Supporting Information

to

Chemical Feasibility of the General Acid/Base Mechanism of *glmS* Ribozyme Self-Cleavage

Matúš Dubecký,[1] Nils G. Walter,[2] Jiří Šponer,[3,4] Michal Otyepka[1] and Pavel Banáš[*,1,3]

[1]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, tr. 17 listopadu 12, 771 46, Olomouc, Czech Republic
[2]Department of Chemistry, Single Molecule Analysis Group, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055.
[3]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic.
[4]CEITEC – Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic.
* corresponding author: phone +420 585634769, e-mail: pavel.banas@upol.cz

Convergence of the QM/MM energies with extending of the QM region

In order to verify the convergence of the QM/MM energies with respect to the completeness of the QM region, we have also performed large-scale QM/MM computations with a set of extended QM regions, at the limits of contemporary computer facilities. Specifically we explored convergence using four progressively extended QM regions (see Figure S1):

i) A *minimal QM region* comprising 65 atoms including the deprotonated nucleobase G40$^-$ capped by a C1' methyl group, part of the GlcN6P cofactor without the methylphosphate group, and part of the sugar-phosphate backbone ranging from the ribose of A-1 up to C4' carbon of the G1 ribose.

ii) A *middle1 QM region* comprising 149 atoms including additionally the C1' methyl-capped G39 forming a base phosphate (BPh) 4BPh interaction with the scissile phosphate, a C1' methyl-capped U51 forming a hydrogen bond with the ammonium group of the cofactor, a C1' methyl-capped G65 hydrogen-bonding cofactor as well as the scissile phosphate, G66 including its ribose moiety forming an interaction with G40$^-$ and the A-1(2'-OH) nucleophile, and one water molecule.

iii) A *middle2 QM region* comprising 176 atoms including additionally the remaining part of the GlcN6P cofactor, the 5'-terminus of A-1 forming a hydrogen bond with G40(O6), and a (solvated) Mg^{2+} ions located near the GlcN6P phosphate group.

iv) An *extended QM region* comprising 248 atoms including additionally the extension of the sugar phosphate backbone up to C2(O5') including the complete G1 nucleotide forming a 4BPh interaction with the GlcN6P and C2 phosphates, a sodium ion bridging the C2 phosphate and G39(O6), a second (solvated) Mg^{2+} ion located near the GlcN6P phosphate group, and an additional eight water molecules bridging the parts of the QM region.

