Discovering Monoterpene Catalysis Inside Nano-capsules with Multiscale Modeling and Experiments

Efrat Pahima,a Qi Zhangb Konrad Tiefenbacher,b,c Dan T. Major a,*

a Department of Chemistry, Bar-Ilan University, Ramat-Gan 52900, Israel
b Department of Chemistry, University of Basel, Mattenstrasse 24a, 4058 Basel, Switzerland
c Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 24, 4058 Basel, Switzerland

KEYWORDS: QM/MM simulations, monoterpenes, nano-capsules, nano-reactor, cation–π interactions

ABSTRACT: Large-scale production of natural products, such as terpenes, presents a significant scientific and technological challenge. One promising approach to tackle this problem is chemical synthesis inside nano-capsules, although enzyme-like control of such chemistry has not yet been achieved. In order to better understand the complex chemistry inside nano-capsules, we design a multiscale nano-reactor simulation approach. The nano-reactor simulation protocol consists of hybrid quantum mechanics-molecular mechanics-based high temperature Langevin molecular dynamics simulations. Using this approach we model the tail-to-head formation of monoterpenes inside a resorcin[4]arene-based capsule (capsule I). We provide a rationale for the experimentally observed kinetics of monoterpene product formation and product distribution using capsule I, and we explain why additional stable monoterpenes, like camphene, are not observed. Based on the in-capsule I simulations, and mechanic insights, we propose that feeding the capsule with pinene can yield camphene, and this proposal is verified experimentally. This suggests that the capsule may direct the dynamic reaction cascades by virtue of π-cation interactions.

Introduction

Terpenes are natural products derived biosynthetically from isoprene units. The basic molecular formula of terpenes is (C₅H₈)n, where n is the number of isoprene units linked together in the form of linear chains or rings to form monoterpenes (n=2), sesquiterpenes (n=3), diterpenes (n=4), etc. Terpenes constitute one of the largest classes of natural products, and possess great structural diversity.3–7 Terpenes also have a broad range of biological activities, such as antibacterial and anti-inflammatory activities,8–10 and are used in treatment of human diseases, such as malaria11 and cancer.12–15 Terpenes also serve in the defense of many plants, animals and microorganisms against predators, pathogens and competitors.6,8–9,14–15 and more.5,6

In nature, terpene chemistry follows a common mechanism, involving a sequence of electrophilic cyclizations, rearrangements, methyl transfers and H⁺/H⁻ transfers via highly reactive carbocation intermediates. These intermediates must be guided in the correct reaction direction and protected from nucleophilic attacks and premature quenching, in order to minimize side product formation. This extremely challenging task is performed by terpene synthases (or cyclases), which are able to bind the acyclic terpene substrates in a product-like conformation, and guide the highly reactive carbocations towards the target product via specific electrostatic interactions.7,16–22 Additionally, terpene synthase reactions are driven by the inherent reactivity of carbocations.23

Terpene cyclases can be divided into two classes, which differ in the way the cationic cyclization cascade is initiated.24–25 In class I terpene cyclases, an allylic cation is generated by heterolytic cleavage of a diphosphate leaving group with the help of divalent metal ions (e.g. Mg²⁺), in a so-called Tail-to-Head Terpene (THT) cyclization.7,17,22,26–28 The activation at the tail end leads to the formation of a susceptible allyliccation, which is then chaperoned by the terpene cyclase towards the final product in a selective fashion.7,29 In solution, it has proven to be difficult to attain reaction specificity with THT synthesis, due to premature quenching by nucleophilic attack or elimination. Therefore, artificial catalysts able to perform selective THT cyclizations are lacking.30–34 Class II terpene synthases utilize a Head-to-Tail Terpene (HTT) cyclization strategy,35–36 where cyclization is initiated by direct protonation of a double bond or an epoxide. In contrast to THT, HTT chemistry has been intensively studied, and has been successfully reproduced in bulk solvent.37–41

Production of terpenes via bioengineering of metabolic pathways within microbial hosts like E. coli and S. cerevisiae were studied within the context of advanced biofuels synthesis.42–47 However, substantial yields were not obtained, although current research directions are constantly evolving and improving.28,46–53
Supramolecular capsules constitute an alternative to standard test-tube or microbial syntheses of terpenes. Molecular capsules provide means to mimic the basic working principles of the structurally complex natural enzymes by enabling the isolation of suitable substrates from bulk solvent in a hydrophobic cavity (i.e. the capsule). For example, in the resorcin[4]arene-based capsule (Fig. 1), the carbocation intermediates inherent to terpene chemistry are stabilized via cation–π interactions. The importance of cation–π interactions in nature were recognized by Kebarle and co-workers, Diederich and co-workers, Burley and Petsko, and Dougherty and co-workers. The interaction between cations (i.e. monopoles) and π-electron systems (i.e. quadrupoles) is considered to be largely electrostatic in nature. These interactions are ubiquitous in nature and are present in ion channels, receptors, enzymes, and more. For instance, cation–π interactions have been found to be important in terpene synthases. Artificial systems that employ cation–π interactions are widespread, such as supramolecules in aqueous media, organic solvents and solid state. A large variety of capsular host systems constructed from self-assembled monomers held together by metal–ligand interactions, hydrogen bonds, or by hydrophobic effects have been described in recent decades. The catalytic activity of host systems can stem from either the encapsulation of a chemically active species (e.g. a transition-metal catalyst) or the intrinsic properties of the cavity space (non-covalent interactions between host and guest, such as cation–π interactions).

