The structure-derived mechanism of box H/ACA pseudouridine synthase offers a plausible paradigm for programmable RNA editing

Dóra Judit Kiss,[1] Julianna Oláh,[2] Gergely Tóth,[3] Máté Varga,[4] András Stirling,[5] Dóra K. Menyhárd,[6] György G. Ferenczy[1,7*]*

[1]Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Magyar tudósok krt. 2, H-1117 Budapest, Hungary

[2]Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, Szent Gellért tér 4, H-1111 Budapest, Hungary

[3]Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány P. stny. 1/a, H-1117, Budapest, Hungary

[4]Department of Genetics, ELTE Eötvös Loránd University, Pázmány P. stny. 1/c, H-1117 Budapest, Hungary

[5]Theoretical Chemistry Research Group, Research Centre for Natural Sciences, Magyar tudósok krt. 2, H-1117 Budapest, Hungary

[6]MTA-ELTE Protein Modelling Research Group, Institute of Chemistry, ELTE Eötvös Loránd University, Pázmány P. stny. 1/a, H-1117, Budapest, Hungary

[7]Department of Biophysics and Radiation Biology, Semmelweis University, Tűzoltó u. 37-47, H-1094 Budapest, Hungary

e-mail addresses: dora.k.menyhard@ttk.elte.hu (D.K. Menyhárd) ferenczy.gyorgy@ttk.hu (G.G. Ferenczy)

Abstract

The uridine to pseudouridine transformation, one of the most abundant and essential posttranscriptional modification of RNAs, is carried out by pseudouridine synthases (PSUs). Aside from a few, very specific targets, pseudouridylation is performed by a ribonucleo-protein complex, box H/ACA PSU, containing 4 different proteins and a guide RNA. Mutations of PSUs cause serious diseases including dyskeratosis congenita (DC), various types of cancer and nephrotic syndrome. Here we combined homology modelling, classical force field based molecular dynamics and quantum mechanics/molecular mechanics based enhanced sampling free energy simulations to show that reactant destabilization through the severe distortion of the target uridine in the active site of box H/ACA PSU is a key factor in the catalysis of pseudouridylation. We propose a dissociation-rebound mechanism where the uracil detaches from the ribose by the cleavage of the C_1'-N_1 bond leading to a charge separated intermediate. The base rebounds to the ribose with its C_5 carbon with a very small barrier. The subsequent tautomerization step is proposed to be coupled to the tilting of the upper dyskerin region, comprising the thumb loop,

and product release. The proposed mechanism does not impose sequence restriction on the substrate – it only requires a complementary guide RNA coordinated to the protein components of the enzyme complex. We also found that the interactions of the guide RNA with the proteins of the complex in the vicinity of the active site are overwhelmingly formed by the sugar-phosphate backbone, indicating that designed guide RNAs could be applied to carry out pseudouridylation of substrates with a great variety of different sequence motifs. Therefore, the endogenous box H/ACA PSU system may be used to target premature stop codons – for example - to induce their read through serving as a vehicle for RNA editing and therapeutics for gene lesion related diseases.

keywords: box H/ACA pseudouridine synthase, catalytic mechanism, QM/MM free energy simulation, reactant destabilization, guide RNA, RNA editing

1. Introduction

Box H/ACA pseudouridine synthases are protein-RNA complexes that are responsible for the isomerization of uridine to pseudouridine in specific locations of a wide range of substrate RNAs. Coordination and positioning of substrate RNAs within the enzyme complex is accomplished by a small nucleolar guide RNA (snoRNA) carrying both an H-box and an ACA-motif (hence the name), and the catalytic isomerization by the protein components (Figure 1).

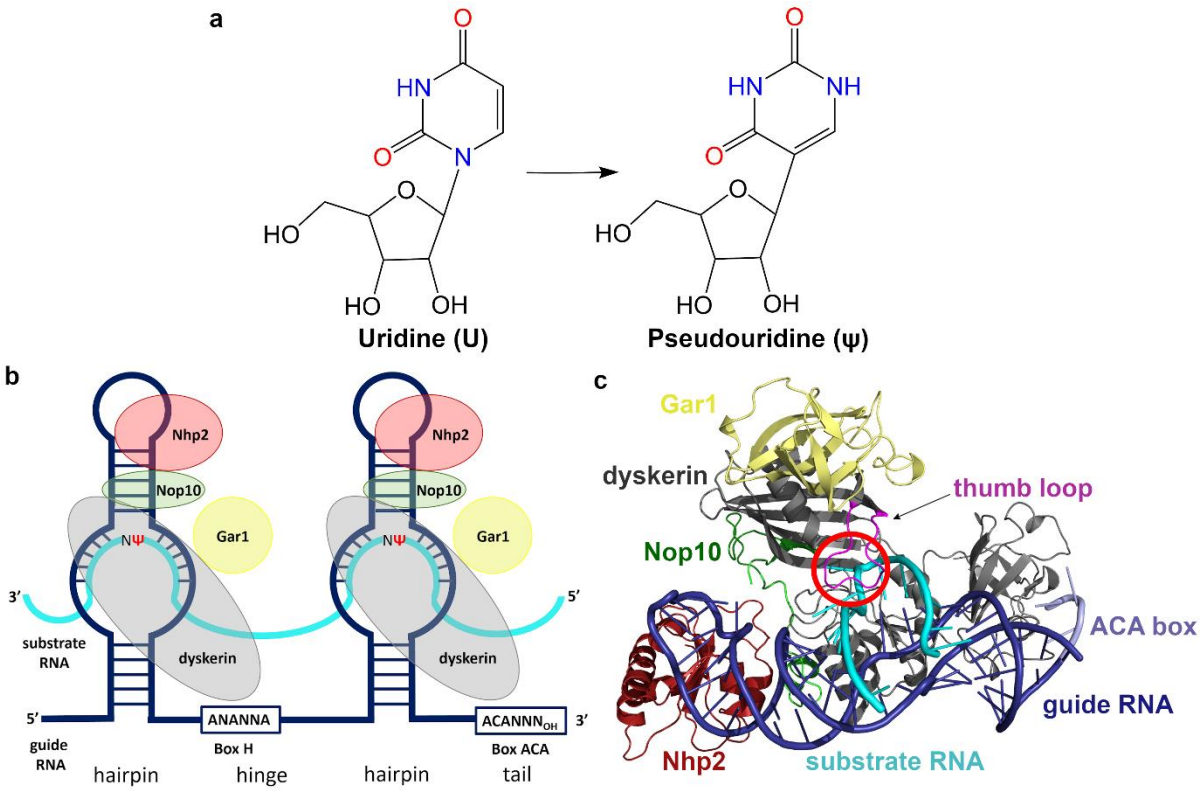


Figure 1. a) Uridine isomerization to pseudouridine liberates an imino group creating a new, non-classical H-bond donor site on the base. b) The schematic representation of the machinery carrying out guide RNA dependent pseudouridylation. Four core proteins assemble the H/ACA RNP: the catalytic unit dyskerin (grey), Nop10 (green), Nhp2 (red) and Gar1 (yellow). The guide RNA (dark blue) exhibits a hairpin-hinge-hairpin-tail structure and thrusts the target uridine into the active site by site-directed base pairing with the substrate RNA (cyan). The red Ψ indicates the pseudouridylation site. c) The 3D structure of the box H/ACA pseudouridine synthase complex. The thumb loop that secures the substrate in the active cleft is colored purple and the substrate RNA light blue. The red circle indicates the pseudouridylation active site of the enzyme complex.

The isomerization of uridine to pseudouridine (Ψ) is the most abundant posttranscriptional modification of RNA,[1,2] so much so, that Ψ is often referred to as the "fifth nucleotide".[3] This modification occurs and is considered to be essential in various types of RNAs[4] - tRNAs,[5,6] rRNAs,[4,7] snRNAs,[8] and mRNAs.[9] Accordingly, the function of multiple ribonucleo-protein complexes, such as the ribosome[10] and the spliceosome[8,11–13] also depends on the normal regulation of pseudouridylation. Box H/ACA PSUs, located in the nucleolus and nucleoplasmic Cajal bodies[14–17], thus take part in such fundamental functions as telomere length maintenance and ribosome biogenesis. Mutations appearing in their protein components are linked to serious illnesses like bone marrow failure, cancer or nephrotic syndrome.[18–21]

Structurally, the U→ Ψ isomerization results in a base carrying an additional hydrogen bond donor, an imino group, modifying and enhancing the interaction network it may form.[22–25] These fortified interactions with non-classical directionality enhance the rigidity and thermal stability of RNAs,[26–29] and promote the formation of turns and pockets. Thus Ψs are often found in functionally important regions of RNAs[30] where they help to stabilize the active conformation. Interestingly, the presence of Ψ in RNAs also substantially contributes to the evasion of immune response triggers such as Toll-like receptors[31,32] and protein kinase R[33] – thus Ψs (and Ψ derivatives) are frequently built into mRNA vaccines.

