates are expected to either play a part in the overall transformation towards the main product, or lead to undesirable side-products.

The procedural generation of these products also gives us stepwise paths for these conversions.

Comparison of computed energy barriers provides a metric for the degree of rate control for the structures along the major synthetic pathway. Rate-determining steps are those along the reaction path that have the largest barrier to overcome.

## 2.0 Methods

We have adapted a computational workflow as reported in **Reiher2015** as a robust method that is able to propose not just a single major product, but many intermediates, in an iterative manner. The workflow is set up in order to grow a graph. In this partial reaction network nodes represent proposed intermediate structures and edges as elementary reaction steps. Abstraction of the network into reaction pathways, and ultimately reaction coordinate diagrams, allows for both qualitative comparison of pathways, and a quantitative description of global reaction energies and activation energies. The robustness of the workflow is used to evaluate three observations from the primary source **Thiyagarajan2015** as guidelines for hypothesis framing.

- The accumulation gradient of lactone during course of reaction
- The final aromatization step can be considered separately
- The lactone produces a comparable\* product distribution when used as the substrate

The workflow is divided in two parts. The first part, A, aims to identify the low-energy intermediates for the overall Brønsted acid-catalyzed transformation, with the explicit acid modeled simply as a proton to generate rational intermediates.

A production run of three cycles is manually curated. The computations are entered & evaluated by hand on basis of structural and energetic descriptors to retain chemically relevant products to avoid a combinatorial explosion.

The direct link between the model Brønsted reaction and a cycle within the workflow is captured visually in the graph representation, and in a descriptor to

retain the relationship between the zeroth-generation reagent and subsequent higher order generations of calculated intermediates, and thereby successive steps in a reaction route.

Subsequently, in part B, elementary reaction steps that describe edges on the partial reaction network are focused on. These are obtained initially by the generational descriptor from the results of part A and are broad but limited by the select calculations. Secondary edges are proposed and evaluated by a set of Linear Transit, Transition State Search, and IR Spectrum calculations. These are costly approaches and reserved for a further small selection of transformations. They also serve as a reference for possible extended research. A third set of edges is proposed by deductive analysis and expands on the shortfalls of the first set.

The results from part B are presented in two ways, as firstly a pathway through the reaction network representing the intrinsic reaction coordinate, and secondly as an out-of-plane slice along the energy axis *i.e.*, the intrinsic reaction coordinate diagram.

To investigate the main hypothesis: that there exists a (competitive) pathway to methyl phthalic anhydride that does not involve the lactone; the initial goal is:

to identify the low energy intermediates that are theoretically accessible within two generational cycles elementary reaction steps from the starting compound MFMA[H] (**1**).

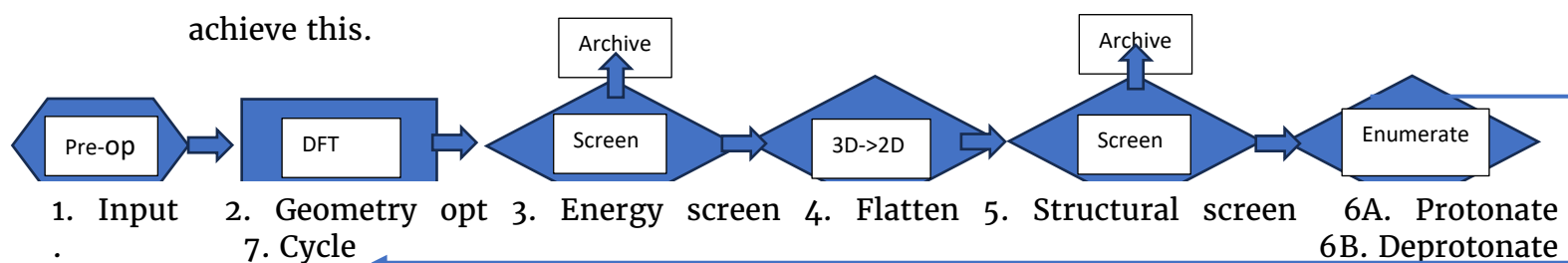
The secondary hypothesis: that Brønsted acid reactions on the lactone substrate produce (precursors to) all products; is secondly explored by applying the

workflow in a separate production run within one step beyond, effectively using the lactone as a zeroth generation.

## 2.1 Part A: Exploration.

Here, the objective is to assess thermodynamic favorability of the lactone in on overall pathway (kinetic features, *i.e.* barriers, will be assessed in part B). As a proxy for extent of contribution to the overall reaction rate, the expected form of results is a list of various ‘reasonable’ intermediates and a quantitative description of their partitioning.

Part A aims to employ a CPI-based strategy to contain combinatorial explosion while producing enough relevant structures. The computational reinforcement workflow, illustrated in Scheme 3 and described in the following section, aims to achieve this.



Scheme 3: Flow scheme for the computational workflow A in seven steps.

The workflow in Scheme 3 makes use of a reinforcement protocol whereby the results of calculation are used to update the search path and iteratively expand the collection of structures. This entails a high degree of manual evaluation of generated results between successive steps & multi-fidelity production cycles. The manual steps are illustrated as diamond-shaped nodes. Programmatic evaluation is desired in future for streamlining the data processing as well as relieving the prohibitive time-cost of manually submitting jobs and screening.

### 2.1.1. Step 1: Structure generation

Structure generation from literature mining as text or as Lewis structure. The procedure is bootstrapped with a zeroth generation starting point from the primary reference, specifically MFMA[H] **1**. Note that the final products from the primary reference 3-methylphthalic anhydride (3-MPA) and (*o,m*)-toluic acids **5** and **6** cannot be directly used in a double-ended string strategy as they have undergone a Lewis acid step as well for final aromatization. Validation of the proposed structures is instead evaluated as a likelihood to precursor/synthon to the final structures upon dehydrogenation over the metal catalyst.

The initial exploration is performed at low fidelity. The starting point structures from a reference or hypothesis are represented as text (common or systematic nomenclature, registry number) or Lewis structures and must be unambiguously converted to a 3D structure with appropriate stereochemistry. This is done manually through a computationally inexpensive model such as VSEPR or a (ML)FF forcefield.

Inputs for the three successive generations are retrieved from iterative application of step 1-6.

The higher fidelity exploration & validation makes use of the results from the lower fidelity GGA optimization run as starting structures.

### 2.1.2 Step 2: DFT geometry optimizations

Without a pre-trained (ML)FF for the unique system of the poly-substituted cyclohexene and cyclohexadiene molecular skeleton we have decided to use DFT geometry optimizations to provide the precision in identifying correct intermediates and accuracy in their relative energies.

As no external reference values and structures exist for the structures we propose in this study, no external benchmarks are available to compare performance.

A first production cycle of 3 generations is performed at the GGA level of the theory using the PBE functional. This is a (historically) popular setting for calculations and therefore selected also to be able to add to the existing body of FAIR research data.

A second production cycle of 3 generations is performed using the B3LYP hybrid functional for increased accuracy of numerical values and precision in replicating GGA results to either provide confirmatory support for the structure or highlight a difference that may indicate a need for higher fidelity calculations yet. Also, this functional has its drawbacks, especially in over-binding, and higher fidelity replications yet may make use of e.g. the range-corrected hybrid or Grimme dispersion correction.

### 2.1.3 Step 3: Energy-based screening

Calculation results are inspected for converged SCF cycles whereby a standard tolerance is selected. Tabulation of results on normalized energy (as a proxy for stability and equilibrium distributions). From the wealth of descriptors that are computed by the DFT workflow only the total energy is used directly in the reinforcement protocol for selection of intermediates to retain



Structures with energy  $E_i > E_{\text{cutoff}=\text{+13kcal}}$  are indexed accordingly and archived. They are not pursued to function as a parent structure as their population even at 200°C would be negligible in further iterations but are taken into the structural screening Step 4 and 5 for characterization.

#### 2.1.4 Step 4: 3D- 2D conversion

For the intermediate assessment of results indicated by the diamond-shaped nodes, structures are represented as their Lewis structures. This is to benefit systematic numbering and indexing as well as for interpretation of structural similarities by bonding patterns. 2D - 3D / file formats & representations

#### Data considerations for analysis & FAIRness

Geometries are represented as cartesian coordinates and atoms have not been aligned or systematically & robustly numbered to date. Use of connection tables / Z-matrix, string representations such as SMILES, fingerprinting or other representational embeddings have not been realized to date. Transforming descriptors from the output file has been restricted to calculating the PA & DPE values as the difference between neutral parent and the corresponding set of protomers, and between a protomer and the various corresponding deprotonated structures, respectively.

Complete calculation results are archived for re-use according to FAIR guidelines.

#### 2.1.5 Step 5: Structure-based screening

Akin to (Morgan) FP a way to characterize the functionality according to motifs and define a metric for similarity. Handhold to future investigation through heuristics on classes of functionalities. Several structures do display structural

correspondence/similarity but are high in energy. These can be flagged. All structures either flagged by the energy-based screening or the structural-based screening are classified as probably correct or probably erroneous and may be candidate for (higher fidelity) replication.