The resorcin[4]arene hexamer (Fig. 1, 2) spontaneously forms in apolar solvents, like chloroform, from six resorcin[4]arene monomer units and eight water molecules. The capsule-like structure, held together by 60 hydrogen bonds, forms an octahedral-shaped cavity of about 1375 Å³. Due to extended cation–π interactions with the aromatic cavity, positively charged intermediates, like carbocations in a THT mechanism, display high affinity for the capsule interior. Besides the capability of reversible guest encapsulation, the resorcin[4]arene hexamer acts as relatively strong phenol-based Brønsted acid (pKa≈5.5–6). Nevertheless, for many reactions, including terpene cyclizations, a Brønsted acid cocatalyst (e.g. HCl) is required. Recent work of Tiefenbacher and co-workers focused on reactions using the precursors geranyl, linalyl and neryl, each with different leaving groups. By studying the influence of the leaving group on the product selectivity and product yield, as well as the relation of the leaving group affinity to the capsule walls, these authors managed to shed light on the mechanism of the reaction inside the capsule. Using this protocol, in addition to standard QM calculations, we obtain mechanistic insight into the many reactions occurring inside capsule I.

**Computational Details**

**Gas-Phase Calculations.** To gain in-depth understanding of the reaction mechanism occurring inside the resorcinarene capsule, it is important to understand the inherent energetics of the chemistry involved. Hence, density functional theory (DFT) gas phase calculations were performed, to shed light on stability, the possible reaction pathways and the energetic profiles of all products observed experimentally (Fig. 3, S2).

Prior to the DFT gas phase calculations, conformational searches were performed in order to determine the lowest energy conformers for each monoterpane, using Schrodinger’s Maestro within the Maestro modeling environment. Relative potential energies were determined using the OPLS force field.

Subsequent DFT geometry optimizations (minima and transition states, TS) and frequency calculations of the monoterpenes were performed in the gas phase using Gaussian 16, revision A.0391 (see Supporting Information Scheme S1, Fig. S1 for all compounds included in the calculations). We employed the meta-hybrid GGA functional M06-2X, which has been shown to be accurate for thermochemical properties, and also gives reliable results for carbocation chemistry relevant to terpenes. The basis set used was 6-31+G**. Normal modes of vibration for the optimized molecules were inspected to verify the nature of the stationary points. To confirm that a given TS connects...
between reactant and product wells, we performed intrinsic reaction coordinate (IRC) calculations. The free energy profiles reported herein were computed using electronic energy combined with zero-point energy and thermal corrections, using statistical mechanics expressions within the harmonic approximation. Entropy errors resulting from not treating free rotations explicitly are ~0.5 kcal/mol per freely rotatable bond (at room temperature), and may be ignored for this study. Possible reactions occurring inside the resorcinarene capsule are summarized in Scheme 1.

Resorcinarene Capsule Modeling. The initial coordinates for the capsule simulations were obtained from the crystal structure of the resorcinarene capsule 1 (Fig. 1). All calculations were performed with methyl groups replacing the R=C_{11}H_{23} chains. The complete capsule 1 nano-reactor system includes the following (Fig. 2): The resorcinarene capsule, composed of six resorcinarene monomer units and 8 water molecules, one HCl molecule (acid catalyst), 5 or 6 chloroform solvent molecules (further details below), and a substrate (either GOAc, NOAc, or LOAc). The resorcinarene capsule, water, and solvent molecules were treated with MM. Force field parameters and residue topology were developed based on existing CHARMM FF C22 and C36 parameters. Water molecules were treated by the TIP3P model. Chloroform parameters were adopted from CHARMM general FF (CGenFF) and the SHAKE algorithm was employed to constrain the solvent geometry. All parameters and residue topology files can be found in the Supporting Information. Most simulations were performed for each substrate in both 5 and 6 solvent molecules capsule 1 systems, since it is unknown how many solvent molecules occupy the capsule when the substrate is present. Indeed, we assume both scenarios are plausible.

**QM/MM In-Capsule Chemical Modeling.** The substrate (either GOAc, NOAc, or LOAc), and in some cases also HCl, were treated using the M06-2X functional, with the following basis set combination: the carbon and hydrogen atoms were treated with the 6-31G* basis set, the oxygen and chlorine atoms were treated with 6-31+G*. The interactions between the QM and MM regions were treated using electrostatic embedding. No cutoff values were employed in computing the MM or QM/MM interactions.

All simulations in the capsule system were performed using a QM/MM scheme that combines the CHARMM program and Q-Chem. QM/MM Nano-Reactor Simulations Protocol Overview. The simulations performed on capsule 1 include minimization and LMD heating, equilibration and production simulations, with nano-reactor simulation protocol settings mimicking the experimental conditions (see similar nano-reactor simulations in Ref. 109). The experiments of Tiefenbacher and co-workers were performed at 30 °C in CDCl₃ with a substrate concentration of 33.3 mM, 3 mol% DCI and 10 mol% capsule. The results were collected over a 72 hours period of time. With today's computational resources, such long LMD QM/MM simulations are not realistic, as they are too expensive. In order to observe reactive events occurring in reasonably short simulation times, we employed a HT-LMD strategy. The simulations commenced by slowly heating the system from 300K to 1200K using a MM PES. Formally, the temperature for our LMD simulations may be defined as 300+(100×n) K, with n running from 0 to 9. In practice, we raised the temperature by 100K every 25 ps simulation time, resulting in an overall simulation time of 0.25 nano seconds. Restraints were added to the capsule to keep it intact at elevated temperatures (further explanation about these restraints are presented in the following section and in the Supporting Information).