Accumulating evidence also suggests that pseudouridylation of mRNAs can result in the incorporation of alternative amino-acids[34] and the read-through at premature stop codons (PTCs).[35–37] As a significant proportion (~10%) of pathogenic human mutations are attributed to PTCs, targeted pseudouridylation potentially provides a clinically relevant strategy to correct these mutations.[38]

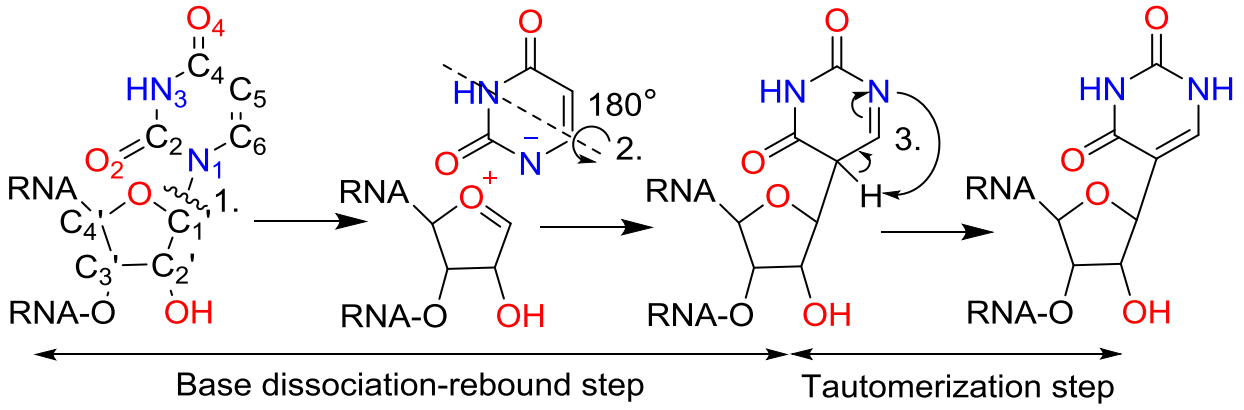
Most of the approaches aimed at the correction of disease-relevant human mutations through RNA editing are based on adenosine deaminases (ADARs) that are usually selective for specific sites in the genome.[39–41] This selectivity is a serious limitation in the putative therapeutic use of the native proteins. Recently two programmable RNA editing systems have been developed, both exploiting the Cas13/CRISPR system to obtain sequence specificity. In the originally described RNA Editing for Programmable A to I Replacement (REPAIR) system, the deaminase domain of human ADARs (ADARDD) has been fused to a deactivated Cas13b (dCas13b) protein.[42] This system was later developed into the RNA Editing for Specific C-to-U Exchange (RESCUE), where ADAR2DDs were *in vitro* evolved into cytidine deaminases.[43] In the case of REPAIR, the resulting chimeric protein enabled the correction of pathogenic human mutations through targeted RNA editing.[42] However, a major drawback of these approaches is the size and potential immunogenicity of the chimeric proteins that need to be delivered. For example, the packing capacity of adeno-associated viruses (AAVs), the most widely used type of viral therapeutic vectors, is limited to only 4.7 kb, thus they can barely fit the functional REPAIR and RESCUE constructs.

It has recently been shown that the Ψ-mediated read-through proceeds independently of the position of the PTC or the surrounding sequence of the mRNA[30], thus if dyskerin-mediated pseudouridylation events not sequence-limited, theoretically almost all PTCs could become targetable by the rational design of synthetic box H/ACA snoRNAs. Relying on the endogenous box H/ACA pseudouridylases, only the appropriate guide snoRNAs would need to be produced and delivered. The delivery could be achieved by the same methods that have been recently developed for microRNA therapeutics and would cause minimal immunogenicity.[44]

Some PSUs are stand-alone enzymes,[45–47] catalyzing the isomerization reaction without the aid of a guide RNA. In contrast, box H/ACA PSUs are complexes built from four protein chains and a unique guide RNA that forms the substrate binding pocket. Substrate RNAs that attach to the guide are positioned so that the uridine to be isomerized is thrust into the active site of the catalytic protein of the complex, dyskerin. Dyskerin is surrounded by three additional, essential core proteins, Nop10, Nhp2 and Gar1, which are all required for optimal reaction turnover.[48,49] The guide RNA is a box H/ACA snoRNA with an evolutionary conserved hairpin-hinge-hairpin-tail structure carrying an ANANNA (box H) sequence in the hinge region and an ACA triplet (ACA box) at the 3' tail.[50] The role of the guide RNA in the pseudouridinase reaction is to recognize the substrate RNA via base pairing interactions and to position it in the active site.[51] Nhp2 is suggested to stabilize the guide RNA and orient the substrate, while Nop10 stabilizes the active site of dyskerin.[20] Gar1 that binds to the complex further from the active site, presumably promotes the catalytic reaction[52,53] and/or facilitates product release.[51,54,55] Dyskerin, the catalytic unit, contains the active site, which is highly conserved in all PSU families (including the stand-alone enzymes).[50,54–56]

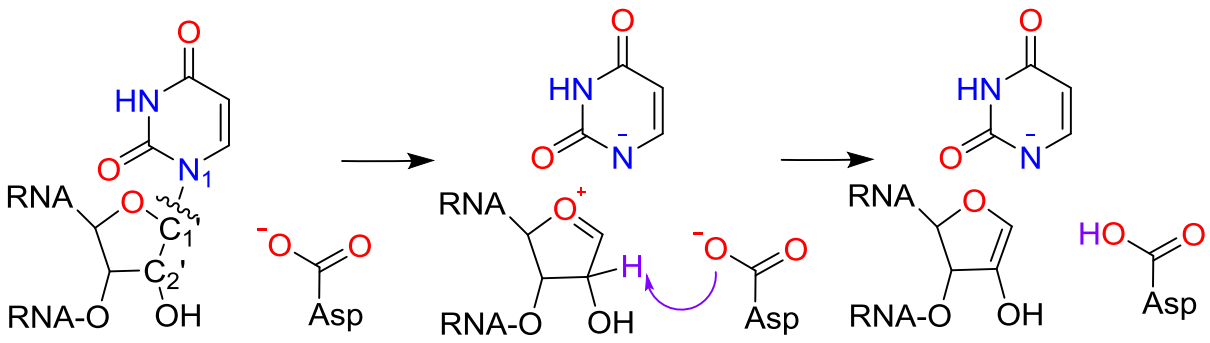
The atomic details of the pseudouridylation reaction mechanism have remained elusive. The reaction involves the cleavage of the glycosidic C_1'-N_1 bond, the rotation of the base and the formation of the new C_1'-C_5 bond. A proton transfer reaction ends the formal description of the reaction (Scheme 1).

Scheme 1. The formal mechanism of pseudouridine formation. The arrows in step three show the movement of the electron pairs.



Among the several mechanisms[57–60] proposed in the literature, the glycal mechanism (Scheme 2) was corroborated by kinetic isotope experiments for stand-alone TruB and RluA PSU enzymes in the presence of the physiological substrate. According to this mechanism, the breakage of the glycosidic bond and C_2' deprotonation of the ribose result in the glycal intermediate (Scheme 2, right side sugar moiety), where either the deprotonation of C_2' or its reprotonation was found to be partially rate-limiting.[61] Due to the planar nature of the glycal intermediate, the double bond between C_1'-C_2' of the sugar can be attacked from both sides. This provides an explanation why along the major *ribo* product, seen in the crystal structures[62,63] a minor product with *arabino* stereochemistry appears when TruB transforms the 5-fluoro uridine (5FU) substrate.[60] However, in case of the physiological U substrate, only the *ribo* product forms,[60] which indicates that the presence of the fluoro-substituent might fundamentally alter the reaction mechanism.

Scheme 2. The first two steps of the glycal mechanism investigated in this study



Previously we performed quantum chemical calculations on small model systems[64] and demonstrated that the base detachment either precedes C_2' deprotonation, or the cleavage and reattachment of the base proceeds entirely without a deprotonation step. The calculations also showed that the previously proposed Michael addition[58,59] and acylal schemes[59] are unlikely.

Although the active site of archaeal box H/ACA and various standalone PSU enzymes exhibit similarities,[50,54–56] the single-turnover rate of H/ACA PSU is two orders of magnitude lower than those of standalone Ψ-synthases like TruB, RluA or TruA.[65] This raises questions concerning the role and purpose of an apparently less effective but significantly more complex enzyme that requires the cooperation of 4 proteins and 2 RNAs for accomplishing a task that can be carried out more efficiently by a single protein. TruB, TruA and RluA, however, impose strict sequence requirements,[66–68] utilizing an extended protein-RNA interaction surface to identify substrates.[26] The complex machinery of box H/ACA PSU allows for a fundamentally different approach, here the substrate RNAs are coordinated almost exclusively by the guide RNA, over a relatively short segment of 12 base-pairs.

To better understand the functioning of PSU enzymes, atomic level insight into the reaction mechanism would clearly be necessary, however, experimental techniques could not reveal all the required details. Molecular dynamics simulations using enhanced sampling methods and hybrid quantum mechanics/molecular mechanics potential based methods (QM/MM) have been successfully applied to study DNA and RNA containing complex systems previously[69–72] but the modelling of a system as large as the complete box H/ACA pseudouridine synthase still remains challenging. Here we aim to build and investigate the structure of the human box H/ACA PSU complex to elucidate the structural and sequential requirements of pseudouridylation and deduce the sequence requirements posed by the catalytic reaction concerning guide RNAs. We explore

the key steps of the reaction mechanism in atomic detail using mixed quantum mechanics/molecular mechanics simulations and the results are discussed in the light of the experimental data primarily obtained for stand-alone enzymes. Furthermore, calculations were also performed for inactive mutants to reveal how mutations affect the structure and the catalytic activity of the complex. These studies contribute to our understanding of the detailed mechanism of pseudouridylation as carried out by box H/ACA PSUs and our results could also be employed in designing targeted RNA editing strategies for the treatment of pathogenic human mutations.