#### 2.1.6 Step 6: Enumeration of reactive acceptor sites (neutral) and donor protons (cations)

In the GGA production cycle all donor / acceptor sites were exhaustively enumerated and submitted for calculation. In the Hybrid production cycle the number of calculations was reduced to include hand-picked selection of these results that seemed most promising (*vide infra*).

The assignment of acceptor sites is done lone pairs on oxygen and  $\pi$  bonds on basis of the Lewis representation taking into regard stereochemistry. The assignment of donor hydrogens (11) is reduced by degeneracy in the product (*i.e.*, methyl hydrogens and the saturated linear chain). Input files are prepared using the optimized parent generation structures, and adding a (reactive) proton in proximity to the respective acceptor site to create a structure that is high up on the potential energy surface in vicinity of a potential well representing the to-be optimized structure.

#### 2.1.7 Step 7: Iteration on 3D structures

Initially at a low-fidelity GGA calculation. Plotting of the structures per generation and with their Lewis representation & Energy. Initially all the possible  $[M+H]^+$  protomers are drawn that result from protonating the various strong nucleophilic positions in a flat (2D) representation. Note that a single nucleophilic group can offer various conformational positions, and thus it is difficult to predict *a priori*

which 2D structures constitute the lowest-energy conformation and therefore must be screened. This 2D collection is preoptimized in 3D using a fast MM routine to allow for easier convergence during the DFT optimization. Jobs are submitted for DFT geometry optimization at a low fidelity PBE functional and ranked according to proton affinity (PA) as inputs for Step 3. The energy of equilibrium and corresponding geometry of each unique protomer are compared to providing a measure of relative proton affinity and to observe the accompanying restructuring of the carbon framework. A *post hoc* filtering was performed to discard fragmentation of the six-membered ring and the structures in excess of 13 kcal/mol relative to the lowest PA within a set. Energy values are transformed to PA and DPE descriptors by taking the arithmetic difference between energy of the optimized protomer energy to the optimized neutral parent structure + energy of the isolated proton. With these values for relative proton affinity then the comprehensive partial reaction network is expanded.

## 2.2 Part B: Intramolecular transformations and pathways.

The aim of part B is to identify intermediates related by single edits, corresponding to elementary reaction steps, and pathways made up of consequent steps.

Additional quantification of the change in energy along a pathway produces a reaction coordinate diagram. The reaction coordinate diagram thus captures the morphology of the potential energy surface along a proposed route and can be used for assigning degree of rate control.

Reaction coordinates, or reaction pathways are sequences of elementary reaction steps. The results of relative ground-state energies in part A represent a

morphological feature of the PES, namely the potential wells. The collection of these stable intermediates represents the partition of equilibrium populations unaccounting for the pathways and corresponding kinetics in which these structures can be formed. This generality provides the possibility to propose elementary steps by arrow-pushing a variety of transformations, and construct input files to interrogate the other morphological feature of interest: saddle-point energies connecting potential wells.

### 2.2.1 Constructing reaction coordinate diagrams.

The primary hypothesis is that we can draw a reaction coordinate diagram through (a maximum of) four intermediates to arrive at a reasonable precursor to the main product, both with and without featuring the lactone.

Three sources of step-proposal were used. Firstly, due to the construction of the workflow, the results of part A represent an abstract but well-defined Brønsted acid reaction. The precise description of the sequence of steps is used to write out the Lewis mechanism for the conversion.

Secondly, select transformations are identified to a second hand-picked set of transformations and are subject to further DFT calculation.

We aim to first construct a reference pathway of purely intramolecular (internal) transformations to classify native stability of the intermediates and assess which transformations can benefit the overall rate and selectivity most, if engineered. Reaction engineering in this sense can be through modification of the external effects (solvation, kinetic dilution, catalyst acid strength & nanoscale zeolite morphology) and has been explored in the experimental studies by the Bruijnincx

group. Internal barrier engineering concerns decorating with molecular functionality to influence *e.g.*, orbital interactions such as HOMO & LUMO gap through EWG and EDG.

The first effort is identifying energetic upper bounds to the individual steps in a non-catalyzed environment, as a reference for comparison of catalytic effects. To gain insight into the energetic barriers for transformation, several elementary steps were selected to further analyze the energetics of intramolecular reaction through proton transfer between closely located donor and acceptor sites.

### 3.0 Results

The research questions:

- (1) Is there a pathway that does not feature the lactone **3**;
- (2) What is the role of the lactone in the overall mechanism;

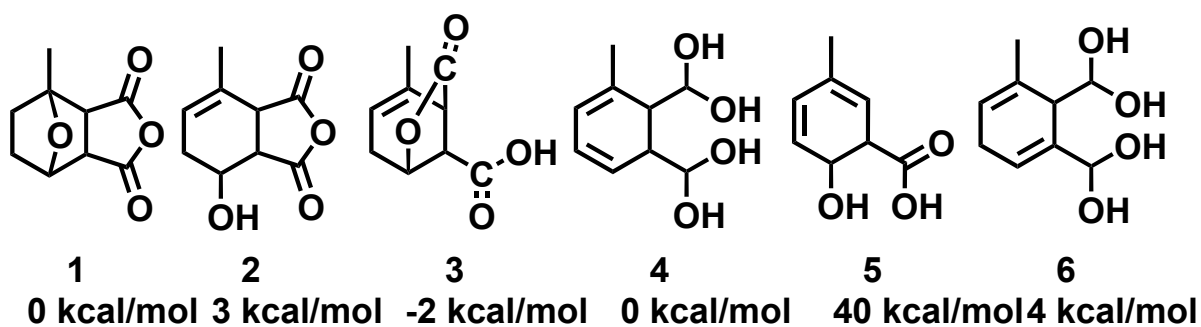
were investigated in two parts. Part A aims to identify and compute the energetics of relevant intermediates. Part B assembles the intermediates in a partial reaction network (hereafter simply network) by connecting them through elementary reaction steps (hereafter simply steps).

#### 3.1 Part A

The proton on-proton off procedure computes by DFT neutral intermediates and their molecular (cat)ion structures as converged cartesian geometries and total energies. Additional descriptors for generation sequence and respective (de)protonation energies are assigned manually.

The geometries and corresponding energetic values are initially calculated in the PBE production cycle. Replication at higher fidelity settings was performed for a select number of structures to limit number of calculations while retaining the relevant ones.

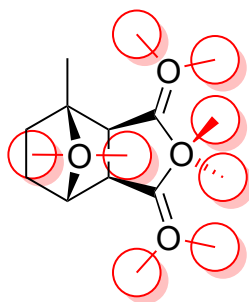
From the final collection of > 100 generated neutral structures there are six structures assigned a key role. These feature in the network. Scheme 4 lists the structures and B3LYP-water energies.



Scheme 4: The main structures for this report and their B3LYP-water energies.

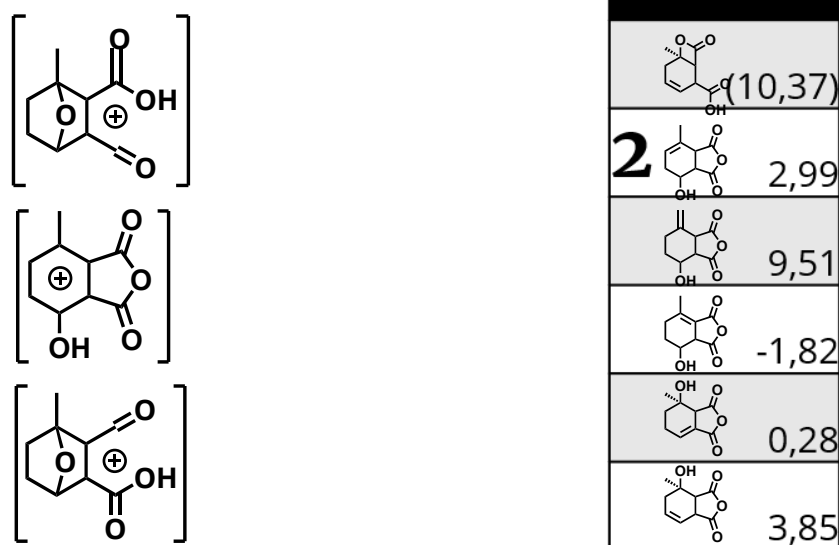
### 3.1.1 Generation 1

Protonation of 1 was performed at eight nucleophilic positions. These are the heteroatom positions indicated in scheme 5.



Scheme 5: Nucleophilic positions considered for protonation.

To aid in reducing data complexity, a *post hoc* interpretation assigns three distinct motifs. These motifs are inspired upon the Comparative Molecular Field Analysis technique (CoMFA) and consider distinct positions as regions which produce (at least one) identical product(s) after deprotonation. This approach is illustrated in Scheme 6 for the first generation of results.



Scheme 6: (left) three motifs; (right) selected stable products.

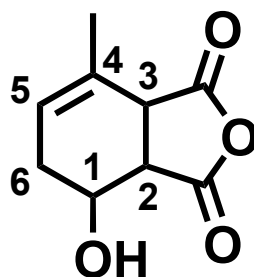
The exhaustive table of generation 1 structures is available upon request.