**Chart 1.** Illustration of the QM/MM nano-reactor simulations protocol.

<table>
<thead>
<tr>
<th>MM LMD</th>
<th>QM/MM LMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>Simulation</td>
</tr>
<tr>
<td>Heating-Equilibration MD</td>
<td>Equilibration-Production MD</td>
</tr>
<tr>
<td>25 ps</td>
<td>0.2 ns</td>
</tr>
<tr>
<td>300K</td>
<td>300K</td>
</tr>
<tr>
<td>400K</td>
<td>400K</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1200K</td>
<td>1200K</td>
</tr>
</tbody>
</table>

Following this heating period, we switched to a QM/MM PES. Each temperature step (i.e. 300K, 400K, ..., 1200K) was further simulated using QM/MM LMD, starting
from the MM heating restart conditions. The simulation time for each QM/MM LMD simulation was 0.2 ns. At each temperature, we performed two simulations for each substrate, using 5 or 6 solvent molecules. It was expected that chemical reactions would occur at elevated temperatures, where the system has sufficient kinetic energy to cross activation barriers, in spite of the short simulation times. This intuitive expectation is based on the known high reactivity of carbocations. Further details about these simulations are illustrated in Chart 1 and explained in the following section.

**QM/MM Nano-Reactor Simulation Protocol Details.** The equations of motion were integrated numerically using Langevin dynamics. The Langevin equation was employed for all atoms beyond 7 Å of the origin (center of capsule). In practice, this means that all capsule **I** atoms (but not solvent, HCl, or substrate) feel a friction force (with a friction coefficient of 7 ps\(^{-1}\)) and also experience an occasional random "kick" to maintain the desired target temperature during the simulation. The atoms inside the capsule (within 7 Å) are thermostatted via stochastic collisions with the capsule atoms. Thus, the Langevin atoms (i.e. capsule atoms), are subjected to NVT dynamics (constant particle-volume-temperature), while the inner atoms (solvent, HCl, and terpene) are formally under NVE conditions (constant particle-volume-energy) dynamics. However, in practice the inner atoms are subjected to NVT dynamics by virtue of their collisions with the capsule.

The thermal motion rapidly breaks apart non-covalent interactions at high temperatures, thus the capsule itself must have restraints so as to not break apart. Nuclear Overhauser effect (NOE) restraints were applied to the capsule's hydrogen bonds to assure that the capsule will not disintegrate during the HT-LMD. The simulations employed the Leap-Frog integration scheme with a time step of 1 fs.\(^{107}\)

**Experimental Details**

**Transformation of β-pinene catalyzed by the resorcinarene capsule I.** Procedure for the cyclization reaction. Chloroform was filtered through basic aluminum oxide prior to usage. To a solution of the resorcinarene capsule **I** (11.1 mg, 1.67 μmol, 0.10 eq) in chloroform (200 μL) were added successively HCl stock solution in chloroform (0.50 μmol, 0.03 eq) and n-decane stock solution in chloroform (20 μL, 167 mmol/L, 3.34 μmol, 0.2 eq). Then additional chloroform was added to reach a total volume of 500 μL for the reaction mixture. After β-pinene (2.63 μmol, 16.7 μmol, 1.00 eq) was added, the reaction mixture was briefly agitated. An aliquot (approx. 10 μL) of the reaction mixture was diluted with 0.2 mL n-hexane and subjected to GC analysis (initial sample). Meanwhile, the reaction was kept at the given temperature (30 °C, 40 °C and 50 °C, respectively). After the indicated time, the reactions were sampled as described above. Cyclization products were identified by comparing the NMR and GC spectra of the reaction mixtures with those of the authentic monoterpene samples. Conversions and yields were calculated as described in Ref. 2. For the preparation of resorcinarene capsule **I** and HCl stock solution, see Ref. 110.
Results

The following section is divided into three parts: (1) Extensive DFT study of the free energy of monoterpenes (the experimentally observed products in capsule I and other known monoterpenes) in the gas-phase; (2) QM/MM HT-LMD simulations, mimicking the carbocation chemistry occurring inside the resorcinarene capsule; (3) Further gas-phase calculations of the carbocation intermediates observed in the HT-LMD simulations.

(1) Thermodynamics of monoterpenes. In order to understand the inherent energetics of the chemistry involved we performed DFT gas-phase calculations at the M06-2X/6-31+G** level. The question guiding us in this part is to what extent the experimental product distribution observed from capsule I synthesis is dictated by the intrinsic thermodynamics of monoterpenes.

Fig. 3 and S2 compare the relative computed free energies of an assortment of known monoterpenes (1-22), as well as novel monoterpenes generated by using some of the carbocation structures found in the study of Jacobson et al.111 The monoterpenes predicted to be most stable are 1, 2 (fenchyl-cation derivatives) and α- and β-camphene (bornyl cation derivatives), followed by experimentally observed α- and γ-terpinene, and additional monoterpenes that possess a terpinyl carbon skeleton (< ~5 kcal/mol of global minimum). Seemingly the relative stability of the monoterpenes is a combination of the number and nature of chemical bonds (σ, π, conjugated), ring strain, and steric crowding.