2. **Computational Methods**

2.1. MD simulations

Full and partial box H/ACA PSUs from *Pyrococcus furiosus* (PDB id: 3hay[51]; 3hjw[54]; 3lwq[73]) and *Saccharomyces cerevisiae* (2lbw[74]; 3u28[75]) were used to build a homology model of the human enzyme (Figure S1) using Schrödinger Modeling Suite,[76] containing residues 60-380 of dyskerin (O60632), residues 1-49 of Nop10 (Q9NPE3), residues 34-153 of Nhp2 (Q9NX24) and residues 67-159 of Gar1 (Q9NY12). The structure thus obtained was subjected to molecular dynamics (MD) simulation at 310K using GROMACS,[77] with the AMBER-ff99SBildnp* forcefield[78] and the parametrization of Steinbrecher et al.[79] for the phosphate moieties. Dodecahedral simulation boxes were used, allowing a buffer distance of 10 Å between the protein and the wall of the box, resulting in box vectors of nearly 120 Å. OPC water model[80] was used (with over 37000 solvent molecules in each system), the total charge of the system was neutralized and physiological salt concentration was set using Mg^{2+} and Cl^- ions (~ 80 Mg^{2+} and ~120 Cl^- ions). (We added Mg^{2+} instead of Na^+ to closely model the physiological environment as the complex is localized mainly in the nucleolus.[81,82]) The final models thus contained a total

of ~160000 atoms. Protonation state of titratable residues at pH=7 was checked using the H++ server.[83] The energy minimization of the starting structures was followed by relaxation of constraints on protein atoms in three steps (holding all non-hydrogen atoms to their position with 1000, 500 and 100 kJ/(mol*nm^2) force constants), an additional unconstrained NVT step at 300 K and a final unconstrained NPT step (all 200 ps) to allow stabilization of pressure (1 bar) before the production run. Trajectories of 600-2500 ns NPT simulations at 310 K and 1 bar were recorded for analysis (collecting snapshots at every 4 ps). The last 300ns of the simulations (Figure S2) were clustered (based on the conformation of the backbone of the protein components using a 1 Å cutoff). In order to investigate the reaction mechanism, 5 systems were selected for MD: the wild type enzyme containing the target U (WT/U) and the product Ψ (WT/Ψ) and three catalytically inactive mutant enzymes, those of D125A, D125N and E206K with a uridine at the target site, referred to as D125A/U, D125N/U and E206K/U respectively. All studied mutations are located in dyskerin; the catalytic unit of the RNP complex.

2.2. QM/MM MD

QM/MM MD simulations were initiated from representative structures of the MD simulations. The box H/ACA PSU complex was extracted from the dodecahedral water box used in the MM GROMACS simulations and was immersed in an octahedral water box generated by the AmberTools package. The total system contained ~120000 atoms. The simulations were preceded by a stepwise relaxation and equilibration. It started with a pure MM minimization procedure (5000 steps with 100 kcal/mol/Å^2 restraints on the protein and RNA residues followed by 10000 steps with the release of restraints) to eliminate steric clashes. Then the system was heated to 310 K in 200 ps and it was further equilibrated at the final simulation temperature for 500 ps in an NPT ensemble. After the initial MM equilibration, we further prepared the system

by running 10 ps of unbiased QM/MM MD simulation applying a 1 fs time step at 310 K in NVT ensemble using a Langevin thermostat with 2 ps^{-1} collision frequency. The QM region included the uridine with the phosphate group of the adjacent nucleic acid, the side chain of Asp125, a water molecule and the Tyr153 sidechain (Figure 2a). Other QM regions were also investigated to check the effect of the size of the QM region on the results (Figures S3 and S4). The link atom approach was used to saturate dangling bonds at the QM-MM interface. The QM part was handled at the DFTB3[84] level of theory and AMBER14 forcefield[78] was used for the MM region. Periodic boundary conditions were applied and long range electrostatic interactions were calculated with the Particle Mesh Ewald method.[85,86] The QM region for the simulation of the tautomerization in bulk water is shown in Figure 2b.

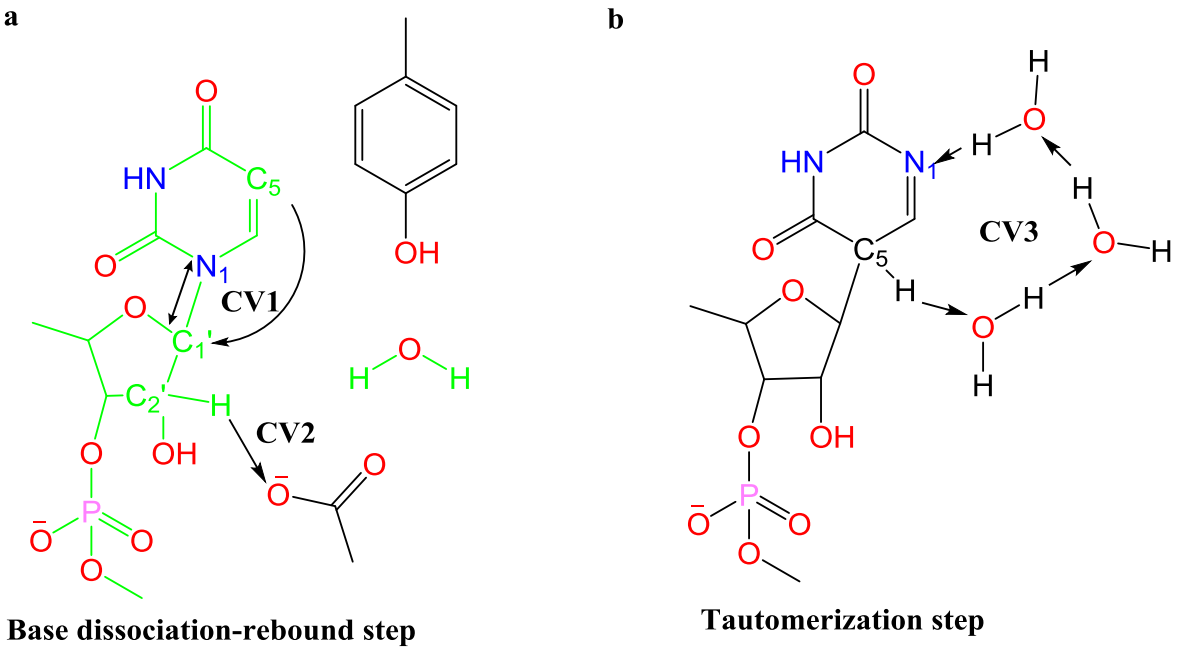


Figure 2. a) The QM region of the QM/MM simulations (with the MM region containing the full protein/RNA environment) applied for the study of the base dissociation-rebound step. The arrows indicate the atomic movements described by the CVs. The first combined CV drives the C_1'-N_1 bond cleavage and the C_1'-C_5 bond formation, while the second CV describes the C_2'

deprotonation by any potential candidate atom including the aspartate oxygens, base hetero atoms or water. In the figure only the catalytic aspartate is indicated for the sake of clarity. The molecules with green bonds were included in the model system for the tautomerization step. b) The QM region in the bulk water simulations for the tautomerization part of the reaction. CV3 describes the proton transfer from the C_5 carbon to the N_1 atom via a water chain.

The free energy surfaces of the reaction steps have been explored by two accelerated free energy methods: conventional metadynamics and steered molecular dynamics simulations using the Amber program package[87] combined with Plumed.[88] The details of the simulations are given in the Supporting Information. To address the possible reaction mechanisms several different combinations of collective variables (CVs) have been tested. This is because the relevance of a chosen set of CVs and the corresponding mechanism can be confirmed only with simulations. For example, we have explored the deprotonation of ribose moiety using CVs describing the proton transfer from the sugar to the nearby aspartate side-chain, because aspartate protonation by the C_2' carbon has been proposed earlier as a necessary ingredient of the formation of pseudouridine. On the basis of the preliminary barrier estimation by metadynamics, however, we have concluded that this route is not feasible under the given conditions.

While steered molecular dynamics and metadynamics employing suitable CV sets for each reaction step are appropriate for the preliminary exploration of the free energy surface, umbrella sampling with the weighted histogram analysis method is able to generate accurate free energy surfaces. Therefore, once the preliminary explorations of the free energy surface of the full reaction route have been carried out, the free energy profile of each step was refined with umbrella sampling and weighted histogram analysis (see details in the Supporting Information). The validity of the located transition states (TS) has been subsequently verified by committor

analysis (unbiased molecular dynamics have been performed from 80 geometries taken from the transition state ensembles with random velocities at the temperature of the simulation). The TS was accepted if approximately 50 % of the simulations reached the initial and 50 %the product states of the given reaction step.