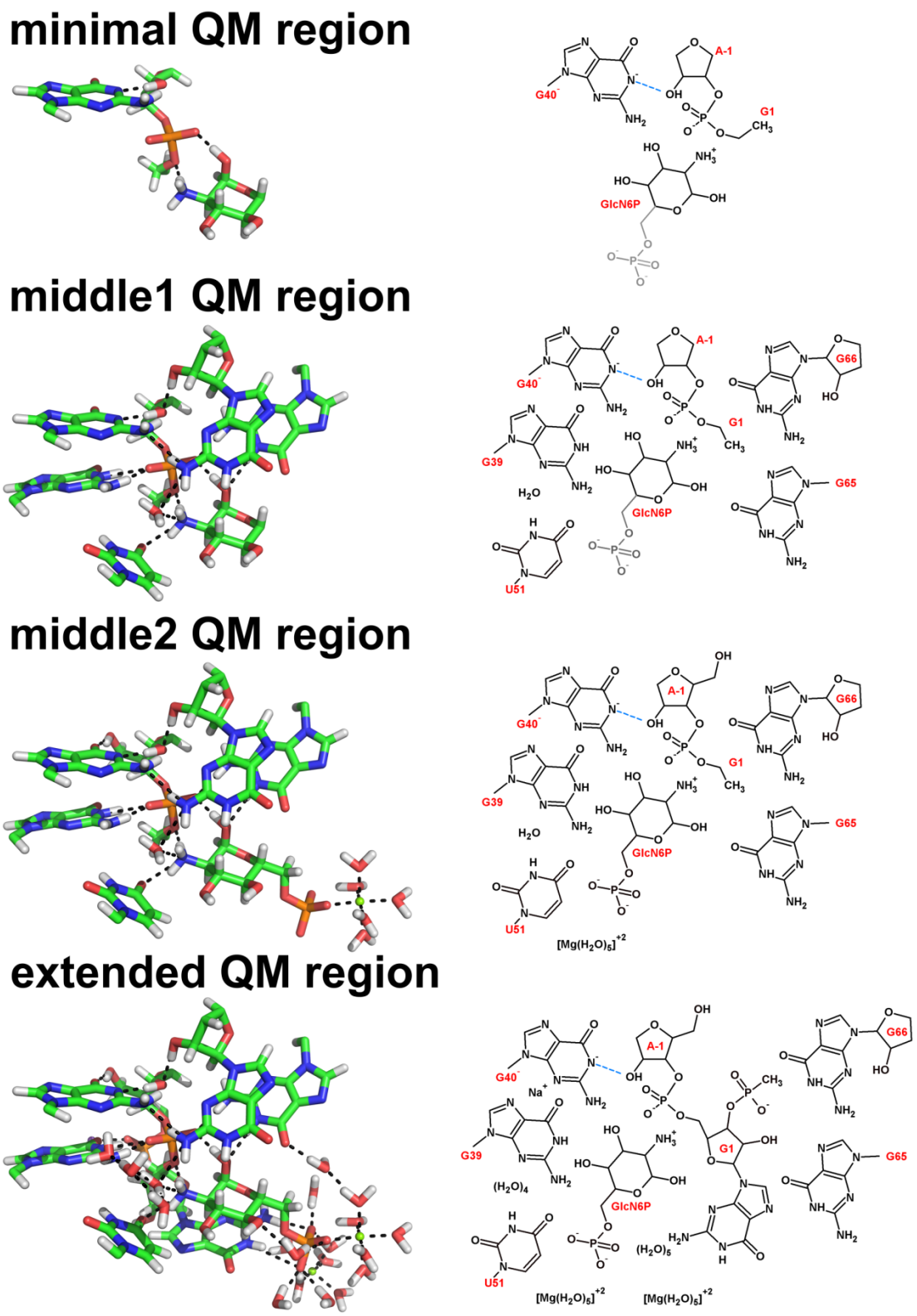

Figure S1. The structures (left) and the schemes (right) of four used QM regions that were progressively extended from the minimal QM region up to the extended QM region discussed in main text.

Table S1. Dependence of the QM/MM energies in kcal/mol on the QM region size. The QM/MM energies of product (P) and transition state like structure (TS-like) are calculated with respect to the energy of reactant (R) and without any correction. Note that the TS-like state corresponds to the true transition state only for the minimal QM region in electronic embedding.

	Electronic embedding		Mechanical embedding	
QM region	**TS-like**	**P**	**TS-like**	**P**
QM minimal	27.2	20.6	-4.9	-19.6
QM middle1	32.8	26.1	5.1	-8.6
QM middle2	29.1	22.4	10.6	-2.2
QM extended	25.7	16.0	19.7	6.9

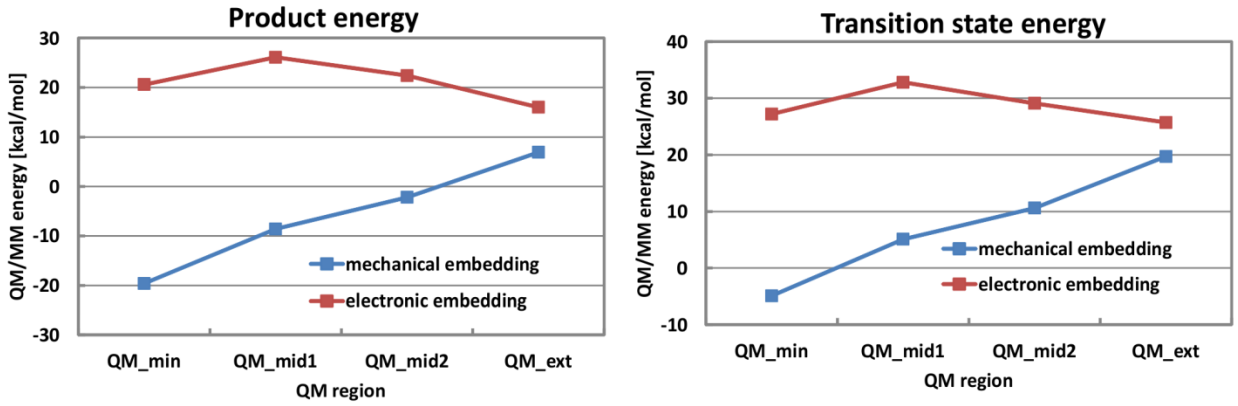

Figure S2. Dependence of the QM/MM energies in kcal/mol on the QM region size in both electronic and mechanical embedding: The energies of product (left) and transition-like (right) states (see also caption of Table S1) are shown with respect to the energy of reactant without any correction in kcal/mol.