This computed monoterpenes free energy landscape may be compared with the products observed experimentally via in-capsule I synthesis.112 The in-capsule I synthesis yielded the following main monoterpenes products: α-terpinene, terpinolene, isoterpinolene, γ-terpinene, and limonene (Fig. 1). Hence, the observed monoterpenes are all within a few kcal/mol of the global monoterpenes minimum. The computed order of stability is α-terpinene (most stable) > γ-terpinene > isoterpinolene > terpinolene > limonene, while experimentally the order is (using NOAc as substrate) α-terpinene > terpinolene > isoterpinolene > γ-terpinene > limonene. We note that the experimental data reported here are those after 72 hours, at which time the products had not yet reached a plateau. Indeed, at 72 hours the amounts of α-terpinene, γ-terpinene, and isoterpinolene were increasing, while the amounts of terpinolene and limonene were decreasing. In particular, the amount of terpinolene was decreasing sharply. Hence, the product distribution seems to be moving towards the thermodynamic partitioning. To validate the accuracy of our DFT approach, we also compared the relative stability of all monoterpenes using the highly accurate CBS-QB3 and M06 methods.113-114 These results confirm the general trends predicted by DFT (Table S4). We may therefore conclude that the in-capsule I synthesis preferably produces low-energy monoterpenes. However, important questions do remain unanswered, namely how come additional low-energy monoterpenes such as α- and β-camphene are not observed experimentally? Under what conditions can they form in the capsule? Additionally, what role does the initial substrate play in the experimental product distribution? To answer these questions, we performed in-capsule I simulations.

(2) In-capsule I HT-LMD QM/MM Simulations. As explained in the Computational Details section, the interactions in the capsule/substrate/chloroform system are treated via a hybrid QM/MM potential, while the dynamics of the system is treated via HT-LMD simulations. For all systems, the room temperature simulations did not produce chemical events during short trial simulations, and below 500K not many reactions occurred (not shown). Hence, our discussion will focus on the HT simulations. Fig. 4,a-c presents the intermediates observed during the simulations performed at a range of temperatures from 500K to 1200K. The results are for three substrates, GOAc, NOAc, and LOAc, and the systems include 6 chloroform molecules. Results for identical systems, but with 5 chloroform molecules, are presented in Fig. S3,a-c.

As expected, the number of chemical events greatly increases with temperature. Beyond ~1100-1200 K, we observe higher energy chemistry that includes fragmentation of the carbocation skeleton structure (which rearranges back to yield different carbon frameworks) and creation of carbanions. Seemingly, the optimal temperature for these monoterpenes HT-LMD simulations is ~900K-1000K. For the purpose of the discussion below, we will treat all HT simulations (T ≥ 500K) with 5 or 6 chloroform solvent molecules as an ensemble of trajectories that reflect what occurs during experiments. Due to the limited simulation time, the diverse product distributions obtained during the various simulations (different temperatures and number of
solvent molecules) do not necessarily reflect meaningful statistical differences between the in-silico simulation conditions. Yet, inspection of the global carbocation pool observed during all simulations, can provide qualitative insight into what occurs in a wet-lab setting.

The in-capsule reactions commence with protonation of the ester oxygen of GOAc, NOAc, and LOAc. Initial C-O heterolytic bond dissociation takes places within the first few ps of the simulations for all T ≥ 500K. This first step results in an allyl cation, which in the case of NOAc is in a reaction-ready Z-conformation B, while for GOAc is in a non-reactive E-conformation A (Scheme 1). In the case of LOAc, the allyl cation conformation obtained depends on the conformation around the C2-C3 bond at the time of C-O bond cleavage. These differences in the substrates in capsule I are reflected in the time required for initial cyclization to form the first intermediate α-terpinyl cation, C (Fig. 4.a-c). In GOAc, the first instance of C formation was observed at 900K, while for NOAc, we observed its formation already in the simulation at 500K. In the case of LOAc, we observe C formation at 900K. Interestingly, we observe that the reaction-ready cation B has a distinct life-time of several psec, which suggests that formation of cation C from the initial substrate is only a partially concerted process. This is in contrast to what has been suggested based on experiments, but is in line with what we have observed in numerous simulations for class I terpene synthases.

On the other hand, the GOAc reaction is distinctly step-wise.

Once cation C is formed, a rich assortment of reactions may take place (Scheme 1, Figures 4.a-c, 5). Direct deprotonation of C would give limonene or γ-terpinene (deprotonations not modeled, as capsule atoms are defined as MM). A 1,2-hydride transfer gives terpinene-4-yl cation, D, which upon deprotonation would produce α-terpinene, γ-terpinene, or terpinolene (Scheme 1). A subsequent 1,2-hydride transfer produces the allylic terpinene-3-yl cation E (E can also form via a 1,3-hydride shift from C). Deprotonation of E would yield α-terpinene or α- or β-phellandrene. Additional hydride shifts are possible, but were not observed in our short simulations. The direct carbocation precursor to isoterpinolene can be formed by a 1,5-hydride transfer or deprotonation/reprotonation mechanism from cation E, and in our simulations was formed via deprotonation/reprotonation by carboxylic acid.

Cation C also underwent electrophilic attack by the C2-C3 double bond to form a pinyl cation, G, which rearranged to give a bornyl cation, H (but not cation M, see Fig. 5.c). Cation H immediately rearranged to yield a tertiary camphyl cation, I. G can give the high energy monoterpenes α- and β-pinene, H can result in bornene or further rearrange to yield α-camphene, while I can result in β-camphene. These results suggest that a variety of terpene products potentially could form in the capsule. However, many products were not observed experimentally, and we will discuss the reason for this below.

Cation D may also undergo electrophilic addition to give sabinyl cation J, which could be deprotonated to give sabine or α-thujene. Methylene rearrangement results in K, and a propyl ring opening of K can give an additional terpinyl cation L, which can give α-, β-, or γ-terpinene upon deprotonation.

