Supporting informations of: Solvation effects drive the selectivity in Diels-Alder reaction under hyperbaric conditions

Daniele Loco,[1] Riccardo Spezia*,[1] François Cartier,[1] Isabelle Chataigner*,[1,2] Jean–Philip Piquemal*,[1,3]
[1]Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France.
[2] Normandie Université, INSA Rouen, UNIROUEN, CNRS, COBRA laboratory, F-76000 Rouen, France.
[3] Institut Universitaire de France, 75005, Paris, France
Contacts: riccardo.spezia@sorbonne-universite.fr, jean-philip.piquemal@sorbonne-universite.fr (SU) isabelle.chataigner@univ-rouen.fr (Rouen)

April 17, 2020

S1 Quantum Chemistry calculations and methods benchmark

In order to define the best electronic structure method to be used in the QM/MM dynamics and energy evaluation, we test different DFT approximations and basis set size and their effect on the relative energies of reactants, products and transition states (TSs). We set the reference values of the energetic barriers to those obtained at MP2 level with an extended basis set. We present this benchmark both in gas phase and in dichloromethane solution, modelled by means of the IEF-PCM implicit solvation model. [1]

All the energetic barriers reported in Table S1 are computed on the reactant, TS or product geometry optimized with the corresponding level of theory. In gas phase the MP2/aug-cc-pVTZ calculations are considered as the reference, while the MP2/6-311++G(d,p) one are used as the reference for the condensed phase, since it is found that the difference between them is negligible ($\sim$ 0.1 kcal/mol).

In Table S1 $\Delta E_{endo/exo}$ refers to the energy difference between the endo/exo product and the reactants, while $\Delta E^{a}_{endo/exo}$ is defined as an activation energy, as the difference between the endo/exo product and the endo/exo TS. From the results it is found that the ΔE are sensitive to the method of choice, both with respect to the functional form and the basis set.

The M06-2X and the ωB97XD functionals show the best agreement with the reference calculations, and in particular they show a better agreement in the ΔE^{a} for the smallest basis set size, and on the contrary with the largest one for ΔE. M06-2X/3-21G level of theory represents a good compromise: while the ΔE are too negative ($\sim$ 6–8 kcal/mol both in gas phase and in solution with respect to MP2), they provide good quality ΔE^{a} ($\sim$ 1 kcal/mol in gas phase and even less in solution with respect to MP2).

Taking these two factors into account the M06-2X/3-21G level of theory is chosen for all the DFT calculations reported in the main manuscript of this work.

It is worth to note that the difference between the energy barriers computed for the endo and the exo configurations is in general very small ($<$ 1 kcal/mol), while the effect of the solvent is to lower, sometimes very slightly, the energy barriers.

It is also the case to remark that it is not uncommon that a smaller basis set can provide results better than a larger one: convergence in basis set sometimes occur for larger basis which are surely too big for the

following molecular dynamics simulations. In molecular dynamics simulations, it is indeed important to have a method which represents a good compromise between accuracy and computational time. Increasing the basis set results in an enormous increase of computing time, which limits the statistical sampling and finally the accuracy of QM/MM results.

Table S1: Method benchmark

	Gas phase				Implicit solvent			
	ΔE_{endo}	ΔE_{exo}	ΔE^a_{endo}	ΔE^a_{exo}	ΔE_{endo}	ΔE_{exo}	ΔE^a_{endo}	ΔE^a_{exo}
B3LYP/6-31G	-14.69	-14.80	20.80	20.47	-15.37	-15.20	18.77	18.86
B3LYP/6-31G(d)	-19.69	-19.79	–	–	-20.14	-19.98	17.12	17.13
TPSS/6-31G	-17.94	-18.10	–	–	-18.40	-18.27	12.67	12.64
TPSS/6-31G(d)	-22.67	-22.83	–	–	-23.13	-23.01	10.73	10.61
M06-2X/3-21G	-37.75	-40.12	10.04	9.69	-41.62	-41.59	7.14	7.45
M06-2X/6-31G	-26.19	-26.26	17.31	17.12	-27.59	-27.29	14.83	15.21
M06-2X/6-31G(d)	-29.31	-30.01	15.22	14.90	-31.20	-30.95	12.92	13.19
ωB97XD/3-21G	-40.36	-40.72	11.21	10.85	-40.69	-40.79	9.78	9.96
ωB97XD/6-31G	-26.03	-26.03	19.23	19.08	-26.74	-26.42	17.23	17.55
ωB97XD/6-31G(d)	-30.54	-30.54	17.43	17.15	-31.02	-30.71	15.72	15.91
MP2/6-311++G(d,p)	-31.74	-31.72	8.91	8.99	-32.57	-32.23	7.53	7.96
MP2/aug-cc-pVTZ	-31.19	-31.18	7.17	7.04	–	–	–	–