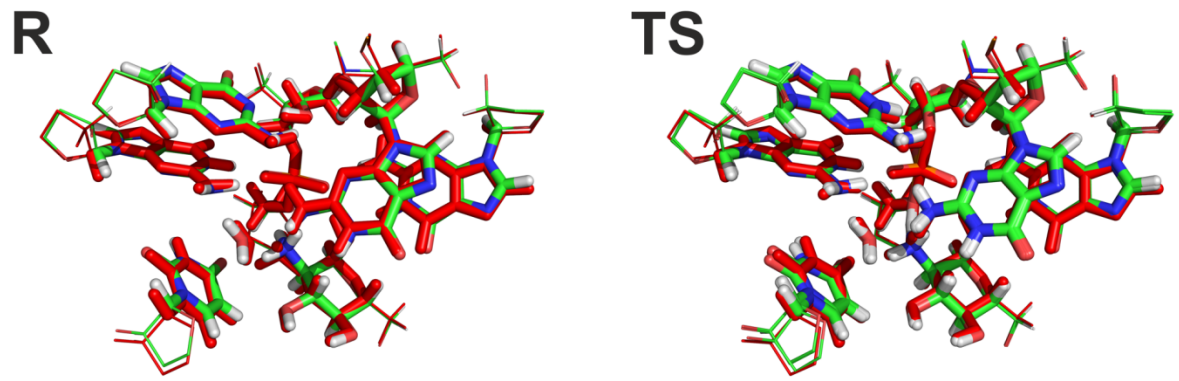

Figure S3. Structures of the middle1 QM region after the partial (not fully converged) optimizations of the reactant (R – left panel) and three structures around the transition state (TS – right panel) compared to the initial structure (in red) obtained by optimizations with minimal QM region. Convergence was not obtained even after two months of calculations using our best performing computational equipment with shared memory parallelization among 16 Intel E5-2660 processors.

Derivation of energy data in Table 2 of the main text and estimation of the associated uncertainties

When comparing the reaction energetics obtained by present QM/MM calculations with the corresponding experimental kinetic data, three important aspects should be taken into account: (i) All energies along entire reaction profile have to be corrected for the thermodynamic penalty due to the rare protonation form of the reactive state (see Method section in the main text), (ii) the energies calculated using minimal QM region (with electronic embedding scheme) should be corrected for the systematic error due to over-polarization effects, and (iii) the sampling over the ribozyme's configurations should be included by averaging the barrier over different starting structures. In addition, the latter two aspects contribute to the total estimate of uncertainty.

Table S2 summarizes reaction barriers calculated for five different starting structures, which are already corrected for the thermodynamic penalty due to the rare protonation form of the reactive state that equals 3.6 kcal/mol, i.e., the starting reactant (pre-cleavage) structure from which our scans were initiated (the rare protonation for containing deprotonated $G40^-$ and protonated ammonium form of GlcN6P) was assigned an energy of +3.6 kcal/mol. The average reaction barrier equals 29.5±4.8 kcal/mol, where the estimation of the uncertainty was evaluated as the half-width of the mean confidence interval as follows:

$$\frac{s \cdot t_{0.975}^{1-n}}{\sqrt{n}},$$

where *s* stands for standard deviation, *t* is quantil Student distribution, and *n* denotes to the number of data. Note that this estimate contains the correction for the rare protonation form of the reactive state and effects of sampling over ribozyme conformations, however, it is based on QM/MM calculations using minimal QM regions, so is still biased due to over-polarization.