The in-capsule I reactions commence with protonation of the ester oxygen of GOAc, NOAc, and LOAc. Initial C-O heterolytic bond dissociation takes places within the first few ps of the simulations for all T ≥ 500K. This first step results in an allyl cation, which in the case of NOAc is in a reaction-ready Z-conformation B, while for GOAc is in a non-reactive E-conformation A (Scheme 1). In the case of LOAc, the allyl cation conformation obtained depends on the conformation around the C2-C3 bond at the time of C-O bond cleavage. These differences in the substrates in capsule I are reflected in the time required for initial cyclization to form the first intermediate α-terpinyl cation, C (Fig. 4.a-c). In GOAc, the first instance of C formation was observed at 900K, while for NOAc, we observed its formation already in the simulation at 500K. In the case of LOAc, we observe C formation at 900K. Interestingly, we observe that the reaction-ready cation B has a distinct life-time of several psec, which suggests that formation of cation C from the initial substrate is only a partially concerted process. This is in contrast to what has been suggested based on experiments, but is in line with what we have observed in numerous simulations for class I terpene synthases.

On the other hand, the GOAc reaction is distinctly step-wise.

Once cation C is formed, a rich assortment of reactions may take place (Scheme 1, Figures 4.a-c, 5). Direct deprotonation of C would give limonene or γ-terpinene (deprotonations not modeled, as capsule atoms are defined as MM). A 1,2-hydride transfer gives terpinene-4-yl cation, D, which upon deprotonation would produce α-terpinene, γ-terpinene, or terpinolene (Scheme 1). A subsequent 1,2-hydride transfer produces the allylic terpinene-3-yl cation E (E can also form via a 1,3-hydride shift from C). Deprotonation of E would yield α-terpinene or α- or β-phellandrene. Additional hydride shifts are possible, but were not observed in our short simulations. The direct carbocation precursor to isoterpinolene can be formed by a 1,5-hydride transfer or deprotonation/reprotonation mechanism from cation E, and in our simulations was formed via deprotonation/reprotonation by carboxylic acid.

Cation C also underwent electrophilic attack by the C2-C3 double bond to form a pinyl cation, G, which rearranged to give a bornyl cation, H (but not cation M, see Fig. 5.c). Cation H immediately rearranged to yield a tertiary camphyl cation, I. G can give the high energy monoterpenes α- and β-pinene, H can result in bornene or further rearrange to yield α-camphene, while I can result in β-camphene. These results suggest that a variety of terpene products potentially could form in the capsule. However, many products were not observed experimentally, and we will discuss the reason for this below.

Cation D may also undergo electrophilic addition to give sabinyl cation J, which could be deprotonated to give sabine or α-thujene. Methylene rearrangement results in K, and a propyl ring opening of K can give an additional terpinyl cation L, which can give α-, β-, or γ-terpinene upon deprotonation.

On the other hand, the GOAc reaction is distinctly step-wise.

Once cation C is formed, a rich assortment of reactions may take place (Scheme 1, Figures 4.a-c, 5). Direct deprotonation of C would give limonene or γ-terpinene (deprotonations not modeled, as capsule atoms are defined as MM). A 1,2-hydride transfer gives terpinene-4-yl cation, D, which upon deprotonation would produce α-terpinene, γ-terpinene, or terpinolene (Scheme 1). A subsequent 1,2-hydride transfer produces the allylic terpinene-3-yl cation E (E can also form via a 1,3-hydride shift from C). Deprotonation of E would yield α-terpinene or α- or β-phellandrene. Additional hydride shifts are possible, but were not observed in our short simulations. The direct carbocation precursor to isoterpinolene can be formed by a 1,5-hydride transfer or deprotonation/reprotonation mechanism from cation E, and in our simulations was formed via deprotonation/reprotonation by carboxylic acid.

Cation C also underwent electrophilic attack by the C2-C3 double bond to form a pinyl cation, G, which rearranged to give a bornyl cation, H (but not cation M, see Fig. 5.c). Cation H immediately rearranged to yield a tertiary camphyl cation, I. G can give the high energy monoterpenes α- and β-pinene, H can result in bornene or further rearrange to yield α-camphene, while I can result in β-camphene. These results suggest that a variety of terpene products potentially could form in the capsule. However, many products were not observed experimentally, and we will discuss the reason for this below.

Cation D may also undergo electrophilic addition to give sabinyl cation J, which could be deprotonated to give sabine or α-thujene. Methylene rearrangement results in K, and a propyl ring opening of K can give an additional terpinyl cation L, which can give α-, β-, or γ-terpinene upon deprotonation.
Hence, a variety of carbocations (A-L) are formed during our HT-LMD simulations, suggesting that not only the experimentally observed monoterpenes can form via their carbocation precursors C-F, but also other monoterpenes such as the α, β-camphene, α, β-phellandrene, sabine and α, β-pinene. Importantly, carbocation precursors for all products observed from in-capsule I experiments (i.e. α-terpinene, terpinolene, isoterpinolene, γ-terpinene, and limonene) are seen in our QM/MM HT-LMD simulations, and this is a key finding of this study.

Now we turn back to a question left unanswered from the previous part. We see various intermediate precursors to other potential low energy products in our simulations, posing the question how come α, β-camphene, or phellandrene are not observed in the experimental work? For this reason, we return to DFT gas phase calculations of the detailed mechanism to study the carbocation chemistry observed in the HT-LMD simulations.