The following sets of CVs have been found suitable to describe the reaction steps: for reaction steps 1 and 2: CV1 describes the C_1'-N_1 bond cleavage and C_1'-C_5 bond formation with a coordination number based collective variable, while CV2 describes the deprotonation of the C_2' sugar carbon by any potential candidate atom including the aspartate oxygens, base hetero atoms or water. CV3 for steps 3 and 4 describes the proton migration from uridine C_5 atom to the nearby water chain and the protonation of the uridine at the N_1 position by the water chain. (The functional form of the CVs and additional sets of CVs describing unsuccessful reaction paths are discussed in the Supporting Information.)

3. Results and Discussion

3.1. Structure of the complexes and conformational strain of the coordinated substrate

Since the structure of the human box H/ACA PSU ribonucleo-protein complex (RNP) has not been determined yet we have derived the initial models for our study by homology modeling, relying on crystal structures of PSU enzymes from bacterial sources in complex with various ligands. These, however, contain either mutated enzymes or modified substrates and therefore capture either pre-reactive or post-reactive stages of the catalytic cycle. Complexes with 5-fluorouracil (5FU) derivatives of the substrate are often frozen in a post-reaction stage with the new C-C bond already in place, while O-halogenated or methylated variants inhibit the reaction. Thus, homology modeling and the subsequent MD simulations were not only required for

implementing the human sequence but to establish the architecture of the active site in the presence of the physiological substrate. Molecular dynamics (MD) simulations were carried out both for the wild type (with substrate RNAs carrying both uridine and pseudouridine at the catalytic position) and catalytically impaired mutant enzyme complexes[20,59,89] (with the D125A, the D125N and the E206K mutant variants of dyskerin) (Figure 3). We found that the structural ensemble generated for the model of the human box H/ACA RNP retains the topology of the *Pyrococcus furiosus* RNP[64] with best fit achieved in the central, catalytic region (Figure S1). We also found appreciable agreement with the AlphaFold predictions[90] although the latter were carried out for the uncomplexed form of each protein (Figure S5). Moreover, the MD derived conformations of the guide and substrate RNAs in the full RNP complex show similarity to the solution state, NMR structure of the (enzyme-free) complex of the human U65 snoRNA (guide) and a segment of the 28S rRNA (substrate) (Figure S6).[91] The observed similarity illustrates that the binding topology and flipping of the substrate uridine is already guaranteed by the complex formation of the two RNA strands.

Guide and substrate RNAs were taken from the crystal structure of the box H/ACA RNP of *Pyrococcus furiosus* in complex with a substrate RNA carrying 3-methyl-uridine at the catalytic position (3lwq) - the closest analogue of the native substrate from among those with an available crystal structure. Following demethylation at N_3, the starting structure of the simulations thus contained a substrate uridine base in an (intermediate) *syn* orientation. During the simulation, the wild-type uridine of the reactant state relaxed to the preferred *anti* conformation. Conversely, in case of the pseudouridine carrying WT complex (product state) or any of the catalytically impaired mutants considered here, the orientation of the nucleobase shifts toward the less preferred intermediate region between the *syn* and *anti* orientation (see Table S1).

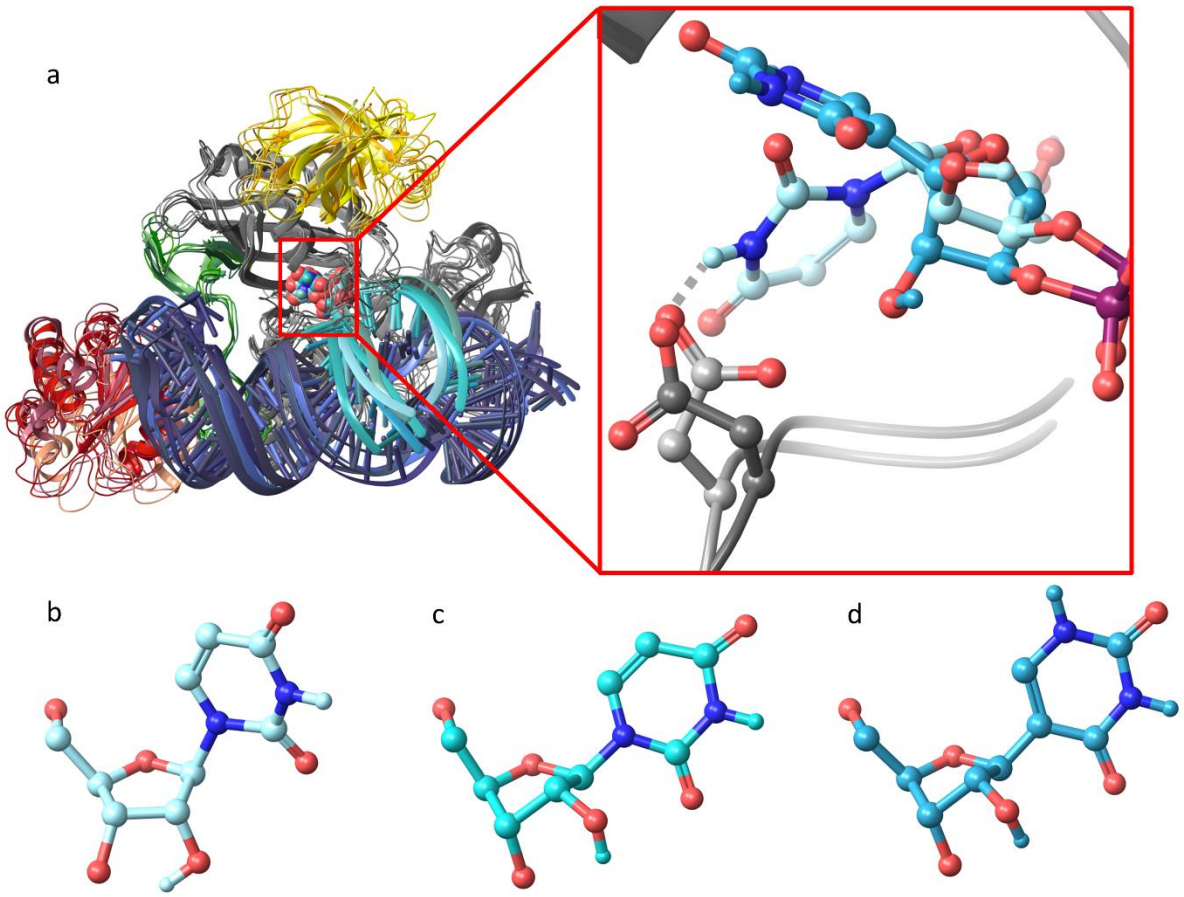


Figure 3. a) The middle structures of the most populated clusters of the equilibrium MD trajectory of the WT/U, WT/Ψ, D125A/U, D125N/U and E206K/U box H/ACA PSU complexes, using different shades of the coloring scheme introduced in Figure 1. In the inset the active site is shown for the reactant WT/U and the product state WT/Ψ (WT/U complex is lighter, WT/Ψ darker). b)-d) Close up of the target nucleotide in the WT/U, D125N/U and WT/Ψ systems, respectively showing the differences in sugar puckering and base orientation.

The WT, uridine-bound complex (the starting state of the pseudouridylation reaction cycle) is unique from another perspective also. In this system the N_3 atom of the base is coordinated to the catalytic Asp125 residue by a hydrogen bond (H-bond) (in 99.5% of the snapshots), while no

such connection or only partial interaction was found in any of the other systems considered; Asp125(OD1, OD2)···N_3 H-bond was seen in 1.6, 0, 14.3, 29.0% of the snapshots of WT/Ψ, D125A/U, D125N/U and E206K/U complexes, respectively. This connection together with the steric confinement of the binding pocket introduces a significant distortion of the bound nucleotide, resulting in the C_4'-C_1'-O_4 angle decreasing from the generally seen ~130° to 107°, while the C_1'-N_1-N_1-C_4 torsion increased to ~210° from ~170° found in the reference set. Neither Asn or Ala in place of the catalytic Asp, nor the Ψ product next to the native Asp induced bending of the bound nucleotide. This only happens if a uridine is docked into the active site containing the WT Asp (see Table S1).

Beside the presence of Asp125 in the catalytic pocket, steric reasons also contribute to the distortion of the reactant state complex of the WT pseudouridinase. When binding to the undisturbed protein structure of the WT box H/ACA RNP, the substrate encounters a very shallow catalytic pocket – limited in size by a "hydrophobic block" created by Val128-Ile226-Leu246 residues (Figure S7). The significance of this interaction is underlined by our subsequent QM/MM calculations. Here the position of the catalytic Asp residue slightly shifted with respect to the substrate U, to a conformation where direct H-bonding to the N_3 of uracil is replaced by a water-assisted interaction, while placing the Asp125 close enough to the O_2'-hydroxyl of the sugar for direct H-bonding (an interaction also present in a minor fraction (4.4%) of the MD snapshots). The orientation and distortion of the base within the crammed active pocket, however, was maintained (see below, and Figures S8 and S9).