Deprotonation of the motifs yields two types of fragmentation products, namely the fragmentation of the six-membered ring, or conversely the fragmentation of the five-membered ether ring (bridge opening). The former consists of variations on tethered DA reagents. The latter is an opening of the oxanorbornane bicyclic bridge. The cleavage of the five-membered ring produces four isomeric products that differ in the position of the hydroxyl group &  $\pi$ -bond.

The structure-based screening rests on CPI and parsimonious structures, which is to say with a preference for shortest sequence of elementary steps, are chosen to proceed upon. Structure **2** bears the most resemblance to the lactone by substitution pattern on the cyclohexene skeleton, and is chosen as the depth-first search candidate.

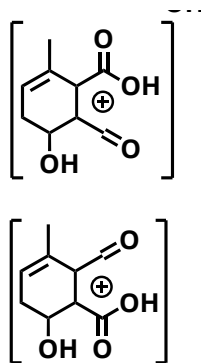
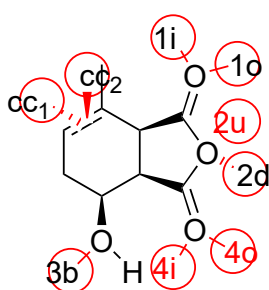
Numbering of the ring substituents in all structures is chosen for consistency based on the nomenclature of **2**, as displayed in Scheme 7.





Scheme 7: numbering of ring positions

### 3.1.2 Generation 2



structure	kcal/mol
<b>3</b>	-2,17
<b>4</b>	-0,35
<b>5</b>	*(40,7)
<b>6</b>	3,54

Scheme 8: (left) Nucleophilic positions on 2; (center) motifs; (right) stable products

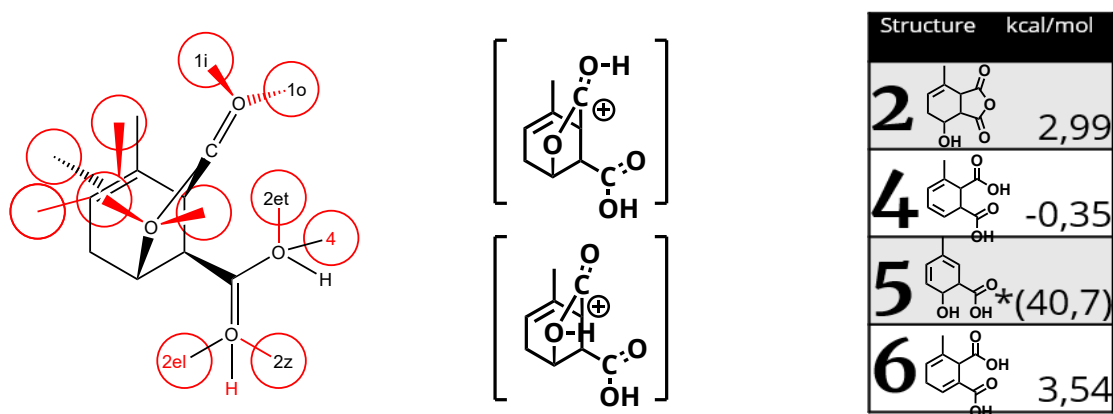
Nine distinct positions are recognized and illustrated in Scheme 8. Four motifs are recognized: two anhydride-protonated motifs and the hydroxyl-protonated one are displayed. The pi-protonated motif does not produce any structures below the 13 kcal/mol cutoff for energy-based screening. Neither does the 3-carboxylic acid motif (3-COOH). The 2-carboxylic acid motif (2-COOH) produces all stable products (energetic screening), and on basis of the structure-based screening the structures labeled **3 4 5 6** are selected.

Although structure **5** exceeds the energetic threshold, after structure-based screening this decarbonylation product is included in this short-list

due to the unique fragmentation pattern that is a precursor to the formation of the *meta*-toluic acid byproduct. Additionally, the 1-hydroxyl protonated motif is hypothesized, but not confirmed in calculation, to eliminate a molecule of water (*vide infra*).

### 3.1.3 Generation 3

The generation 3 most stable products are identical to those obtained in generation 2. Scheme 9 illustrates the major contributing motifs.



Scheme 9: Protonated lactone; motifs; and the main products.

### 3.2 Part B: Connections and elementary steps into pathways

In this section the focus lies on connecting the structures as network nodes with edges representing elementary reaction steps. Through these connections we shall trace pathways and draw the corresponding reaction coordinate diagrams. These diagrams serve in mechanistic analysis to identify global thermodynamic and kinetic attributes of the overall reaction.

Three collections of edges are included. First, steps analogous to the generative proton on-proton off sequence are drawn. Second, intramolecular (internal) transformations are proposed on basis of structural similarity through the transfer of a hydrogen from a donor to an acceptor position (a non-rigorous approach by manual inspection of structures for the identification of single edits). Third, a set of edges that describe the intermolecular Brønsted reaction with an explicit external acid and conjugate base.

The intramolecular transformations of select structures into their tautomers\* are verified through a Transition State Search (workflow B) to obtain reference energetic barriers. Note carefully that the barriers represent a non-catalyzed transformation as the non-protonated structures are optimized along a single constrained dimension. These transformations are represented as a second set of edges in the network, illustrated with directional weighted solid lines which represent the height of these barrier reference energies.

\*We here define tautomerization in a general fashion: the relocation of a hydrogen atom and a corresponding bond, including ring-chain, or annular, tautomers

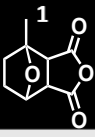
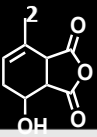
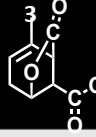
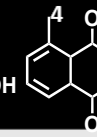
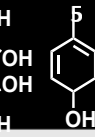
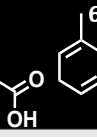
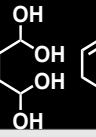
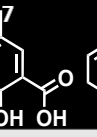
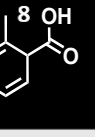
**Stanovnik2001** and not exclusively sigmatropic kind whereby hydrogen migration is complemented by a  $\pi$ -bond.

The third set of edges are included by generalization of the observed patterns on non-computed substrates, as well as a qualitative application of heuristic rules (e.g., hydride transfer reactions). Attempts to hard-code reaction rules from Reaxys, or application of Baldwin's rules and similar rule-based approaches showed little direct promise as they were not able to replicate the proton-on/off results.

Table 2 summarizes the edges from the three sets: (1) from generation sequence. (2) from intramolecular tautomerization. (3) from hypothesized similar steps.

The connection table format illustrates horizontally the generations of selected stable and parsimonious intermediates. The first generation on the first row includes thus the single edge between parent structure 1 and its product 2. Generation two and three on the second and third row, respectively, illustrate multiple intermediates that are generated from their parent.

Table 2: Connection table of network edges. Nodes are numbered 1 -9 in the top row and represent intermediates. Edges are illustrated at the intersection of a row and column. Energetic values from the B3LYP-water production run as reaction energy (A) and barrier height (B). Two decarbonylation steps are hypothesized and not calculated, these are represented as H.

$A(E_r)$ $B(\Delta E_{ts})$ kCal/mol									
1		$\underline{A}(+3)$ $\underline{B}(54)$							
2	$\underline{A}(-3)$ $\underline{B}(51)$		$\underline{A}(-5)$ $\underline{B}(25)$	$\underline{A}(-3)$	$\underline{A}(37)$	$\underline{A}(1)$	H	(1) $\underline{B}(42)$	(10) $\underline{B}(54)$
3		$\underline{A}(+5)$ $\underline{B}(30)$		$\underline{A}(+3)$ $\underline{B}(50)$	$\underline{A}(42)$ $\underline{B}(46)$	$\underline{A}(+4)$	H		

Because of the physical basis of our workflow to represent the Brønsted acid mechanism, the labeled results of generational sequence of part A provide a natural set of connections between the intermediates as nodes in a network. The difference in energy between parent and child structures is reported as an energy of reaction and are represented in the network as a directional weighted line in the final network: Scheme 12.

### 3.2.1 Competition between fragmentation and tautomerization.

Interestingly, the four production runs show that the proton on-proton off procedure can produce divergent products for the same sites, under variation in computational settings. Both of these results are chosen for representation in the network.

The 2-COOH motif on structure 2 yields a decarbonylation product when performed under the water-solvated PBE production run as well as the lactonization product in all other settings. Protonation of the 3-COOH motif on structure 3 also gives both a decarbonylation and a ring-opening of the lactone. The B3LYP production run does not show this which may be attributed a known over-binding as a relic of this hybrid functional choice. **Holthausen2001**.

Whereas all reaction steps are elementary and therefore reversible, the fragmentation reactions are negligibly occurring in a reverse direction and taken simple as irreversible steps toward final, lower mw, products such as the toluic acids.

Not all calculations have been performed on all sequences. Inspection of these results is sufficient to frame similar steps in part C.

### 3.2.2 Native reference barriers

For five reaction steps the computational workflow B was able to produce saddle-point structures listed in Table 3. These represent the intramolecular (internal) transformation by rearrangement of a hydrogen from a donor position into an acceptor position, and the corresponding geometry rearrangement. These barriers are noted  $\Delta E$  and serve as a threshold and reference value to evaluate  $\Delta\Delta E$  in catalysis in a computational procedure inspired upon **Nikbin2013**, **Zimmerman2017** [not in this study].