(3) Free Energy Profiles for Monoterpene Formation. The current approach allows a peek into the real-time dynamics of the terpene chemistry occurring inside capsule I. However, the simulations aren’t sufficiently long to explain the detailed product distribution. Indeed, to explain the experimentally observed product distribution would require either very long simulations, or a swarm of parallel simulations to collect significant statistics for the observed chemical events. In order to gain quantitative insight into the thermodynamic details of the product distribution, we turn to mechanistic DFT calculations along pre-defined reaction paths that lead to products suggested by Tiefenbacher and co-workers and as seen in our HT-LMD

Figure 5. Free energy profile (kcal/mol) for a model reaction for the monoterpene pathway, starting from (A) 2,3 transoid-cation. Blue curves represent the mechanism suggested by Tiefenbacher and co-workers,1-2 grey lines represent alternative mechanisms observed in our QM/MM HT-LMD simulations and intermediates that were less favorable energetically. The calculations were performed at the M06-2X/6-31+G** level of theory.
We note that this analysis is similar to our previous work on monoterpenes and that of Hong and Tantillo.\textsuperscript{10, 121}

In principle, we could compute the free energy barriers inside the capsule, using techniques such as potential of mean force, but such simulations are rather expensive. In the current case, we presume it is sufficient to model the reactions in gas-phase, as the effect of the capsule is expected to be similar for all reaction paths. As will be discussed below, this assumption might not hold for all cases.

As shown in Fig. 5, the transoid A to cisoid B linalyl cation direct rotational transition suggested by Tiefenbacher et al.\textsuperscript{12} has a relatively high free energy barrier of 10.5 kcal/mol. However, this rotational barrier can be significantly reduced if cation A is partially recaptured by acetic acid. Indeed, this is the strategy adopted in class I terpene cyclases.\textsuperscript{112–114} Inspection of Fig. S1 and Scheme S1, the transition going through LOAc has lower activation free energies (A to LOAc and LOAc to B have activation free energies of 1.0 kcal/mol and 4.9 kcal/mol, respectively). However, both direct and acetic acid-assisted pathways were observed in the QM/MM HT-LMD simulations.

B is more stable than A by 11.1 kcal/mol, which may be due to the folded structure of B, with a short C1 and C6 distance (1.65 Å). This state of B resembles the α-terpinyl cation, C, (C1-C6 bond length: 1.53 Å), but with the positive charge resonating between the C1-C2-C3 carbons and the C6-C7 π-system. Following generation of B, terpinyl cation C is readily produced via C1-C6 bond formation (-15.0 kcal/mol). The following C to D 1,2,4-hydrate shift has a 4.9 kcal/mol activation free energy. C and D are nearly isoergic (reaction free energy of -1.1 kcal/mol), as both consist of a similar tertiary carbocation. Based on our QM/MM HT-LMD simulations, we discovered an additional plausible pathway from C to D, going through the stable allylic carbocation terpinen-3-yl-cation, E, with a -10.7 kcal/mol reaction energy for the C to E reaction and an activation free energy of 7.5 kcal/mol. The subsequent transition from E to D is endergonic (free energy barrier of 13.9 kcal/mol and reaction free energy of 9.6 kcal/mol). However, the E carbocation can also lead to α-terpinene, which is the most abundant product in Tiefenbacher and co-worker’s experiments.\textsuperscript{115, 118} Considering the lower barrier of the C→D transformation, this is the more likely pathway.

Pinyl cation, G, results from C via TS, with a free energy barrier of 5.8 kcal/mol, which is slightly higher than the C→D transition, and with an endergonic reaction free energy of 3.8 kcal/mol. Deprotonation of cation G can lead to the α- or β-pinene product. Earlier work has shown that bornyl cation, H, is a high-energy secondary carbocation (transition state).\textsuperscript{19, 121} Here we compute the activation free energy for the G→H transition to be 6.7 kcal/mol (Fig. 5, 6). Similarly, fenchyl cation, M, which has also been shown to be a first-order saddle point,\textsuperscript{19, 121} can form from G with a similar activation free energy (7.0 kcal/mol). Interestingly, both H and M have been shown to be bifurcation points,\textsuperscript{19} and can branch in the directions of camphyl (I) and P, and cations N and O, respectively.\textsuperscript{18, 29, 115, 118} Camphyl cation is a precursor for monoterpenes α- and β-camphene, while fenchyl cation is a precursor for α- and β-fenchene (Fig. 3). Additionally, bornyl cation H can form the high-energy cation P (precursor for 41),\textsuperscript{19} while fenchyl cation M can form low-energy cation N (pre-cursor for 1 and 2). Our earlier gas-phase work\textsuperscript{18} suggested that both bornyl H and fenchyl M cations should be formed with equal probability in the absence of a capsule, suggesting we should observe low-energy cations I, N, and O in our HT-LMD simulations. We do observe camphyl cation, I, (and the preceding bornyl cation, H), but do not observe fenchyl cation, M, or N and O in our HT-LMD simulations. In order to test these assumptions, i.e. H and M should form with equal probability, we performed further QM/MM HT-LMD simulations (Fig. S4). By using the initial 900K restart conditions for pinyl cation, we implemented an algorithm for changing the momenta of all atoms in the simulation via random kicks, similar to transition path sampling.\textsuperscript{125} Hence, we sampled additional regions of phase space in the vicinity of pinyl cation in an attempt to observe fenchyl cation M. The results are presented in Figure S4, and show that no fenchyl cation was observed.