All calculations were performed with Gaussian09 suite of programs.[2]

S2 Polarizable molecular dynamics

A total of 7 sets of MD simulations, encompassing reactants, endo/exo TS, pre–TS and products, are run, using the geometries obtained from quantum chemistry optimizations, then solvated with 603 molecules of CH_2Cl_2 using the Packmol software.[3] For CH_2Cl_2 we used the AMOEBA force field developed by Mu et al.[4]. For the solute we kept its atomic positions fixed to let the solvent relax around, using the AMOEBA non–covalent parameters for organic molecules, as reported by Ren et al[5], to treat the solute–solvent interactions along the equlibration procedure and the following MD trajectories. The solvent was then equilibrated for all the systems at a constant temperature of 300 K and a constant pressure of 1 and 10000 atm (10 katm). The classical Newton's equation of motion were integrated numerically with the Velocity–Verlet algorithm[6] with a 1 fs time step. NPT ensemble trajectories were generated applying the Bussi–Parrinello algorithm[7] to maintain constant temperature and pressure.

Each system, was first equilibrated for 10 ps, then each dynamics has been propagated for at least 2 ns. After an equilibration time of less than 10 ps the systems at the two different pressures reach their different equilibrium densities, namely 1.419(3) and 1.776(4) g/mL. In Fig. S1 we have reported the initial equilibration of the densities of all systems, both endo and exo, at high and low pressures along the solvent dynamics. The low (1 atm) and high (10 katm) pressure systems thus correspond to simulation boxes of different sizes as shown in the main manuscript.

All the AMOEBA polarizable MD have been carried out with the Tinker-HP package[8,9].

S3 Polarizable QM/MM simulations

S3.1 Principles of the method

The polarizable QM/MM we employed[10] is based on a variational formulation of the embedding scheme[11]. The classical environment (the solvent in this work) is treated at the AMOEBA FF level, while DFT is