The procedure was set up to evaluate the possibility of internal rearrangement by a singly-constrained optimization along a straight vector. The five barriers that have been verified by a single imaginary frequency in the harmonic analysis. A sensitivity to orientation of the hydroxyl group showed that the competition between fragmentation and restructuring is not necessarily a bifurcation as **Houk2008** describes because the product is determined before the saddle-point.

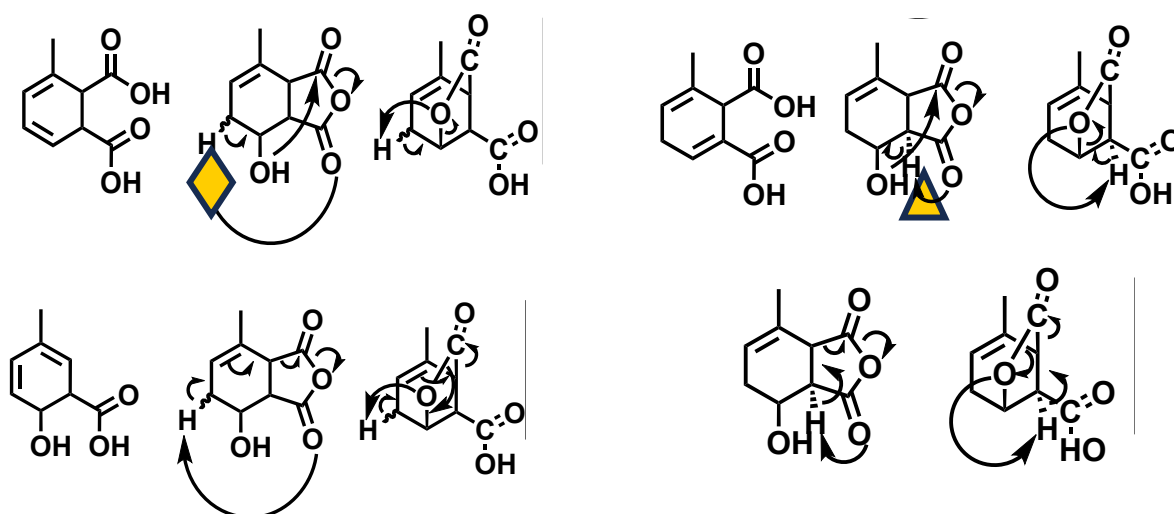
Table 3: Saddle-point structures

ID	$\Delta$ Energy (B3LYP-water)
TSS 1-2 [ring opening]	53,56 kcal/mol
TSS2-3 [lactonization]	25,61 kcal/mol
TSS3-8[decarboxylation]	47,34 kcal/mol
TSS3-4[rearrangement]	42,59 kcal/mol
TSS2-9[dehydration]	53,51 kcal/mol

### 3.2.3 Motivation for decarbonylation at uniquely the 3-position

The substituted cyclohexene conformation is restrictive to exclusively forming gamma-lactones. The 1,3-configuration for the lactonization that has been reported by Kato<sup>1977</sup> is seen also in our results as the 1,2 beta-lactone has energies exceeding our filtering threshold. The 4-methyl substituent which plays a part in stabilizing the oxanorbornane ring-opening in step 1, thereby also controls the 2-COOH motif and ultimately the formation of 3.

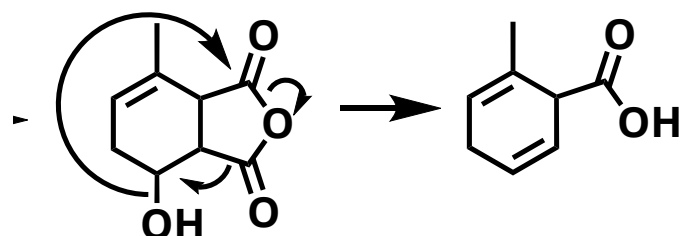
Scheme 10 illustrates the decarbonylation of the 3-position acyl substituent in an arrow pushing mechanism. The 3-COOH motif lacks the conjugated pi-system to accomplish a comparable fragmentation.



Scheme 10: Selective decarbonylation of the 2-COOH motif in 2 and the corresponding bridge motif in 3. Left, for deprotonation at the 6-position indicated with a diamond; right the 2-position indicated by a triangle.

### 3.2.4 Motivation for decarboxylation at uniquely the 2-position

The decarboxylation was identified during the TSS workflow as a neutral fragmentation. Scheme 11 illustrates this.

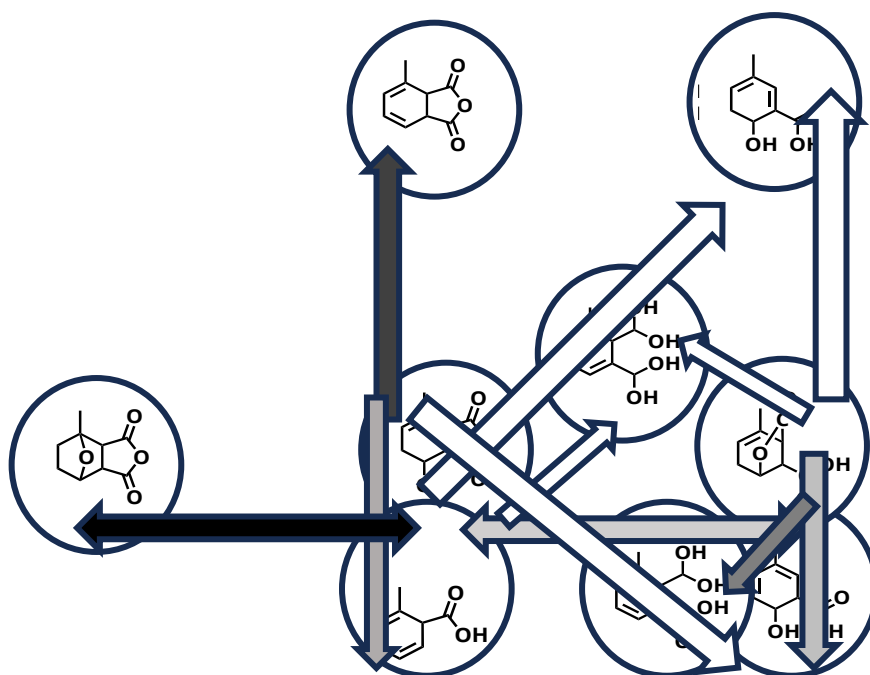


Scheme 11: Proposed decarboxylation step

### 3.3 Partial reaction network

Combining the results from above described parts we now draw the network in order to connect steps and trace pathways.

Scheme 12: Network of nine structures and (energy-weighted by grayscale) connections between them.





### 3.4 Tracing Pathways

The network in scheme 12 of nodes (pt A) and edges (pt B) is inspected to propose pathways of connected steps for the overall reactions into main product 3-MPA (3.4.1 & 3.4.2) and (*o,m*)-toluic acid side products (3.4.3 & 3.4.4).