It seems that the box H/ACA PSU enzyme uses the energy of substrate-binding to weaken the glycosylic bond and facilitate the reaction – and also to recognize substrates to be isomerized. This observation is in line with the catalytic mechanism proposed for human uracil DNA

glycosylase (UDG), an enzyme that initiates DNA repair through base-excision.[92] The crystal structures of the UDG enzyme in complex with 2′-deoxypseudouridine (dΨ), selected as an uncleavable substrate analogue, revealed a strained active site conformation with the pseudouridine ring rotated halfway between the *syn* and *anti* orientation along with the sugar puckering distorting to a mild C_3'-exo conformation stabilized by several hydrogen bonds and π-π stacking. This orientation was proposed to induce the p-σ* overlap with O_4' (anomeric effect) and also to facilitate the overlap of the glycosidic bond and the carbonyl π system. This conformational strain and stereoelectronic effects were suggested by the authors to ease the cleavage of the glycosidic bond. Another analysis of the destabilization of the ground state of the UDG complex emphasized the role of tautomeric strain created by nearby H-bonding residues and long range electrostatic interactions.[93] In our results, in case of box H/ACA PSU, the distortion of the substrate base of the WT/U complex was shown to be a key feature of the starting state of the pseudouridylation reaction (further discussed in connection with the QM/MM simulations).

In the product state WT/Ψ PSU complex and in case of the mutant enzymes, a base distortion in the other direction was seen. In these systems - in the absence of the Asp125-N_3 connection - the nucleotide bound to the active site is flipped to an unusually upright position, with C_4'-C_1'-O_4 angles of 159.1 ± 6.9°, 148.3 ± 4.7°, 140.9 ± 6.1° and 143.1 ± 8.3°, in the WT/Ψ, D125A/U, D125N/U and E206K/U complexes, respectively. This upright position and the emergence of the N_1-H_1 new H-bond donor on the Ψ ring allows the pseudouridine of the product state complex to form new H-bonds with the upper region of dyskerin that communicates both with Nop10 and Gar1, responsible for facilitating product release. In over 75% of the snapshots, N_1-H_1 forms H-bonds with either the sidechain hydroxyl group of Tyr153 or the backbone carbonyl oxygen of

Thr224, positioning the ring so that O_2 (opposite from the newly formed sugar-base C_1'-C_5 bond) reaches the backbone amides of Ile226 and Arg227 (Table S2). Comparing the calculated B-factors of the complexes revealed an unexpected consequence of these interactions: in the pseudouridine bound system the upper region ("fluctuating domain": residues 150-252 (see Figure S10)) of dyskerin and the entire Gar1 molecule become more flexible than in any other system considered here (an increase of 2.6- and 3.0-fold was seen in dyskerin (150-252) and Gar1, respectively, compared to the B-factors of the same regions of the WT/U complex) (see Figure S11). The loss of the Asp125-N_3 connection in itself leads to a slight increase in the conformational heterogeneity of these regions, as seen in case of the D125A/U, D125N/U and E206K/U systems (1.4- and 1.3-, 1.4- and 1.1-, 1.0- and 1.2-fold increase in the B-factors of dyskerin(158-243) and Gar1, in case of the D125A/U, D125N/U and E206K/U complexes, respectively) but it is considerably magnified in case of the pseudouridine bound form. The sampled conformations describe the concerted tilting of the upper dyskerin region, comprising the thumb loop (residues 184-192) that locks and fixes the substrate to the protein components of the complex, and Gar1.

Thus, it seems that the shift of the pseudouridine ring to a more relaxed, upright position simultaneously relieves the conformational strain of the distortion and offers new interaction possibilities for the emerged N_1-H_1 H-bond donor of the Ψ ring. This however requires the slight dislodging of the "fluctuating domain" to restructure the "hydrophobic block". This suggests that product release might take place - maintaining the strong interaction between these regions and the substrate – by "peeling" away the substrate from the guide snoRNA in a coupled backward motion, rather than by the simple opening of the thumb loop and the spontaneous scission of all the H-bonds between the 12 base-pairs that exist between the substrate and the guide.

We also examined the coordination-mode of the guide and substrate RNAs to the enzyme complex to uncover any sequence specific requirements. (We found these interactions to be nearly identical in all studied systems, thus the numbers presented here are taken from the simulation of the WT/Ψ complex which had the longest running time.) A total of 51.9±4.2 H-bonds are formed between the guide RNA and the various protein components of the box H/ACA PSU, mostly between the sugar-phosphate backbone of the guide and various protein components. Out of these 20.5±2.0 H-bonds are sequence specific (nucleobase···protein type), while the rest is formed between the sugar-phosphate backbone and protein atoms. 8.0±1.2 of the base-specific interactions concern the coordination of the ACA motif (and its surroundings) to the PUA region of dyskerin (distant from the active site, see Figure S12), 10.0±1.7 are formed between hairpin region of the guide RNA and Nhp2, while only 2.5±0.6 between the pseudouridylation-pocket and the active site of the enzyme. These latter H-bonds are formed with long, thus flexible sidechains of Glu104, Lys144 and Gln147 of dyskerin, which do not impose a strict sequence requirement on the guide, as opposed to over half of the H-bonds that contribute to the recognition of the ACA-motif, which are formed with main-chain atoms of the protein (Tyr311, Lys314, Gly361, His360), thus with the fold of the enzyme itself. We found no base-specific requirements for the coordination of the substrate RNA by the protein. Nearly 80% of the approximately 16 H-bonds formed between the substrate and dyskerin (the substrate only interacts with the guide snoRNA and dyskerin) are formed between the sugar-phosphate backbone and the protein, the only nucleobase that forms base-specific association with the protein matrix is the uridine to be isomerized. Taken together these findings suggest that guide snoRNAs which will be able to align to and activate the box H/ACA PSU enzyme maintaining the basic topology seen here can be varied rather freely in the pseudouridylation-pocket region

indicating that pseudouridylation can be induced in a great variety of RNA substrates. Therefore, the box H/ACA PSU enzyme seems – based on the MD simulations – to be an ideal vehicle for RNA editing.

3.2. Investigation of the catalytic mechanism by QM/MM calculations

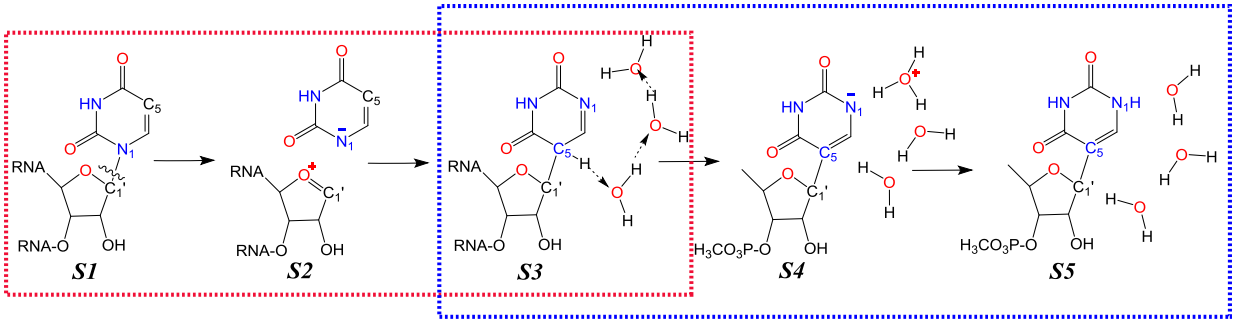
Steered molecular dynamics and metadynamics simulations were performed to explore the mechanism and umbrella sampling simulations were used to generate reaction free energy profiles.

3.2.1. Base dissociation-rebound step

To clarify a potential cooperative mechanism of both the sugar-base dissociation and rebinding step and sugar C_2' deprotonation and reprotonation step, we ran 2D metadynamics QM/MM simulations containing a CV describing the C_1'-N_1 cleavage and C_1'-C_5 bond formation and another one describing the C_2' deprotonation by any potential proton acceptor nearby (Figure 2a).

Results of the simulations indicate that the reaction likely follows a dissociative mechanism for the base dissociation-rebound step (Scheme 3). The rotation of the uracil ring and the new bond formation closely follows the C_1'-N_1 bond cleavage without proton detachment from C_2'. The bond cleavage results in a ring plane orientation almost parallel to the sugar plane, similarly to that observed in some crystal structures.[94,95] The reorientation of the ring is linked to the bond cleavage and formation of the C_1'-C_5 bond rather than being a separate step. We observed that various ring rotations can lead to a ring position suitable for reattachment with C-C bond formation.

Scheme 3. The reaction mechanism of box H/ACA pseudouridine synthase proposed by the QM/MM free energy calculations. The reaction starts with the base dissociation-rebound step (the



steps inside the red dashed rectangle) modelled using the full protein-RNA environment. The tautomerization reaction (the steps inside the blue rectangle) was calculated using a model system in water. The proton on the C_5-atom (***S3*** state) is transferred to the N_1-atom (***S5*** state) via a water chain. The ***S4*** state appears as an intermediate when the water chain is modelled with three water molecules (see the text for discussion).