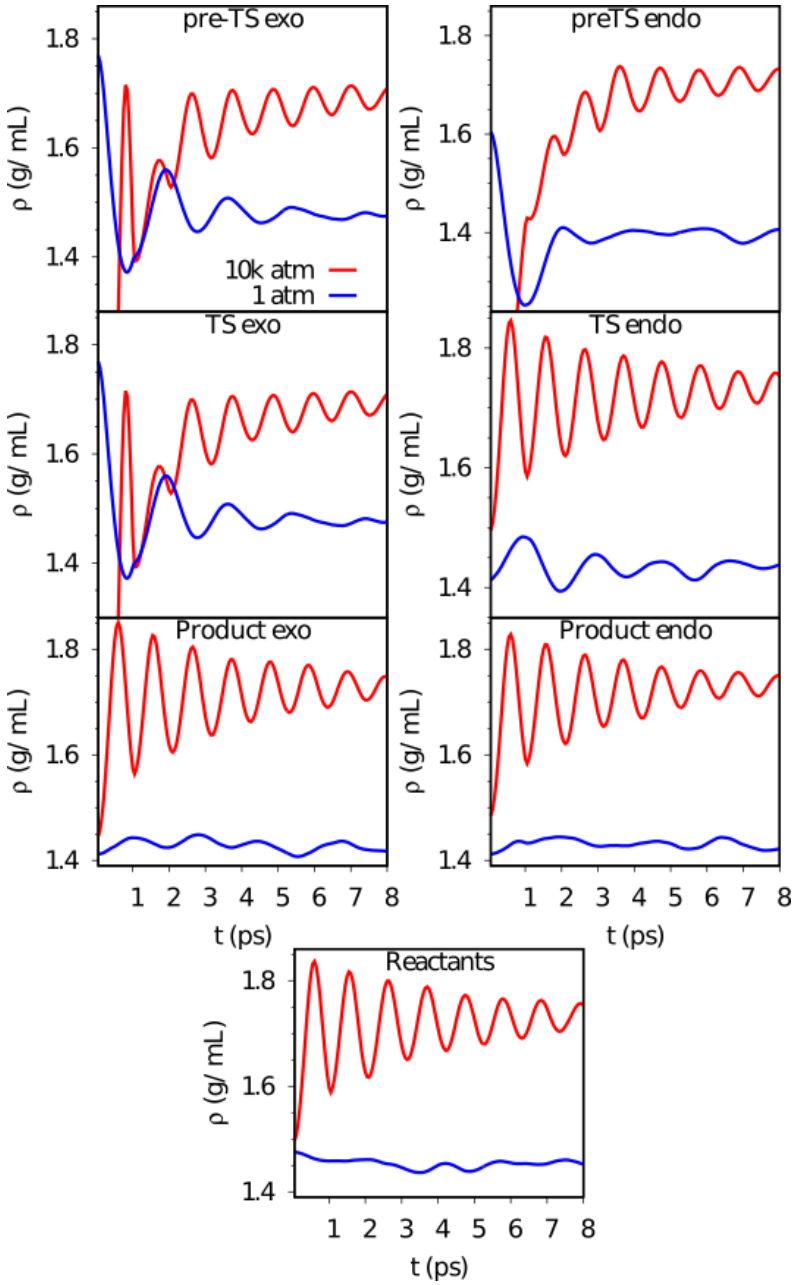

Figure S1: Solution density's time–series along the initial part of the equilibration dynamics run for all the chemicals analyzed in this study (pre-TS, TS, product and reactants), each in a CH_2Cl_2 box of solvent molecules. In red are reported the densities for the 10 katm MDs, while those for the 1 atm MDs are in blue, for both endo and exo conformations

used for the QM subsystem. The variational formalism couples the classical induced dipoles, representing the polarization of the MM subsystem, and the electron density of the QM one in a self-consistent way, both for energy and forces evaluation.[12,13]

The classical atomic point dipoles, computed in the framework of the smeared Thole's damping interaction scheme[14], are induced in response to the electric field generated by both the QM subsystem, as well as the MM sites themselves, which bear distributed multipoles up to quadrupoles. When the environment is polarized, this additional set of dipoles is induced at each classical site, polarizing back the QM density, giving rise to the so–called mutual polarization effect, which is solved in a self–consistent field(SCF)–based procedure, as in standard DFT implementations.

The total energy of the system includes contributions arising from the QM/MM coupling and both from the interactions between QM atoms only and MM atoms only. In a very general form it can thus be written as:

$$E^{\mathrm{Tot}}(\boldsymbol{P},\boldsymbol{\mu}) = E^{\mathrm{MM}} + E^{\mathrm{QM/MM}}(\boldsymbol{P},\boldsymbol{\mu}). \tag{1}$$

The first term of the sum accounts only for interactions modeled with the classical AMOEBA potential, to compute bonding and electrostatic energy contributions between classical atoms, as well as dispersion/repulsion interactions. Also QM/MM dispersion/repulsion interactions are included in this way. The second term of eq. S3.1

$$E^{\mathrm{QM/MM}}(\boldsymbol{P},\boldsymbol{\mu}) = E^{\mathrm{QM}}(\boldsymbol{P}) + E^{\mathrm{El}}(\boldsymbol{P}) + E^{\mathrm{Pol}}(\boldsymbol{P},\boldsymbol{\mu}), \tag{2}$$

includes all the quantities that are computed in the SCF procedure, namely: i) the QM energy expression of the isolated system, ii) the QM/MM electrostatic interaction between the AMOEBA multipoles and the DFT density and lastly iii) the mutual polarization contribution to the QM/MM energy. For each term is highlighted the explicit dependence on either the QM electron density, represented through the one–body density matrix $\boldsymbol{P}$, or the classical induced dipoles $\boldsymbol{\mu}$. Those are the terms which are variationally minimized during the SCF procedure, to reach a self-consistent relaxation of both the MM induced dipoles and the QM electronic density.