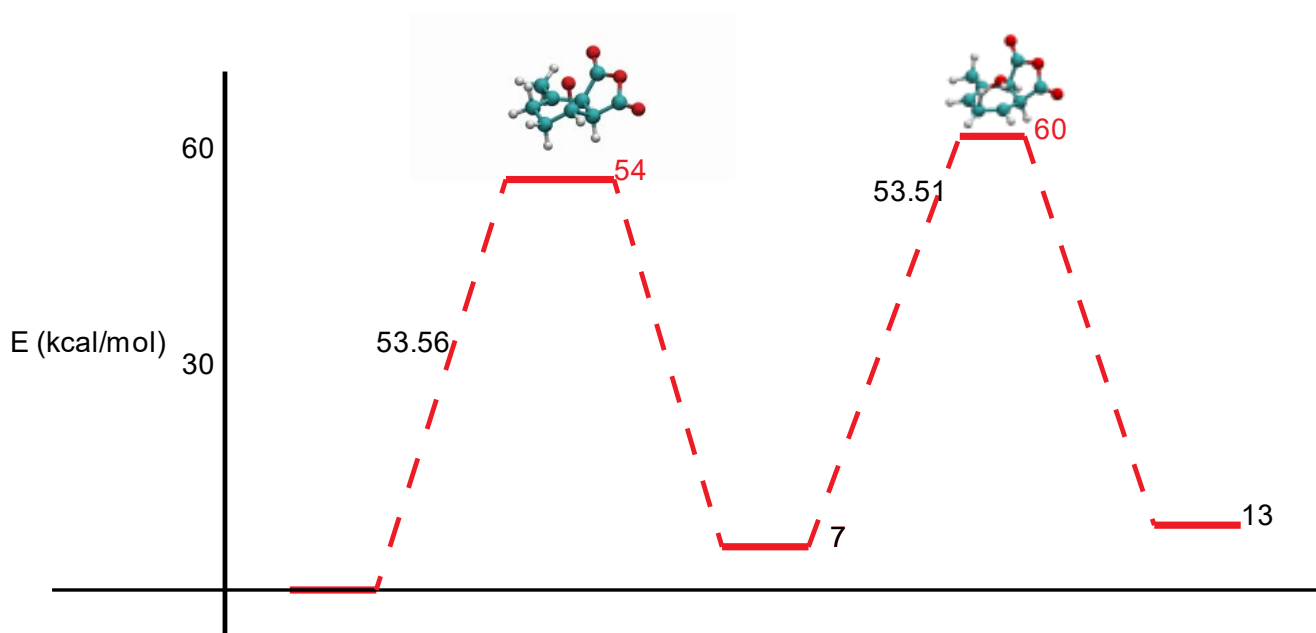
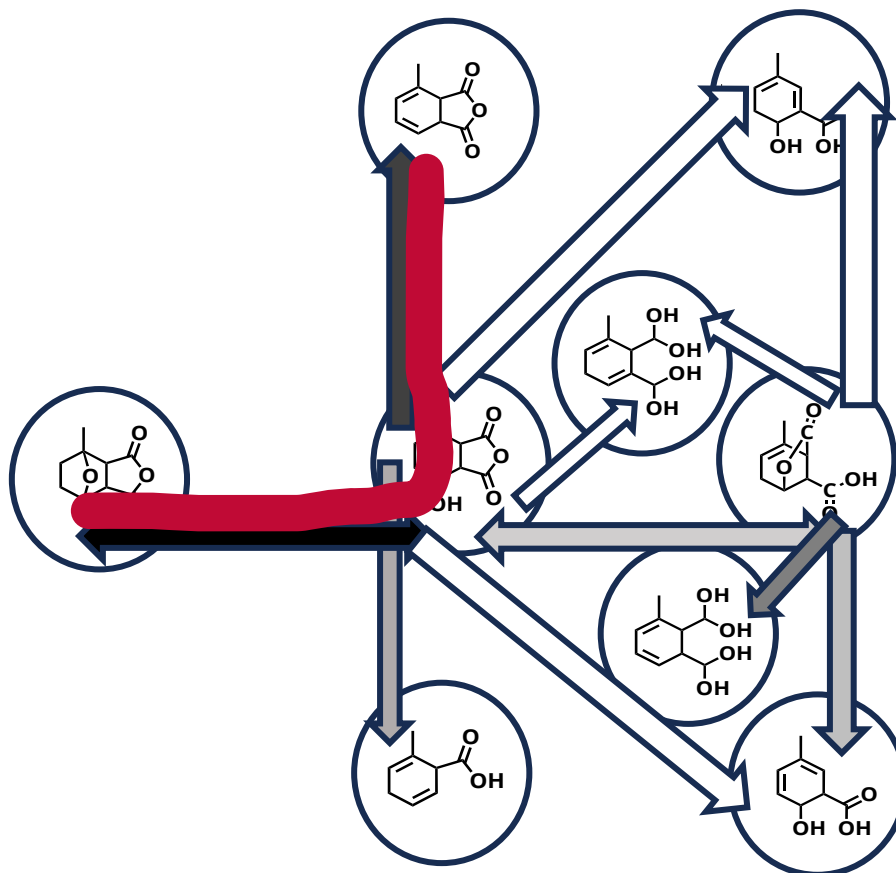
The observations in **Genuino2014** report the aromatization step to be catalyzed by either Lewis or Brønsted acid. Endpoints of the proposed routes are structures **4**, **6**, **9** as main product precursors to 3-MPA; structure **5** as endpoint for a *o*-toluic acid byproduct; and **7,8** for a *m*-toluic acid byproduct.

#### 3.4.1 The direct dehydration

This two-step pathway is the shortest route to a 3-MPA precursor and omits the formation of the lactone. This pathway features relatively large native barriers for all steps; and has the greatest activation energy of all. The uphill thermodynamics make this reaction unfavorable, although a true dynamic equilibrium will not be attained due to the fragmentation reaction and removal of water which draw the balance to the product side. Scheme 13 illustrates the direct dehydration network path and the corresponding reaction coordinate diagram.

Scheme 13: Pathway for direct dehydration (top) & reaction coordinate diagram

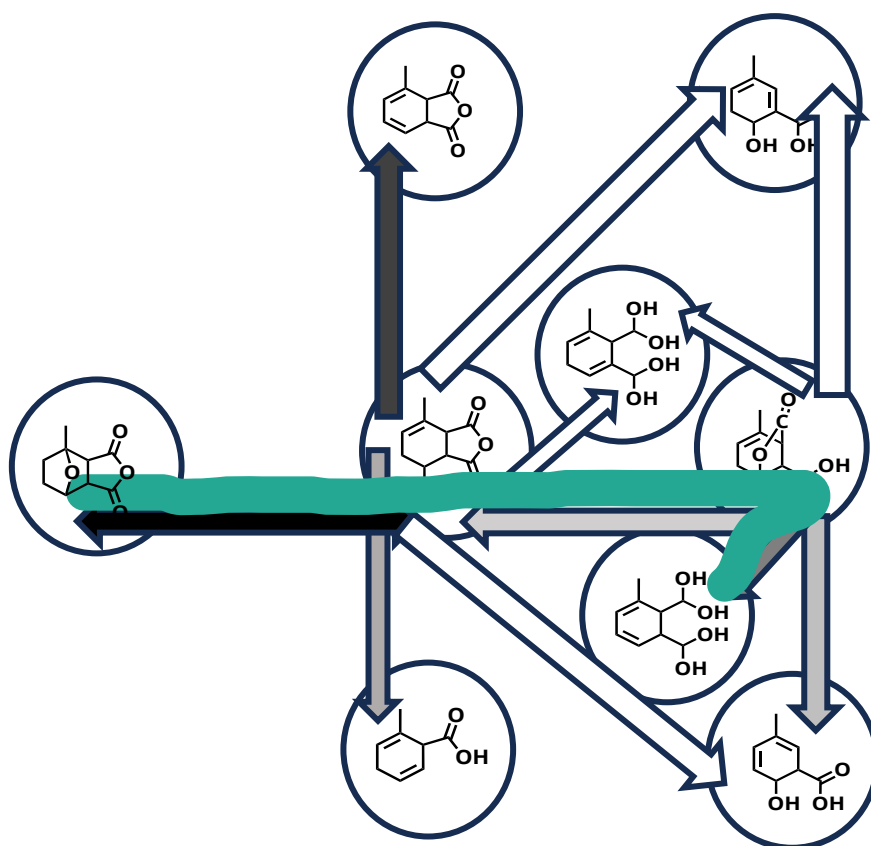
(bottom)

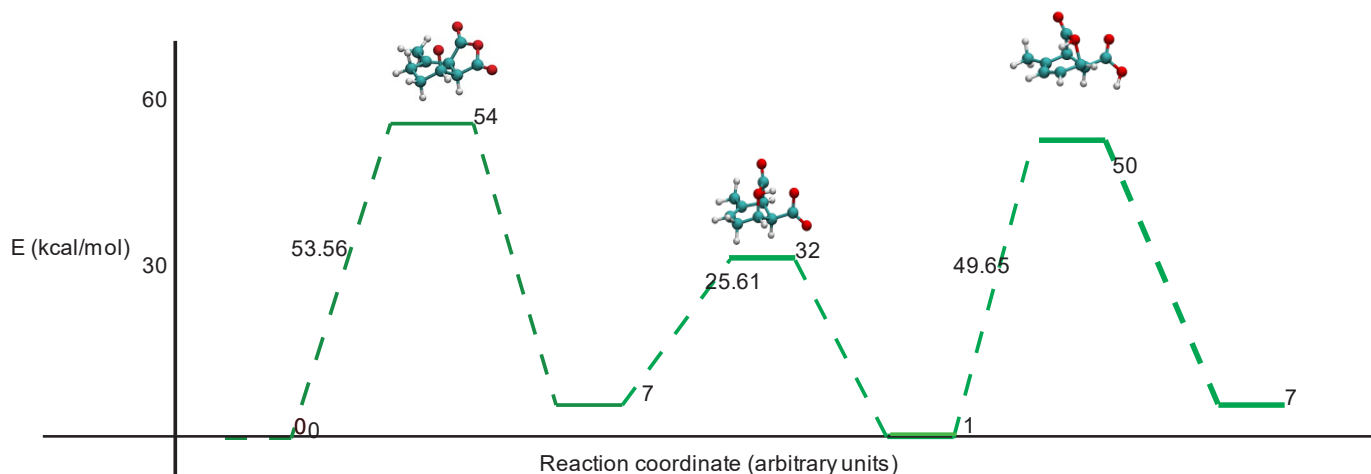


### 3.4.2 Via Lactone

The lactone is the most stable intermediate with a relative energy of  $-2.99$  kcal/mol compared to the starting reagent **1**. This high relative stability rationalizes the observed pooling of this intermediate and the observed concentration gradient over the course of the experimental reaction in **Thiyagarajan2015**. Scheme 14 illustrates the pathway and reaction coordinate diagram.

Scheme 14: (top) Network pathway into the phthalic anhydride main product via lactone; (bottom, next page) reaction coordinate diagram.





A comparison of the pathways into the main product by direct dehydration vs. via lactone in table 4 shows a considerable energy difference of 29kcal/mol. The Boltzmann partition of a direct dehydration, even at 200°C, is negligible\*.