The umbrella sampling simulations estimate a dissociation barrier of 14 kcal/mol for the base detachment. We note here that according to our test calculations DFTB3 underestimates the energy by ~10 kcal/mol compared to B3LYP/6-311+G** (see the Supporting Information). Still, the obtained barriers are similar to that found in case of the UDG dissociation step at the AM1 level of theory (14.9 kcal/mol).[96] The free energy curve between the ***S1*** and ***S3*** states includes a plateau-like region with a shallow minimum for the ***S2*** state located 10.7 kcal/mol above the reactants, and separated by only 0.6 kcal/mol barrier from the carbon reattachment. We note that this latter barrier is within the accuracy of the simulations. After the new carbon-carbon bond formation (***S3***), the energy drops to 1.5 kcal/mol compared to the reactants forming a stable intermediate. The representative structures for the ***S1***-***S3*** states (Scheme 3) and the transition

states with the most important geometric parameters are shown in Figure 4. The adequacy of the QM region was validated by repeating the calculations with more extended QM regions that include nearby charged residues and H-bonding partners of the phosphate adjacent to the target uridine (Figures S3 and S4). The free energy profiles are similar with a ~2 kcal/mol barrier increase for the larger QM regions. These results show that the applied QM region (Figure 2a and WT-QM1 in Figure S4) with its interacting charged and H-bonding partners in the MM region represents a good model for the base dissociation-rebound step using QM/MM calculations with electrostatic embedding.

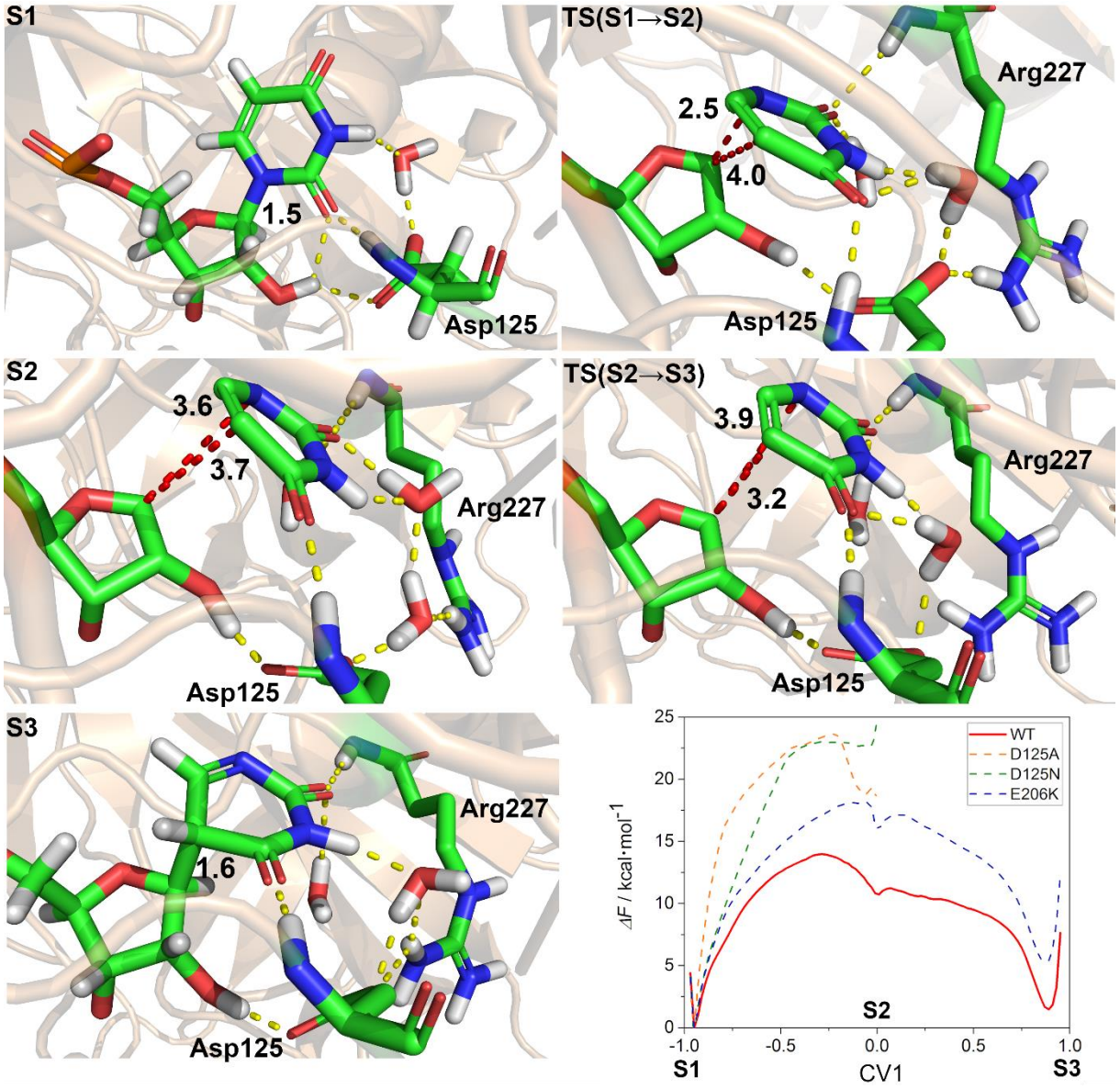


Figure 4. Representative geometries for the stationary points and the energy profiles of the base dissociation-rebound step. The approximate values of the most important structural parameters characterizing the states - the C_1'-N_1 and C_1'-C_5 distances - are indicated (in units of Å). Yellow dashed lines display the hydrogen bonds. The energy profiles obtained for the WT (red), D125A (orange, dashed), D125N (green, dashed) and E206K (blue, dashed) enzymes by umbrella sampling are shown. In the case of the inactive D125A and D125N mutants the **S2** (detached

base) state was reached with high barriers and these routes were not investigated beyond the **S2** state.

The distorted conformation of the substrate uridine ring positions the C_5 carbon closer to the ribose moiety and initializes the catalytic reaction. In the transition state, several hydrogen bonds formed between protein side chains, and coordinated water molecules promote the lowering of the transition state energy. In the intermediate state, both the C_1'-N_1 and C_1'-C_5 distance is above 3.6 Å, while the major hydrogen bonds observed in the transitions state still remain intact. In the second transition state, the C_1'-C_5 distance decreases, though otherwise the structure strongly resembles ***S2*** in line with the energy profile showing rather a plateau than a deep minimum. The fairly confined active site containing mainly by hydrophobic side chains hinders the movement of the base forcing it to rotate instead of just moving away during the bond cleavage step.

The dissociative mechanism is in line with the proposal of Drohat et al.[97] that most ribonucleotides are hydrolyzed via a dissociative mechanism. For uracil-DNA glycosylase (UDG), an enzyme that plays a crucial role in the base excision repair mechanism, a similar stepwise S_N1 mechanism was proposed by kinetic isotope effects based transition state analysis[98] as well as QM/MM calculations.[71,96] We have to note here that while the catalytic reaction of UDG enzymes starts with a base detachment step as in the case of PSUs, it is not followed by rotation and reattachment of the ring to achieve isomerization of the nucleotide, but by water attack on the C_1' carbon and the subsequent release of the hydrolyzed nucleobase product.

To further clarify the role of the catalytic aspartate we initiated simulations for the D125A and D125N as well as the E206K inactive mutants. We applied the same simulation parameters in these computations, so the results are directly comparable. Overall, the mutants display similar energy profiles compared to the wild type enzyme, however the barriers are increased to 23.5

kcal/mol, 23.0 kcal/mol and 18.2 kcal/mol for the D125A, D125N and E206K mutants, respectively (Figure 4), owing to the missing conformational strain.

To estimate the energetic cost of the distorted binding of uracil, we carried out scans (B3LYP-D3/6-31G** C-PCM water solvation) of the C_1'-N_1-N_3-O_4 dihedral of the nucleoside, fixing its "backbone" atoms (O_3', C_3', C_4', C_5' and O_5') to their position in the RNP complex. A comparison of the energies of the scan and the C_1'-N_1-N_3-O_4 dihedral angle ranges of the QM/MM simulations indicates an energetic cost of ~8 kcal/mol for the distortion of uracil in the WT/U complex, while marginal energetic cost is attributed to the dihedral range found in case of the mutant complexes (Figure S13).

We emphasize that the energy is increased locally as a fraction of the binding energy is stored in the conformational energy of the substrate, and the total energy of the Michaelis complex is lower than that of the reactant state with non-interacting reactants, i.e. the Michaelis complex represents the resting state,[99] and it was used as a reference in barrier calculations. The distortion of the C_4'-C_1'-O_4 angle and the C_1'-N_1-N_3-O_4 torsion angle of uracil toward the orientation occurring in the transition state can be seen in the wild type complex, while this cannot be observed in the mutated complexes (Figures S9c,d). This is in line with the conclusions drawn based on our classical MD simulations (Figures S9a,b) and experimental data showing that aspartate to alanine or asparagine mutation of the WT Asp125 leads to a complete loss of activity (even mutation to glutamate is not tolerated for most species),[59,89,100] and the dysfunctionality attributed to the E206K variant.[3] These results support the significance of the distorted uridine conformation forced on the substrate upon binding and emphasize the catalytic role of Asp125 residue.