The purely classical contribution E^{MM} in eq. S3.1 does not enter in the SCF procedure, since it does not depend on either the electronic density or the classical induced dipoles, being only dependent on the nuclear configuration.

From eq. 2 are derived analytically the forces used to integrate the classical Newton's equations of motion in a mixed QM/MM MD scheme. The widespread velocity–Verlet algorithm is used for the numerical integration of the Newton's equations.[13]

The method has been implemented in a developing version of the Gaussian09 suite of programs[2], interfaced with the Tinker MD package[8,9] The energy defined in eq. 2 is used in the analysis of the results obtained from QM/MM energy calculations and dynamics, either extracting it from single–point calculations or as an average from the computed trajectory.

Gaussian is used to solve the QM/MM equations and to compute the electrostatic and polarization QM/MM energies and forces. This allows us to minimize communication between Gaussian and Tinker. Tinker/Tinker-HP computes all MM nonelectrostatic terms and, given the total forces, integrates the equations of motions. The QM/MM driver handles the communication between Gaussian and Tinker and works directly on the input/output of each program.

S3.2 Polarizable QM/MM dynamics technical details

The classical polarizable AMOEBA trajectories, for each chemical species involved in the different stages of the reaction, either in exo or endo configuration for both high and low pressure, have been sampled extracting solute–solvent configurations to be used as starting geometries for as many short (from $\sim$ 400 fs to less than 2 ps) hybrid polarizable QM/MM MDs in the NVE ensemble, with a 0.5 fs time–step, using the original simulation cell as a droplet for free (nonperiodic) boundary conditions. This is not a problem in our case because of the long equilibration of the box of solvent in the full FF–based MD