\*The Boltzmann ratio of products as a function of the difference in internal energy at 200°C  $N_2/N_1 = e^{(E_2-E_1)/RT} \approx 1 \cdot 10^{-21}$

The lower native barrier heights represent a kinetic favorability towards the via lactone route.

A measure of centrality (number of connections in the network) acts as a proxy for the number of products that can be obtained from an intermediate in a single step. A greater number of connections over comparable barrier heights implies a greater number of byproducts or alternate pathways that said structure is involved in. A lower degree of centrality therefore is more desirable as it represents a purer distribution of products.

The bridge opening of diverse substituted oxanorbornane systems has been reported by Lajunen1991 where both unimolecular (E1) and bimolecular (SN2)

mechanisms have been suggested to be rate limiting in the acid catalyzed hydrolysis.

In comparison of main product pathways, we discern that the lactone is incapable of producing the decarboxylated product directly. Thus the ‘via lactone’ route represents a more linear pathway toward the desired main product, whereas direct dehydration is subject to a more branched pathway and the formation of a decarboxylated *o*-toluic acid.

Table 4: Comparison of thermodynamic & kinetic global descriptions for the main product by direct dehydration and via the lactone

	Direct dehydration	Via lactone
Step 1	Ring-opening $\Delta E_{ts}=54$ kcal/mol	Ring-opening $\Delta E_{ts}=54$ kcal/mol
Step 2	Elimination $\Delta E_{ts}=54$ kcal/mol	Lactonization $\Delta E_{ts}=25$ kcal/mol
Step 3	-	Isomerization $\Delta E_{ts}=50$ kcal/mol
Activation energy	$E_{activation}=57$ kcal/mol	$E_{activation}=54$ kcal/mol
Reaction energy	$E_{reaction}=13$ kcal/mol	$E_{reaction}=7$ kcal/mol
Centrality	7	5

We conclude that the role of the lactone in the proposed pathway is the kinetic and thermodynamically favored intermediate, and moreover can produce the main product more consistently than a direct dehydration could.

The reported observations of product distribution in the experimental report are comparable but not identical. The values of product distribution under the two substrates are reproduced from **Genuino2015** in table 5.

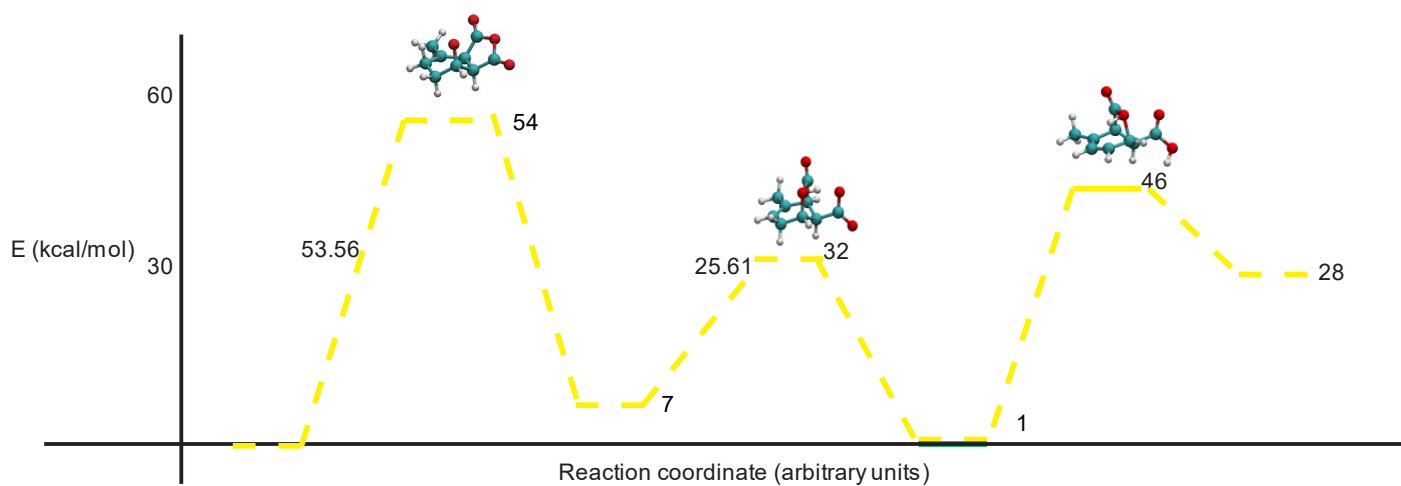
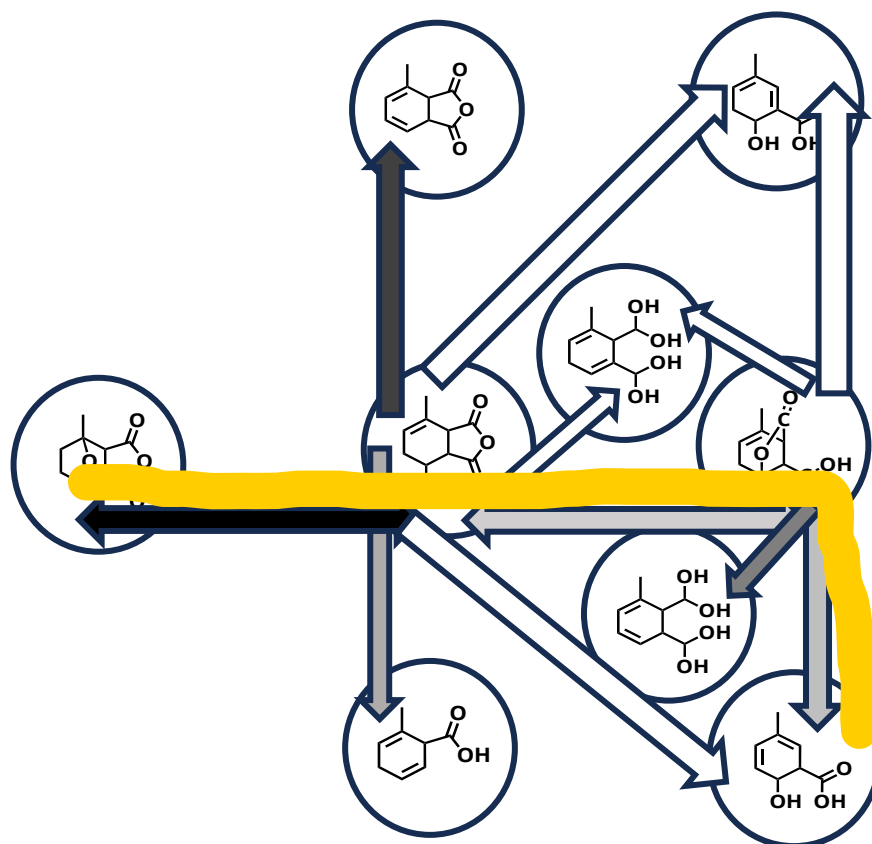
Table 5: Product distributions for replacing substrate **MFMA[H]** with **3**, from **Genuino2015**.

substrate	3-MPA	<i>o</i> -TA	<i>m</i> -TA
<b>MFMA[H]</b>	56%	17%	2%
<b>lactone</b>	52%	28%	4%

### 3.4.3 *m*-toluic acid

Due to limitations of the methodology for examining the intramolecular transformations, the decarbonylation of **2** does not have a quantified barrier. The decarbonylation of **3** is illustrated in scheme 14. The sequence of steps for lactonization is in competition with this decarbonylation where the fragmentation barrier of 48 kcal/mole exceeds the lactonization with value 25kcal/mol. It would be desirable to suppress the fragmentation in favor of lactonization to avoid formation of (excess undesired) toluic acid.

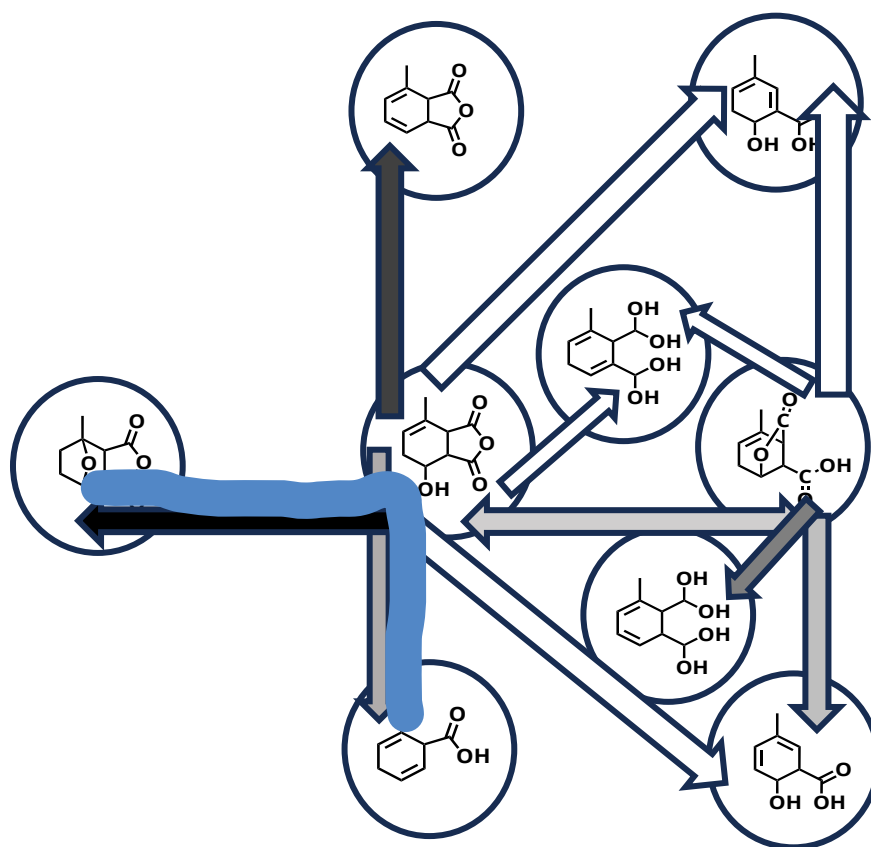
Scheme 14: (top) Network pathway for decarboxylation via 3; (bottom) reaction coordinate diagram.