The proper positioning of the target uridine is assured by base pair formation between the substrate and guide RNA within the pseudouridylation pocket; only the target uridine and one adjacent nucleotide is free from these interactions. This reduces significantly the conformational freedom of the bound substrate uridine within the box H/ACA pseudouridine synthase and contributes to reactant destabilization through the observed strained conformation that weakens the C_1'-N_1 bond. In contrast, stand-alone pseudouridine synthases that directly bind their substrates are unable to restrict the conformational freedom in the same way and this may lead to a different catalytic mechanism in these enzymes. Indeed, although the active sites of PSUs are fairly similar, experimental evidence suggests that PSU enzymes may not share exactly the same reaction mechanism. A striking example for their different behavior is their reaction with 5-fluorouridine, as some Ψ-synthases are inhibited by 5FU while others efficiently handle it.[95,101,102] According to our computational results the base dissociation and rebound step in case of the human box H/ACA pseudouridine synthase is not accompanied by the C_2'H deprotonation and this is at variance with the mechanism proposed for TruB and RluA where small first order kinetic isotope effect was reported for the C_2'H proton.[60,103]

Another factor appearing to contribute to substrate strain is the closed conformation of the thumb loop of dyskerin. The thumb loop adopts a closed conformation upon substrate binding[51,54] and the open conformation in the absence of substrates,[104] while crystal structures formed with nonreactive targets or modified products revealed intermediate conformations.[73,105] Furthermore, the thumb loop was proposed to sense the progression of the reaction[101] and also plays a role in product release by opening up upon pseudouridine formation[53] (as indicated by our classical MD simulations, where we also detected the backward flip of the "fluctuating domain" of dyskerin – which contains the entire thumb loop - in the product state). The incomplete closure of the loop

might explain the nonreactive nature of the 5-bromouridine[105] and 3-methyluridine[54] substrate analogues corroborating our finding that the positioning of the target uridine in the active site is a major factor for the catalytic reaction.

3.2.2. Tautomerization step

The rebound base has to undergo tautomerization to arrive at the pseudouridine product. The tautomerization includes the deprotonation of C_5 resulting in an sp^2 arrangement around this carbon atom, and the protonation of the N_1 atom of the base (Scheme1 last step). QM/MM MD simulations initiated for the proton transfer via proton relay systems included either the conserved Asp125 residue or water chains combined with the Tyr153. The inclusion of Tyr153 was motivated by the observation that the Y153A mutation in HfCbf5 and PfuCbf5 leads to partial uridine transformation resulting in both unmodified uridine and pseudouridine[89] suggesting that Tyr153 participates in the reaction, although its presence is not essential.[73] Tyr153 is also one of the H-bond partners of Ψ in our classical MD simulations. However, none of the models led to feasible reaction profiles as the barrier heights exceeded 30 kcal/mol for each pathway tested.

These simulations were run with a closed thumb loop conformation thus in a fairly small active site, which is unfavorable for the formation of an effective proton relay chain. The thumb loop and the attached Gar1 plays a major role in the catalysis and product release.[51,53] Both our results and those of Wang et al. indicate that the conformational change of the thumb loop is triggered by the U to Ψ transformation and Gar1 amplifies the differences between the dissociation of uridine and pseudouridine containing RNA chains.[52] We also observed several fold increase of the B-factors of the thumb loop and Gar1 in the MD simulations with the pseudouridine complex compared to the uridine complex. These findings and the high barriers obtained for the tautomerization step in the closed active site suggest that a more open protein conformation

would be appropriate for the formation of a proton chain and lowering the barrier of these steps. The proton transfer reaction does not require large atomic movements, and the local environment may determine the pathway. Therefore, the final steps of the mechanism were investigated by model calculations in bulk water. This approximation was adopted as we were unable to reliably model the partially open protein conformation where sufficient space would be provided for the proton transfer while also allowing for the participation of protein components in the reaction. Thus, the tautomerization barrier of our simulations in bulk water represents an upper limit. Moreover, these calculations cannot take into account the structural differences between mutants. Although we found a high barrier for the initial C_1-N_1 bond cleavage for each investigated mutant (*cf.* with free energy profiles in **Figure 4**), we cannot exclude that the C_1-N_1 bond cleavage and C_1-C_5 bond formation occurs in a mutant that reaches **S3**, the initial state of the tautomerization. Whether the tautomerization is affected by the structural differences of mutants cannot be addressed by our simulations in bulk water.

The simulated reaction pathway is shown on Scheme 3. The free energy simulations disclosed that C_5-H deprotonation by water precedes N_1 protonation. The step led directly to the **S4** state without additional barrier for the proton exchange between the water molecules. The umbrella sampling estimates an 18.9 kcal/mol barrier (Figure 5) with respect to the **S3** state. The **S4** state corresponds to the deprotonated base and the presence of a labile H_3O^+ species generated by the barrierless proton transfer from the first water (that extracts the C_5-H proton) to the third water, via the second water between them (Figure 6). The **S4** state is reached with a 14.5 kcal/mol free energy drop with respect to the transition state and it transforms to the **S5** state with an additional 14.7 kcal/mol free energy gain. The barrier between the **S4** and **S5** states is around 1 kcal/mol according to our calculations. This small barrier is within the accuracy of our simulations thus it

cannot prove or disaffirm the existence of the **S4** intermediate that contains the H_3O^+ species. We note, however, that while the emergence of H_3O^+ species has been shown by experimental[106,107] and computational[108,109] means, its occurrence in a proton relay may be model dependent.[110] We also note that the deprotonation of C_5 restores the planarity of the base in the **S4** state. The total reaction (**S1→S5**) is exergonic as it is expected, and the total free energy gain is nearly 9 kcal/mol (Figure 7).

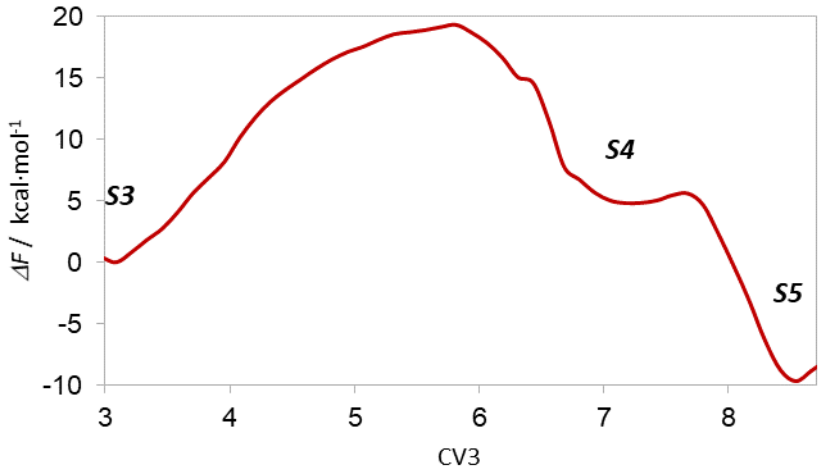


Figure 5. Energy profile for the **S3→S4→S5** reaction steps as obtained by umbrella sampling for a model system in water. The model includes a chain of three water molecules in the QM region. The reaction barrier corresponds to the C_5-H bond breakage. A shallow minimum (**S4**) with a H_3O^+ species appears before the proton transfers to the N_1-atom (**S5**).