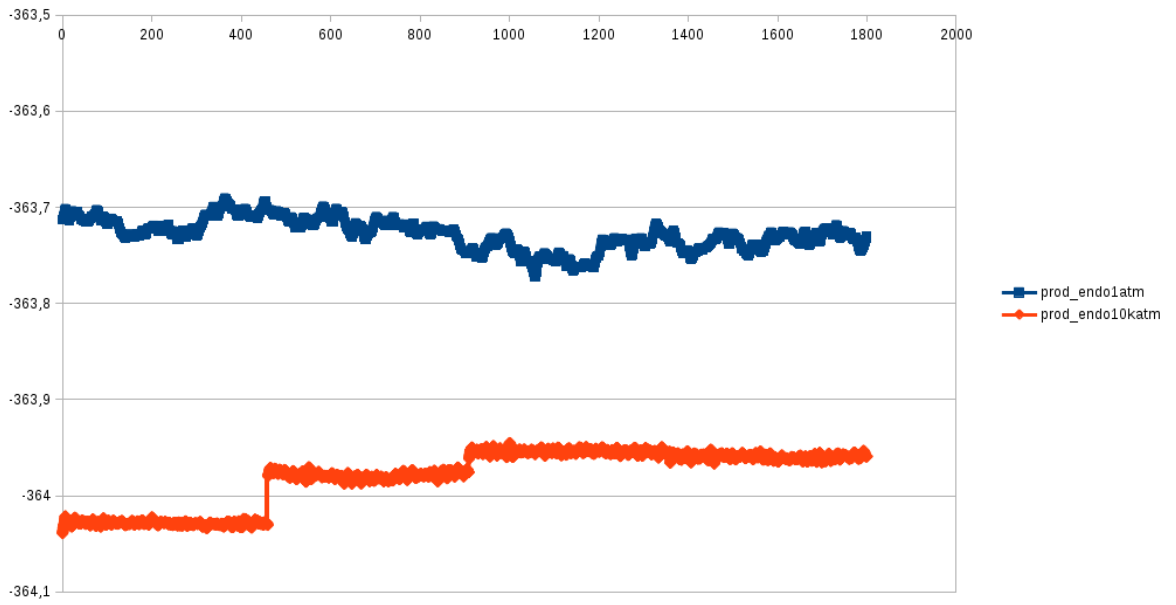

Figure S2: QM/MM energy of the solute along ∼1.8 ps of the hybrid polarizable QM/MM trajectories for the endo product at 1 atm (blue line) and 10 katm (red line, composed by three short trajectories). Y-axis reports energies in atomic units, while x-axis time in femtoseconds.

compared to the short hybrid QM/MM MDs, ranging in the time–scale of nanoseconds the former and of picoseconds the latter (*vide infra*). The short time scale is sufficient to sample the ground state energies of the reacting species, since they are in equilibrium with the solvent environment and no major (long lasting) solvent molecules reorientation and reactant nuclear displacements take place, as shown by the very small fluctuations of the QM system energy (see Figure S2). On the other end the short time scale of each singular QM/MM MD allow us to avoid problems like solvent evaporation related to the free boundaries of the droplet.

Since no barostat can be applied in the QM/MM dynamics, even trajectories as short as 1 ps rapidly induce the system at 10 katm to go back to the original normal atmosphere condition, with an increase of the energy, which is in fact observed to be generally lower at high pressure (see Table S2). For this reason for each reacting species at 10 katm the phase space spanned by the long full AMOEBA dynamics is sampled picking 10 points, well separated in time, and from those configurations very short ($\sim$ 500 fs) polarizable hybrid QM/MM dynamics are run, averaging then the energies obtained from each of them. Since the main interest in performing QM/MM MD is in the evaluation of environment effects on the QM subsystems, this procedure ensure the sampling of a variety of MM environments. The short length of each QM/MM trajectory at 10 katm is compensated with their number, to improve the statistic for the average energy and properties computation, obtaining a total simulation length quite larger then the case at 1 atm (4.5 ps instead of 1.8 ps).

Comparing the 1 and 10 katm energies time–series in Fig. S2, within a compatible time scale, they show quite similar behaviors, and the high pressure ensemble of dynamics does not show energy drift, while it appears to fluctuate around a constant average in its whole length (see blue curve in Fig. S3). The energy sampling along each short QM/MM MD is just stopped before the effect of the adiabatic expansion on the energy starts to appear, as it is also visible in the picture reported (no drift occurs during a short MD run). Also the exo product at 10 katm is reported in Fig. S3. The jumps seen in the energy time series correspond to the different starting points used for each short QM/MM MD runs, taken from each 100 ps of the 1 ns solvent equilibration dynamics described in Section S2.

The situation is quite different for pre–TS and TS exo/endo, where for both low and high pressure conditions no dynamics is theoretically practicable, since it would, in principle suddenly move the TS–like