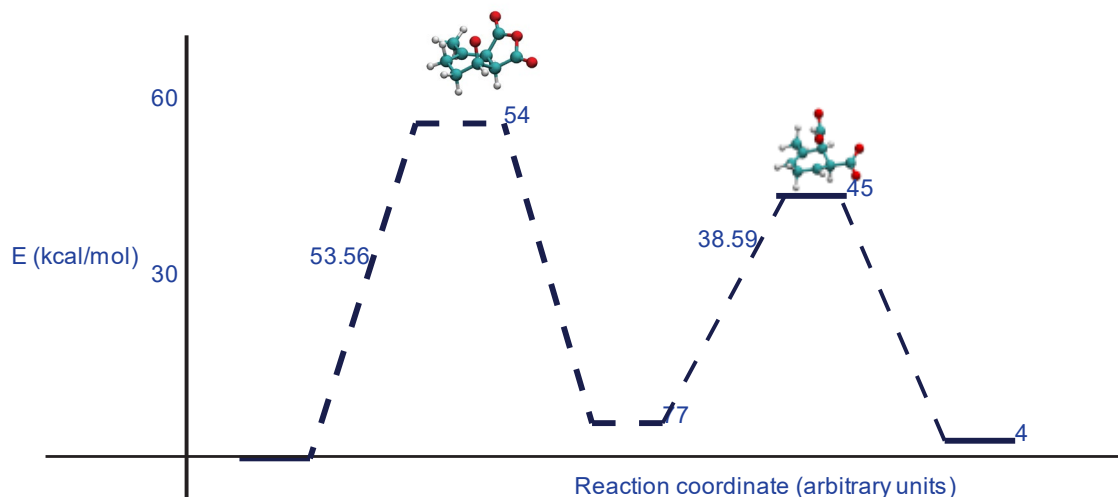
### 3.4.4 Decarboxylation

The decarboxylation of 2 as a precursor to *m*-toluic acid was uniquely identified through the workflow B. Scheme 15 presents the network path and reaction coordinate diagrams.

The decarboxylation, or its 2-position fragmented product, does not appear in the results of workflow A. Nevertheless, as a precursor to *o*-toluic acid this pathway is included in the network. The thus hypothesized decarboxylation into the *ortho* substituted product, and the decarbonylation into the *meta* substituted product offer a hypothesis to qualitatively inspect the headspace-gas chromatogram from the instrumental analysis in **Thiyagarajan2015** for a correspondence with the volatile products carbon monoxide and carbon dioxide.







Scheme 15: (top) network pathway for decarboxylation into *o*-toluic acid precursor; (bottom) reaction coordinate diagram.

### 3.5 Global results and outlook

The results of part A in the form of the total energy of accessible intermediates indicate a thermodynamic favorability towards forming the lactone as it possesses the top 1 lowest overall energy of all computed structures.

The  $\beta$ -lactone appears to be too strained under the effect of the cyclohexene confirmation to form between the 1-hydroxyl and 2-acyl group. The 2-COOH motif therefore does not represent any products on the major pathways. Conversely 3-COOH motif is key in the major pathway toward phthalic anhydride. The 1,3-substitution pattern with both substituents axially oriented has been reported by **Kato1977** where the authors conclude that this substitution pattern is needed for lactonization. To omit the lactonization in their total synthesis of fujenoic acid, the researchers proceed with the exocyclic product instead.

In the DFT output files many more descriptors are available *e.g.*, partial charge models Hirschfeld and Mulliken, which can be used in a more programmatic evaluation of sites in a future continuation of this work. The employed manual positioning of the proton has non-convergence issues that may be alleviated by a programmatic approach.

On the basis of the performed calculations, we are interested to scope out the possibility for application of machine learning in proposing the geometry and energy of additional intermediates and of saddle points. We believe that the produced data in this work provides a sufficiently diverse body for training purposes when not only the final SCF cycle, but the relaxation trajectories are included.

One of the shortcomings of the current approach is a lack of expected explicit acid-base assisted reactions in a more refined model to be more representative of those found in Nature.

Not all steps calculated in the +H/-H approach have a corresponding intramolecular transformation due to the distance separating donor and acceptor sites. The selection of transformations with quantified barriers was limited by the approach and not all could be modeled with the transition state search. The possibility for assisted transformations between intermediates is qualitatively described where applicable but lies outside the scope of the current investigation.

WJ van Weesep acknowledges:

Teaching and supporting staff at the department of chemistry, and students & postdocs at ICC and MCC groups.

Specifically, the following individuals for motivation to scientific excellence:

N Artrith and RE Bulo for going above and beyond the supervisory role and providing personal mentorship. They are role models in professional achievements as well as being exemplary educators.

Professor PCA Bruijninx

Professor ETC Vogt

Asst. Professor P Ngene

JM Boereboom

K Draijer

D Evans

J He

P Keizer

O Kerkenaar

M Lauwerse

M Lazemi

F Mohammad

Y Pat

M Roelofs

T Tchakoua

S Simko

K Stanciakova

## Bibliography

- Belgacem2008** J. Polymer Res., 2014, **21**:340
- Chang2014** Green Chem., 2014, **16**, 585–588
- Ciok2021** Green Chem., 2021, **23**, 367–373
- Ennaert2016** Chem. Soc. Rev., 2016, **45**, 584–611
- Genuino2014** ChemSusChem, 2017, **10**, 277 –286
- Gupta2016** RSC Adv., 2016, **6**, 12932–12942
- Holthausen2001** Wiley, 2001, 2nd Ed.
- Houk2008** Angew. Chem. Int. Ed., 2008, **40**, (47), 7592–7601
- Kato1977** J. Chem. Soc. Perkin Trans. 1, 1977, 206–210
- Kucherov2021** ACS Sustainable Chem. Eng., 2021, **9**, (8), 3011–3042
- Lajunen1991** Acta Chem. Scand., 1991, **45**, 578–582
- Lancefield2020** Angew. Chem. Int. Ed., 2020, **59**, (52), 23480–23484
- Li2022** Catal. Sci. Technol., 2022, **12**, 1902–1921
- Mahmoud2014** Green Chem., 2014, **16**, 167–175
- Nikbin2013** J. Catal., 2013, **297**, 35–43
- Reiher2015** J. Chem. Theory Comput., 2015, **11**, 5712–5722
- Stanovnik2001** Adv. in Heterocyclic Chem., 2001, **81**:254.
- Thiyagarajam2014** ChemSusChem, 2014, **8**, 3052–3056
- Thiyagarajam2015** Angew. Chem. Int. Ed., 2015, **55**, 1368–1371
- Williams2014** ACS Catal., 2012, **2**, 935–939
- Zimmermann2017** WIREs Comp. Mol. Sci., 2017, **8**, (2), 1354

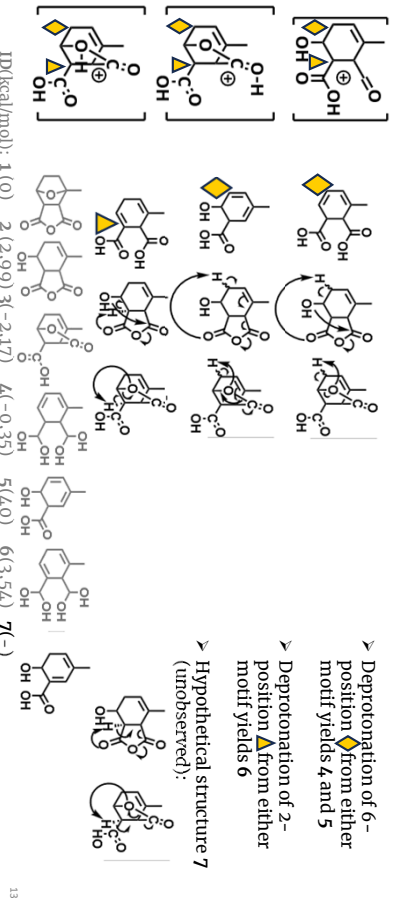








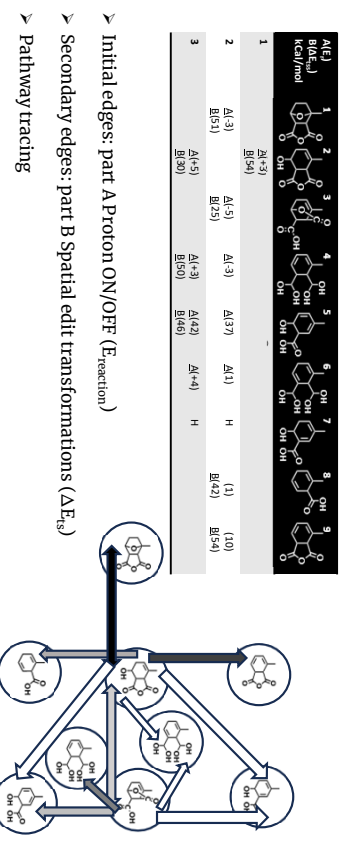
## Results B: Transformations generation 2 & 3