Table S2. The QM/MM energies of reactant (R) and transition states (TS) obtained in the calculations starting from different MD snapshots are shown in kcal/mol. The statistics summarize the averaged value of the reaction barriers together with the uncertainty estimated by the half-width of the mean confidence interval.

snap ID	R[a]	TS
1[b]	3.6	32.0
2	3.6	23.9
3	3.6	28.1
4	3.6	29.3
5	3.6	34.0
statistics		29.5±4.8

[a]The energy of reactant is elevated by the correction to the rare population of the reactive state. Note that also the TS energy is elevated by the same correction (see methods of the main text).
[a]This snapshot was used for detail exploration of the potential energy landscape of the self-cleavage reaction. The other snapshots were reoptimized using R and TS state from snap 1 as an initial guess.

In order to take into account also the over-polarization effect in the small QM region, the systematic error originating from this effect should be evaluated. The over-polarization (in electronic embedding) can be eliminated upon increasing the size of the QM region. We studied the convergence of the calculated energies with the size of the QM region using one particular

starting structure (see Table 1 in the main text, Figures S1, S2, and the first section of this Supporting Information). The energies calculated using the extended QM region are supposed to be the most accurate since the QM-MM interface is sufficiently far from the chemical changes occurring in the active site, and thus the QM-MM coupling is less error prone. It is worth noting that both electronic as well as mechanical embedding are limited in accuracy of QM-MM coupling description although the electronic one is likely more realistic. In other words, while mechanical embedding completely neglects the polarization effects, the electronic one is prone to overestimate polarization. Thus the most reasonable energy estimate for a given particular starting structure is expected to be in the middle of these values, and might be expressed as an averaged value of energies obtained by both embedding schemes in the extended QM region, i.e., as follows:

$$E = \left(E_{extended,EE} + E_{extended,ME}\right)/2$$

where $E_{extended,EE}$ and $E_{extended,ME}$ are the energies obtained using an extended QM region with electronic and mechanical embedding, respectively. Using the data from Table 1 in the main text, the reaction barrier in the extended QM region is calculated to 27.5±1.5 kcal/mol, where the uncertainty due to the residual effect of the QM-MM description inaccuracy in the extended QM region is evaluated as half of the difference between the mechanical and electronic embedding energies:

$$\delta = \left|E_{extended,EE} - E_{extended,ME}\right|/2$$

The barrier height estimated using the extended QM region is by 4.6 kcal/mol lower than the corresponded barrier height obtained for the same starting structure using the minimal QM region and electronic embedding, i.e., the systematic error originating from the over-polarization in minimal QM region can be estimated as -4.6±1.5 kcal/mol. In other words, the reaction barrier estimation comprising sampling over multiple starting structures and the penalty for the rare protonation form of the reactive state should be corrected also for this systematic error originating from over-polarization. Thus, the total estimate of the reaction barrier is 24.9 ± 5.0 kcal/mol (see Table 2 in the main text). The total estimation of uncertainty was derived from the two uncertainties mentioned above (±4.8 kcal/mol due to sampling over different starting structures and ±1.5 kcal/mol due to residual inaccuracy of the QM-MM description in the extended QM region) using error propagation:

$$\delta_{tot} = \sqrt{\delta_{coupl}^2 + \delta_{start}^2},$$

where δ_{tot}, δ_{coupl}, and δ_{start} are the total uncertainty, uncertainties originating from the residual effect of the approximate description of the QM-MM coupling, and the different starting structures, respectively.

In contrast to the reaction barrier, i.e., the relative energies of the R and TS states, we did not calculate the energy of the product (P) state for the additional starting structures. Therefore, we do not have any information on the variability of the product energy due to the sampling over different starting structures. However, our study is focused on the energy feasibility of the forward reaction mechanism, so we are primarily interested in the reaction barrier between states R and TS. For the product energy, we need to consider only the correction for the rare protonation form of the reactive state (+3.6 kcal/mol) and the over-polarization effect. Since we calculated P energies only for one starting structure, for which we also tested the effect of the increasing size of the QM region, the best estimate of the P energy is thus the value obtained using the extended QM region averaged over both embedding schemes, while half of the difference between these energies (from distinct couplings) estimates the uncertainty due to

residual limitations of the QM-MM coupling description, i.e., the estimate of the P energy is 15.1 ± 4.6 kcal/mol (see Tables 1 and 2 in the main text).