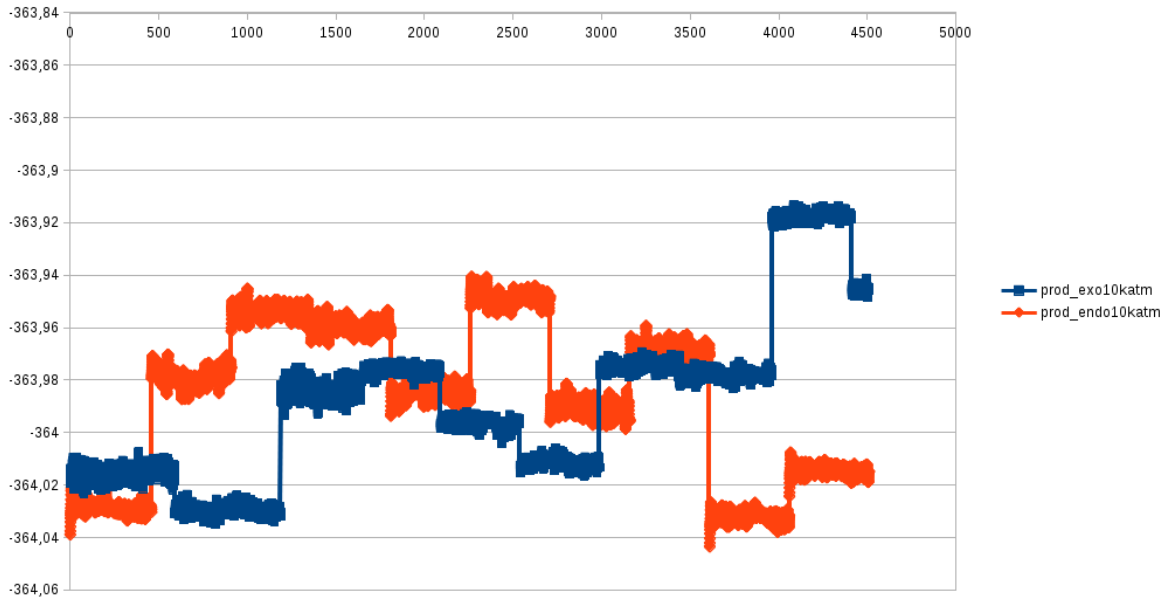

Figure S3: Whole solute QM/MM energy time–series along a total of 4.5 ps trajectory (composed by ∼10 short 450 fs trajectories) of the endo (blue line) and exo (red line) products at 10 katm. Y-axis reports energies in atomic units, while x-axis time in femtoseconds

structure towards reactants or product. The corresponding AMOEBA MDs are sampled and QM/MM energy–evaluations are performed directly on the extracted structures, without any further refinement of the configuraion of the system, by extracting 10 structures each, in analogy with what as been done for high pressure MDs. For each stage of the reaction, at the two different pressures, the reacting species are treated at the M06-2X/3-21G level of theory, as this is the level of theory selected for the reasons discussed in Section S1. The solvent box is modeled with the same AMOEBA FF used for the full FF–based dynamics (see Section S2).

The energy shown here, and used in the paper to define the chemical pathway of the Diels-Alder reaction at different pressures, is the solute QM/MM one, meaning that all the energies used are strictly including only the internal energy of the QM subsystem and its interaction energy with the classical solvent. In other words the energy discussed does not include at all the whole energy of each solvent molecule or any energy interaction not involving the QM subsystem. Since all structures (obtained for each species) have been used in the energy evaluation used to determine the reaction pathway (reported in the main manuscript of the article) the resultant statistical uncertainty is pretty low, due also to the small energy fluctuations along the reported NVE trajectories.

S4 Average absolute energies from QM/MM simulations

In Table S2 the QM/MM energies absolute values are reported. Only the components of the hybrid QM/MM energies taken into account to build the energy profile for the reaction under investigation reported in the main article are considered. In particular these components are: i) the solute QM energy modified by the electrostatic interaction with the environment classical multipoles through the variational coupling with the MM subsystem (E^{solute}(Elec), *i.e.*: terms 1 and 2 of eq. 2), ii) the QM/MM polarization between the QM solute and the MM solvent (E^{solute}(Pol), *i.e.*: term 3 of eq. 2) and iii) their sum E^{solute}(Elec+Pol).

Notably the superscript solute is meant to stress that all these energy components are strictly including only the internal energy of the QM subsystem and its interaction energy with the classical solvent, so that

the energy discussed does not include at all the whole energy of each solvent molecule and any energy interaction not involving the QM subsystem. This is in fact the energy considered in all the analysis conducted in the present work for the chemical reaction pathway under study.