Additionally, we calculated the average distance of the following carbocations from the capsule’s walls: pinyl, G, camphyl, I, α-terpinyl, C, bornyl, H and fenchyl, M. All distances were measured from the charged carbon to the nearest resorcinarene monomer. Specifically, the distances were defined from the cation to the three carbon atoms of the four phenyl rings, connected to the OH groups (named CZ1, CZ2, CD2 in the SI rtf files). See Fig. 7a,b and SI Table S5 and Fig S6 for further details. We found that pinyl cation G’s C3 position is 5.4 Å (±0.8 Å), camphyl C’s C3 is 5.8 Å (±0.7 Å) and α-terpinyl cation C C7 is 5.4 Å (±1.0 Å) from the capsule’s walls. To probe the distance between the secondary carbocations bornyl, H, and fenchyl, M, and the capsule wall, special distance restraints were imposed on the bonds to avoid rearrangement (further information can be found in the SI). Analysis of the MD trajectories revealed that the bornyl C2 ensemble averaged distance from the capsule

![Figure 6](https://example.com/figure6.png) Possible carbocation pathways in the gas-phase (blue arrows) and in capsule I (red arrows). Length of arrows correspond to probability of step (see kinetic and thermodynamic data in Figure 5).
Wall is 5.6 Å (±1.0 Å), and the fenchyl C2 distance from the wall is 5.9 Å (±0.5 Å) (Fig. 7a,b). We ascribe this difference to the greater steric hindrance in fenchyl cation, due to the additional bulky propylene group in the position adjacent to the cation. Hence, we conclude that an unhindered secondary carbenium, H, may be preferentially stabilized by capsule I. Since both bornyl and fenchyl cations are bifurcation points, we consider this capsule I stabilization of the bornyl pathway a dynamic effect.

To probe whether these in silico observations are reproducible in the lab, we performed in-capsule experiments by feeding the capsule with β-pinene (precursor for cation G, see Fig. 6). The rationale for using β-pinene as a starting material is that the effective barrier to camphene formation is smaller than when starting from a linear precursor. In this case, we observe formation of significant amounts of camphene, yet no fenchene was formed (Table 1, Fig. S5). We also observe formation of limonene, terpinolene, α- and γ-terpinene, which are a result of the ring-opening of pinyl cation, G, to terpinyl cation, C. Hence, there is full agreement between experiments and theory. We speculate that the interaction between the capsule π-system and the secondary cation carbon in bornyl, H, is stronger than for fenchyl cation, M, due to greater steric hindrance in proximity of the cation in the latter. This is supported by the ensembled averaged distance presented above. Hence, bornyl cation may be preferentially stabilized by the walls of the capsule. This suggests that the capsule can direct the dynamic reaction cascade in the direction G→H, away from G→M, by virtue of π-cation interactions. We have previously shown that the enzyme bornyl diphosphate synthase can similarly modulate bifurcation behavior in monoterpene systems.19

Sabinyl cation, J, results from D via TS_D-J with a 1.9 kcal/mol activation barrier and an exergonic reaction free energy (−6.5 kcal/mol), and can lead to sabinene or α-thujene. Carbocation K is formed by a methylene transfer from C6 to C3, and might proceed to cation L via an endergonic reaction (8.7 kcal/mol), which can lead to various terpinenes. Although cations J and K are relatively stable, the corresponding monoterpenes (sabinene, α-thujene, 25, 27) are not (Fig. 3), and thus these are not observed experimentally although their carbocation precursors may exist in capsule I.

**Figures 7a,b.** Cation-π interactions between monoterpene carbocations and capsule I. The left figure (a) shows bornyl cation, I, and the right figure (b) shows fenchyl cation, M.

**Table 1:** Experimental data for the capsule I-catalyzed conversion of β-pinene as substrate (standard reaction conditions, for more information see SI). The reaction was performed at three different temperatures (30-50 °C).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30 °C</td>
<td>19%</td>
<td>5%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>40 °C</td>
<td>72%</td>
<td>24%</td>
<td>12%</td>
<td>5%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>50 °C</td>
<td>99%</td>
<td>33%</td>
<td>18%</td>
<td>9%</td>
<td>5%</td>
<td>2%</td>
</tr>
</tbody>
</table>

[a] Conversion after 41h, quantified by gas chromatography; [b] Yields of products after 41h; products were identified by gas chromatography and 1H NMR spectroscopy and quantified by gas chromatography.
Discussion

In this section we will discuss two important questions relating to natural product synthesis inside nano capsules:

1. Observed product distribution. Specifically, what dictates the product selectivity inside capsule 1?
2. Influence of substrate. Specifically, can we explain the difference in the product distribution between the substrates geranyl, neryl and linalyl acetate?

First, we tackle the product selectivity question. Specifically, are the experimentally observed monoterpenes also the lowest energy ones? For this reason, we performed DFT M06-2X gas phase calculations for an assortment of monoterpenes (Fig. 3, S2). We found that the in-capsule I observed monoterpenes are all within a few kcal/mol of the global monoterpene minimum. The monoterpene product with the highest experimentally observed yield is α-terpinene, 4, which may be explained by 4 being the monoterpene with the lowest free energy among the different products observed experimentally.β-Limonene, 14, has the highest free energy among the experimentally observed products, and thus has a very low experimental yield. High-level CBS-QB3 and G4 ab-initio calculations corroborate these DFT results. However, low energy monoterpenes, such as products 1, 2, α and β-camphene and α and β-phellandrene were not observed experimentally. This inspired us to perform QM/MM HT-LMD simulations in an attempt to observe the chemistry as it takes place inside the capsule.

The simulations showed that a broad spectrum of monoterpene products can form inside the resorcinarene capsule, via carboxylation chemistry. Additionally, the data suggests that when provided with sufficient kinetic energy, such as in our HT-LMD simulations, kinetically disfavored pathways to thermodynamically preferred monoterpenes are attainable. For instance, camphyl cation, 1, en route to camphene, was observed. This suggests that the designed HT-LMD simulation protocol can readily describe the different types of carboxylation chemistry occurring inside nano capsules. However, camphene was not observed experimentally and to answer this we turned to detailed DFT gas-phase studies of the mechanistic possibilities suggested by Tiefenbacher and co-workers and by our HT-LMD simulations.