13



## Results B: Edges & Partial Reaction Network



14

- > Initial edges: part A Proton ON/OFF (E<sub>reaction</sub>)
- > Secondary edges: part B Spatial edit transformations (ΔE<sub>ts</sub>)
- > Pathway tracing

14

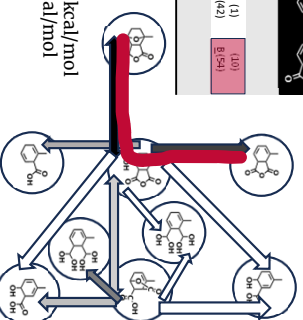


## Results part B: Pathways – direct dehydration

AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9

- > Uphill thermodynamics: E<sub>reaction</sub> = 13 kcal/mol

- > Two high barriers: Ring-opening ΔE<sub>ts</sub> = 54 kcal/mol  
Elimination ΔE<sub>ts</sub> = 54 kcal/mol  
E<sub>activation</sub> = 57 kcal/mol



15

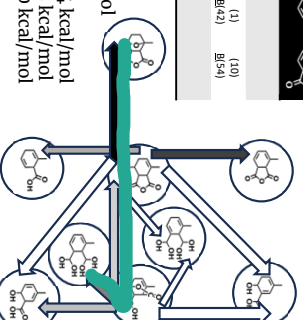


## Results part B: Pathways – via lactone

AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9
AE (E <sub>act</sub> ) kcal/mol	1	2	3	4	5	6	7	8	9

- > Lactone most stable intermediate E<sub>reaction</sub> = -2 kcal/mol

- > Lactonization low barrier: Ring-opening ΔE<sub>ts</sub> = 54 kcal/mol  
Lactonization ΔE<sub>ts</sub> = 25 kcal/mol  
Isomerization ΔE<sub>ts</sub> = 50 kcal/mol  
E<sub>activation</sub> = 54 kcal/mol



16

15

16



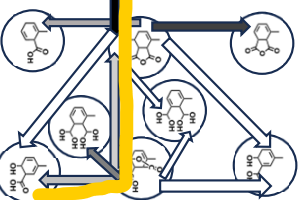


## Results part B: Pathways – *m*-toluic acid

Introduction ● Methods ● Results ● Conclusion

Alt	Alt	Alt	Alt	Alt	Alt	Alt	Alt	Alt
1	2	3	4	5	6	7	8	9
Alt(1) E(54)	Alt(2) E(51)	Alt(3) E(25)	Alt(3) E(37)	Alt(3) E(1)	Alt(4) E(42)	Alt(4) E(54)	Alt(4) E(30)	Alt(4) E(10)
1	2	3	4	5	6	7	8	9

- Lactone most stable intermediate  $E_{\text{reaction}} = -2$  kcal/mol
- Elimination competitive:
  - Ring-opening  $\Delta E_{\text{rs}} = 54$  kcal/mol
  - Lactonization  $\Delta E_{\text{rs}} = 2$  kcal/mol
  - Elimination  $\Delta E_{\text{rs}} = 48$  kcal/mol
  - $E_{\text{activation}} = 54$  kcal/mol
- Elimination at position 2 not calculated due to limitations of the method



17

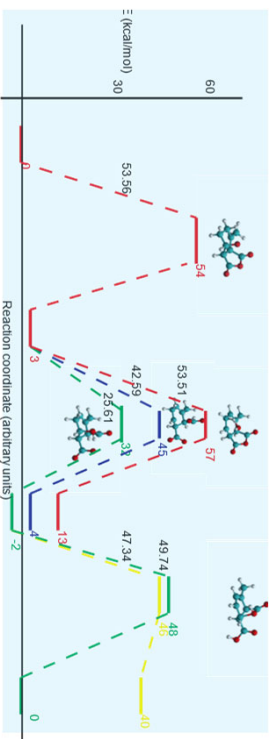
17



## OVERVIEW

Introduction ● Methods ● Results ● Conclusion

- Lactone's favorable kinetics and thermodynamics corroborate the experimental findings



Comparison of reaction profiles for direct dehydration (red); decarboxylation (blue), lactonization-isomerization (green), decarboxylation (yellow)

19

19

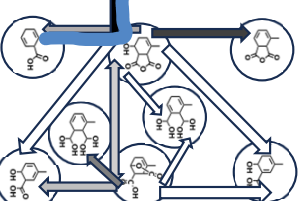


## Results part B: Pathways – *o*-toluic acid

Introduction ● Methods ● Results ● Conclusion

Alt	Alt	Alt	Alt	Alt	Alt	Alt	Alt	Alt
1	2	3	4	5	6	7	8	9
Alt(1) E(54)	Alt(2) E(51)	Alt(3) E(25)	Alt(3) E(37)	Alt(3) E(1)	Alt(4) E(42)	Alt(4) E(54)	Alt(4) E(30)	Alt(4) E(10)
1	2	3	4	5	6	7	8	9

- Decarboxylated product more stable than decarboxylated lactone  $E_{\text{reaction}} = 1$  kcal/mol
- Elimination competes with dehydration but outcompeted by lactonization
  - Ring-opening  $\Delta E_{\text{rs}} = 54$  kcal/mol
  - Elimination  $\Delta E_{\text{rs}} = 42$  kcal/mol
  - $E_{\text{activation}} = 54$  kcal/mol



18

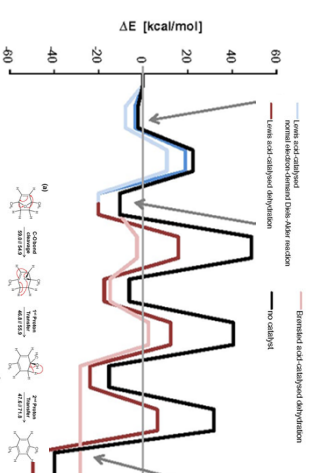
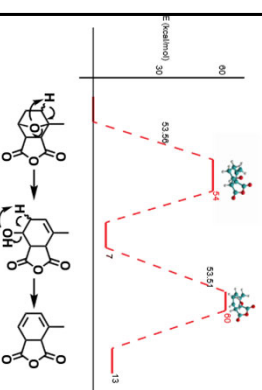
18



## Direct dehydration vs. *p*-xylene

Introduction ● Methods ● Results ● Conclusion

Direct dehydration:



Restricted frozen core B3LYP-TZP Becke97/Good cosmo water (rel. permitt. 78.39) with ADP2014

20

20

## Conclusions & outlook

A *cis* 1,3 -relationship is amenable to a ring closure producing the lactone, which does not form between 1,2 substituents.

The proposed primary intermediate 2 and the lactone 3 are both able to produce doubly unsaturated 4, 2-dicarboxylic acids 4 & 6 as a precursor to final aromatization.

Fragmentation and isomerization are both reported from the same ON-OFF sequence and display similar transition structures that differ chiefly in hydroxyl orientation.

Further investigations towards elucidating the catalyzed mechanism and selectivity control are needed to describe the experimental results

21

21

## Take-home message

Underreported field of polyfunctionalized DA adducts but vital for the transition into biobased building blocks requires optimization for selectivity

A variety of substituents among the oxanorbornane (and oxanorbornene precursor) systems are typical for biomass derived furanics

Exploitation of biobased building blocks for drop -in chemicals and functional or novel replacements

22

22



## Acknowledgments

Supervision by: Nong Arruth, Rosa Bulo  
Daily supervision by: Jelle Boereboom

Examinor: Peter Ngene

Theory @ MCC group: Andrés Borello Mendez

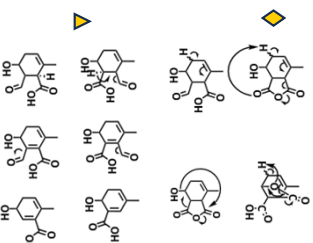
Qijung Che  
Koen Draijer  
Frank de Groot  
Jian He  
Pim Keizer  
Masoud Lazemi  
Fabian Mohammad  
Theophile Tchakoua  
Zixiong Wei  
Xiaotong Xu

MCC teaching & support staff



23

23



➤ breaking of symmetry by the  $\alpha$ -unsaturated 4-methyl substituent enables the decarbonylation of the 3-position more than the 2-position.

24/11/2023, 16:58

24

24