Table S2: Absolute QM/MM energies, reported in atomic units, computed as an average over a series of structures, for all the species involved in the reaction. The structures are obtained as described in Subsection S3.2

1 atm							
	Substrate	$Prod_{exo}$	$Prod_{endo}$	TS_{exo}	TS_{endo}	Pre–TS_{exo}	Pre–TS_{endo}
E^{solute}(Elec+Pol)	-363.667	-363.729	-363.727	-363.637	-363.649	-363.652	-363.655
E^{solute}(Pol)	-0.905	-0.897	-0.900	-0.876	-0.887		
E^{solute}(Elec)	-362.762	-362.832	-362.830	-362.761	-362.762		
10 katm							
	Substrate	$Prod_{exo}$	$Prod_{endo}$	TS_{exo}	TS_{endo}	Pre–TS_{exo}	Pre–TS_{endo}
E^{solute}(Elec+Pol)	-363.929	-363.983	-363.986	-363.947	-363.957	-363.962	-363.959
E^{solute}(Pol)	-1.165	-1.151	-1.157	-1.184	-1.195		
E^{solute}(Elec)	-362.764	-362.832	-362.829	-362.761	-362.761		

S5 Absolute molar volumes

In Tab. S3 the average molar volumes of each species involved in the investigated reaction are reported. The average is computed over the structures obtained by short-runs of QM/MM MDs, but for the TS and pre–TS, where the structures sampled from FF–based MDs for the solvent equilibration, while the structures of the reagents are fixed in their QM optimized geometries.

From the table it can be noticed how the molar volumes are not generally affected by the change in pressure, as well as that exo/endo stereoisomers do not show significant differences between them. As an example the difference between endo and exo products (ENDO and EXO in Tab. S3) at both 1 and 10 katm is negligible (less than 0.5 cm^3mol^{-1}), as well as the difference between the endo products at 1 and 10 katm or the exo ones.

Notably only for the pre–TS structures the pressure effect is noticeable, and in particular the pre–TS exo increase its molar volume passing from low to high pressure condition of approximately 5 cm^3mol^{-1}, leading to the differences in the in activation volumes discussed in the main article.

These volumes are used in the paper to build the relative activation and reaction volumes table. It is also worth noticing that the electronic densities used are the final SCF densities, including all the effects due to the QM/MM interactions which have been discussed in Section S3.

References

[1] J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct.*, 1999, **464**, 211 – 226.

[2] M. J. Frisch and *et al*, *Gaussian Development Version, Revision H.36*, Gaussian Inc. Wallingford CT 2010.

[3] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

[4] X. Mu, Q. Wang, L.-P. Wang, S. D. Fried, J.-P. Piquemal, K. N. Dalby and P. Ren, *J. Phys. Chem. B*, 2014, **118**, 6456–6465.

[5] P. Ren, C. Wu and J. W. Ponder, *J. Chem. Theory Comput.*, 2011, **7**, 3143–3161.

Table S3: Molar volumes for each species involved in the reaction at high and low pressures. The volumes are computed from the electronic densities and their mean and confidence interval (c.i.) are reported in cm^3mol^{-1}

	1 atm		10 katm	
	$\langle V \rangle$	c.i.	$\langle V \rangle$	c.i.
Reactants	106.080	0.610	106.271	0.612
pre-TS Endo	103.766	1.722	101.673	3.599
pre-TS Exo	99.702	3.005	104.730	3.249
TS Endo	97.427	2.566	97.268	3.168
TS Exo	95.993	3.515	97.327	3.513
Product Endo	95.316	0.535	95.154	0.579
Product Exo	95.836	0.553	95.676	0.550

[6] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.

[7] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[8] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory and Comput.*, 2018, **14**, 5273–5289.

[9] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder and J.-P. Piquemal, *Chem. Sci.*, 2018, **9**, 956–972.

[10] D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci and J.-P. Piquemal, *Chem. Sci.*, 2019, **10**, 7200–7211.

[11] F. Lipparini, G. Scalmani, B. Mennucci, E. Cancès, M. Caricato and M. J. Frisch, *J. Chem. Phys.*, 2010, **133**, 014106.

[12] D. Loco, Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal and B. Mennucci, *J. Chem. Theory Comput.*, 2016, **12**, 3654–3661.

[13] D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2017, **13**, 4025–4033.

[14] B. Thole, *Chem. Phys.*, 1981, **59**, 341–350.