As shown in Fig. 5, camphenes are obtained via pinyl cation, G. The TS forming G from C, TSCG, is slightly higher in free energy (by 0.9 kcal/mol) than the competing pathway TSCb, which forms terpinene-4-yl-cation, D. Hence there is a slight kinetic preference for the latter pathway. Additionally, the pinyl to bornyl pathway, G→H, has a free energy barrier of 7 kcal/mol, and the combined barrier from C→H is 10.8 kcal/mol. Similarly, the barrier for C→E is significantly higher than C→D (7.5 kcal/mol), explaining why phellandrene isn’t observed. This indicates a strong kinetic preference for C and D, and therefore for all the subsequent deprotonation species that are observed experimentally.

Thus, for the first question we can answer that the selectivity of the products is dictated by a balance of thermodynamic and kinetic control: the experimentally observed products are low-energy species with low-energy carboxylation precursors; and the height of the free energy barriers dictates which pathways are accessible. This proposal is further supported by in-capsule I experiments starting with β-pinene, which yielded camphene as a main product. It seems that the interaction between the capsule π-system and the secondary cation H favors this pathway over the pathway via M (Fig. 6).

In order to tackle the second question, regarding the difference in product selectivity for the different substrates, we focus on a specific case: namely the difference in formation of α-terpinene and terpinolene using GOAc and NOAc. When using GOAc, the amounts of α-terpinene and terpinolene were 35% and 5%, respectively, while NOAc as a substrate produced 34% and 24%, respectively (Fig. 1). We assume these products are formed primarily from cation D, which is slightly more stable than C (1.1 kcal/mol). Analysis of our HT-LMD simulation trajectories reveals that in the case of GOAc, acetic acid remains in the vicinity of carbon C3, as partial reattachment (i.e. electrostatic stabilization) to this position is necessary to form the reactive 2,3-cisoid cation. Hence, deprotonation at position C1, producing mostly α-terpinene, is a likely scenario. In the case of NOAc, acetic acid remains in the vicinity of C1, as the 2,3-cisoid cation is formed directly from C1-OAc cleavage. Here, deprotonation at either C1 or C7 is a possibility, yielding significant amounts of both α-terpinene and terpinolene. This might explain the elevated amounts of terpinolene when using NOAc. We measured the distances between C1 or C7 and the OAc oxygens and found that indeed, the acetic acid tends to remain near these carbons when NOAc turns into C or D (further information can be found in the SI, Fig. S7 a-c). We note that the difference between GOAc and NOAc can also be explained due to different chemical mechanisms, i.e. S1 and S2/S2', respectively. This explanation is in line with the current simulation results, which suggest that GOAc clearly reacts in a stepwise fashion, whereas NOAc reacts in a more concerted fashion.

In conclusion, the current results suggest that the product distribution and the dependence on the initial leaving group is a combination of thermodynamic and kinetic factors. This suggests that if one starts with different initial substrates, one might bypass kinetic bottlenecks to reach more stable products. For example, one might be able to synthesize α-camphene by feeding the capsule with pinene, as we showed in this work.

Conclusion

In this work we presented a multiscale nano-reactor simulation approach. The nano-reactor simulation protocol consists of hybrid QM(DFT)/MM-based HT-LMD simulations. Using this approach we model the production of tail-to-head formation of monoterpenes inside a resorcin[4]arene-based capsule, and observe the chemical mechanisms as they occur inside the resorcinarene nano-environment. The HT-LMD simulations are able to generate all relevant carboxylations that are precursors to the experimentally observed monoterpenes, as well as additional carboxylations. We find that the best temperature for performing these types of QM/MM HT-LMD simulations of monoterpenes, are 900K – 1000K. Even short simulations, such as the ones performed in this work, can produce a variety of carbocations appearing inside the resorcinarene capsule. We combine this approach with extensive DFT calculations
of monoterpene stability, as well as detailed reaction mechanisms. Using this extensive computational platform, we explain two central questions related to terpene synthesis in nano-capsules: (1) What determines the observed product distribution inside capsule? (2) What is the role of the substrate leaving group? The current results suggest that the product distribution is a result of a balance of thermodynamic and kinetic control: the experimentally observed products are low-energy species with low-energy carbocation precursors; however, the height of the free energy barriers dictates which pathways are accessible, and certain very low-energy monoterpeneS are not observed. Specifically, most products formed experimentally are among the most stable known monoterpeneS, suggesting thermodynamic control. However, some of the most stable monoterpeneS were not observed, due to kinetic control. The effect of the initial leaving group (i.e. the difference in the product distribution between each initial substrate) is possibly due to a change from a stepwise to concerted mechanism. Based on this in silico platform, we suggested an initial substrate that was able to overcome the kinetic control and reach a thermodynamically more stable terpene. Additionally, we show that cation interactions with the capsule may direct the dynamic reaction cascades.

ASSOCIATED CONTENT
Supporting Information. Details of gas-phase calculations. Free energy for all monoterpeneS. Comparison of gas-phase energies for all monoterpeneS using DFT and high-level ab initio calculations. Details of the QM/MM HT calculations. Details of the QM/MM HTLMD simulations. Additional experimental data. Distance analyses from the HT-LMD trajectories. CHARMM topology and parameter files.

AUTHOR INFORMATION
Corresponding Author
* majort@biu.ac.il

Author Contributions
All authors have given approval to the final version of the manuscript.

Funding Sources
Israeli Ministry of Science and Israeli Science Foundation (Grants # 2146/15 and 1683/18).

ACKNOWLEDGMENT
This work was supported by the Israeli Ministry of Science and Israeli Science Foundation (Grants # 2146/15 and 1683/18).

REFERENCES