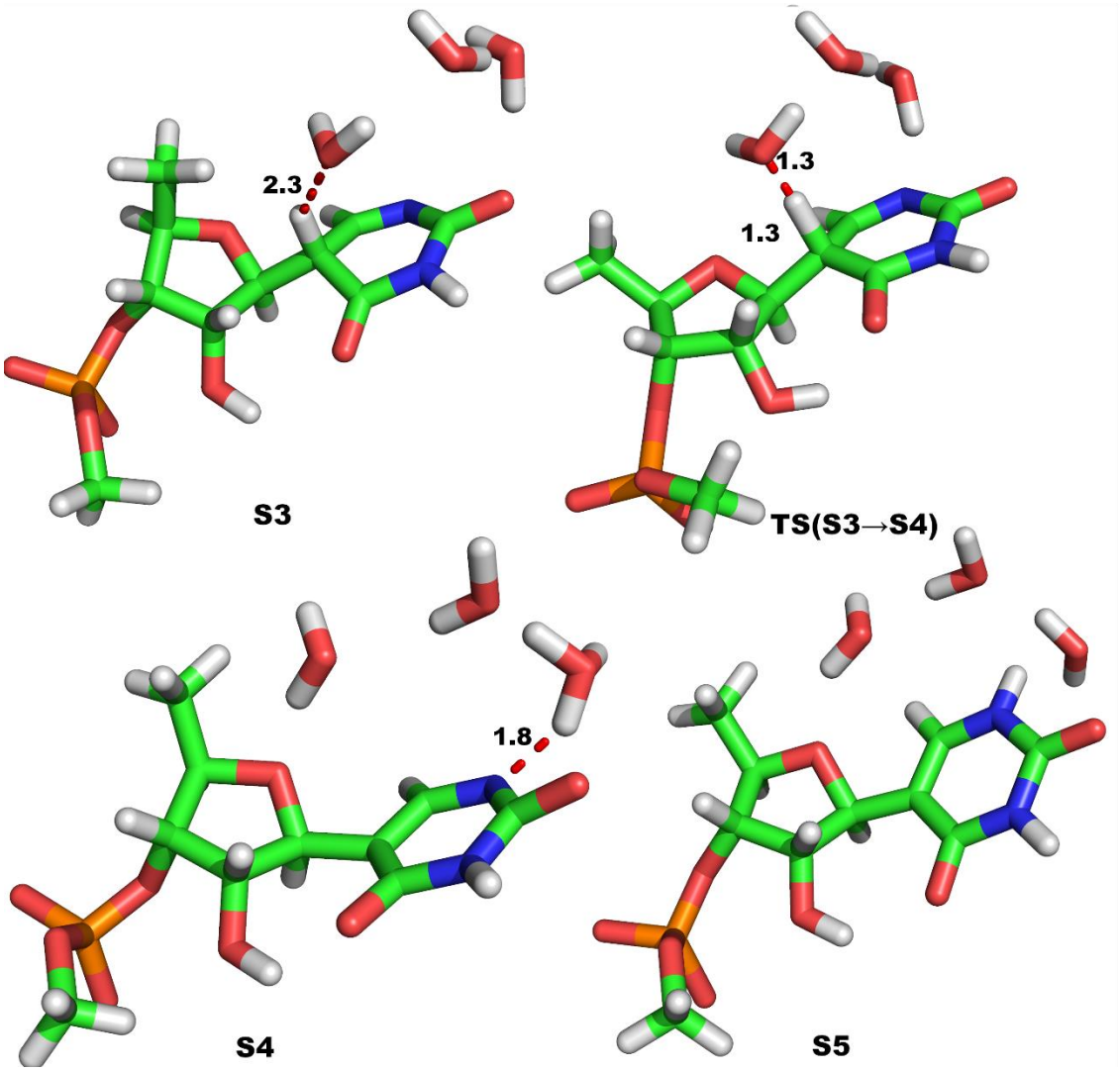


Figure 6. Representative structures of the stationary points of the tautomerization step. Characteristic interatomic distances are shown with dashed red lines. The **S4** and the **TS** between the **S4** and **S5** states are highly similar and only the former is shown.

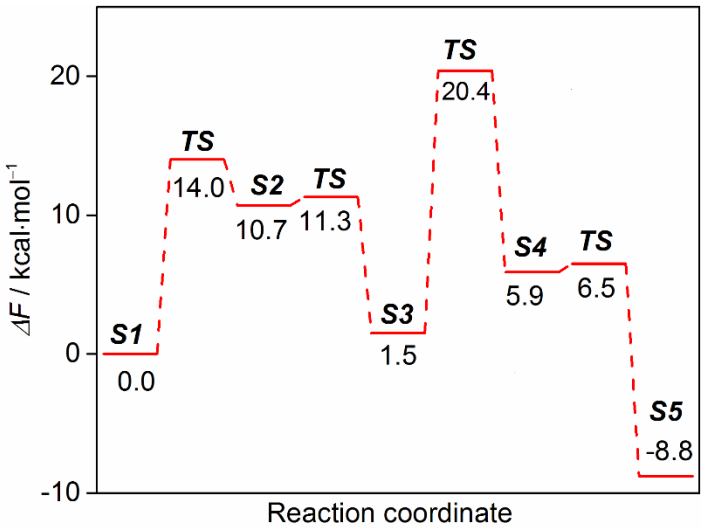


Figure 7. The energy profile of the full reaction pathway estimated by the umbrella sampling calculations. The model system contains the full solvated RNP complex for steps **S1→S3** and selected atoms (see **Figure 2**) in water for steps **S3→S5**.

The processes are localized, and the movements of the molecules participating in these reaction steps are fairly limited. Therefore, the results obtained for the model are expected to be valid for an open, water accessible protein active site. The elevated barriers can be related to the drastic changes in the electrostatic field due to the movement of charges.

The turnover rate of H/ACA RNP (k_{cat}) has been determined[53] as 0.004 s^{-1} at 27 °C. This is two-order of magnitude smaller compared to standalone Ψ-synthases like TruB, RluA, TruA[65] or Pus10[111]. This can be attributed to multiple factors, like the temperature difference of the measurements, the variations in the catalytic mechanisms between Cbf5 and stand-alone enzymes, and the different modes of substrate binding and product release. Duan et al. estimated a turnover rate of 0.32 min^{-1} for the initial steady-state phase of the reaction,[51] which approximately conforms to a 18 kcal/mol rate limiting barrier height. Thus, kinetic experiments

generally propose a slow turnover rate for this multi-turnover[112] enzyme. We have to be careful with direct comparison of the kinetic parameters with our reaction barriers as we do not estimate barriers for substrate loading and product release, which could be rate-limiting for H/ACA RNPs. Quantitative comparison of calculated and experimental kinetic data is also hindered by the application of the semiempirical DFTB3 method necessitated by the complexity of the system and the catalytic process, and the approximate treatment of the tautomerization in the lack of structural data for the open conformation of the protein. Nevertheless, the overall exergonic reaction profile and the disfavored closed protein conformation pseudouridine product complex as obtained from MD simulations are in line with an expected irreversible reaction. Moreover, the proposed mechanism is in accord with experimental mutational data and with the ability of H/ACA RNPs to perform pseudouridylation of a wide range of substrates.

4. **Conclusions**

We had studied the structure and reaction mechanism of human box H/ACA pseudouridine synthase using classical and QM/MM molecular dynamics simulations with enhanced sampling methods. The simulations revealed a dissociative mechanism for the base dissociation-rebound step when the uracil detaches from the ribose by the cleavage of the C_1'-N_1 bond. This leads to a shallow minimum on the free energy surface that corresponds to an intermediate with detached base. The base rebounds to the sugar moiety with its C_5 carbon with a very small barrier. Analyses of structural data of both the classical and QM/MM MD simulations of the wild type complex indicate that the target uridine of the substrate RNA is distorted to a strained conformation. The strained conformation was not observed in any of the investigated inactive mutant enzymes (D125A, D125N, E206K) or in the product state complex containing

pseudouridine. The base detachment in the wild-type complex proceeds with a reasonable 14 kcal/mol barrier. In contrast, the investigated mutant enzymes (D125A, D125N, E206K) that do not distort the bound uridine exhibit high barriers for the base detachment step making the reaction unlikely, in line with their experimentally observed inactivity. These results indicate that the enzyme applies reactant state destabilization by using the free energy of substrate binding to weaken the cleaving C_1'-N_1 bond and to facilitate the first step of the reaction. Similar effect was formerly proposed to ease the cleavage of the glycosidic bond in uracil DNA glycosylase, an enzyme catalyzing an analogous reaction step.[92,93]

High barriers were estimated for the subsequent tautomerization when modelled within the active site. This is attributed to the spatially restricted binding pocket where the proton relay chain formed with water molecules and required for the proton transfer between the C_5 and N_1 atoms of the base cannot assemble easily. Based on this result we propose that a more open, water accessible active site is required to facilitate this step. This is in line with the observation that product formation is accompanied by thumb loop opening yielding a more spacious binding pocket and with our finding that the thumb loop and Gar1 exhibit significantly increased B-factors in the pseudouridine-bound complex compared to reactant state. We demonstrated with free energy calculations that tautomerization via proton relay occurs with reasonable barriers in water.

The structural analysis of the human box H/ACA pseudouridine synthase model revealed that the binding of the snoRNA to the RNP complex, and specifically into the catalytic pocket of dyskerin, is not sensitive to the actual nucleotide sequence of the snoRNA as only a few base specific interactions connect the bases to dyskerin, and even these interactions are formed with long, flexible side chains incorporating adaptability to the system. Moreover, since the substrate

RNA only associates with the guide snoRNA and dyskerin, and only the target uridine forms base specific interactions with the protein, – in presence of the appropriate guide RNA – ligand binding seems to be virtually independent of sequence-context. These results also explain why the prediction of pseudouridylation sites has proved to be such a challenge.[113] We also found that the pseudouridylation reaction does not require the participation of any neighboring nucleotides – neither from the guide nor from the substrate itself. Since neither substrate binding nor the uridine to pseudouridine transformation presents sequence requirements for the RNA components, we propose that engineered guide snoRNAs that can associate with the pseudouridine synthase complex inherently and endogenously present may be applied to facilitate the pseudouridylation of nearly any substrate. This suggests the opportunity of programmable RNA edition for a wide range of sequences.

Our structural analysis and reaction mechanism calculations show that the key element of the uridine to pseudouridine transformation is the correct positioning of the target base to achieve the conformational strain that drives the reaction, fortified through the tight binding between the guide and the substrate RNA which requires exact sequence match between them.[114,115] Though still much work is ahead, these findings indicate a hitherto unexploited therapeutical potential for RNA-guided editing by box H/ACA pseudouridine synthases.

Supporting Information

Computational details; definition of collective variables; details of umbrella sampling simulations, analysis of uridine structural distortion in wild type box H/ACA PSU ribonucleo-protein complex RNP); structural comparison of the human RNP with those of *Pyrococcus*

furiosus;and AlphaFold2 model; MD derived B-factors; Reaction barriers with varying QM regions; model calculations comparing DFTB3 and DFT results for the C-N bond cleavage, C-C bond formation and tautomerization; DFTB3 and DFT model calculations for the C-N bond cleavage in uridine and the proton transfer between ribose and aspartate

Acknowledgements

We thank Drs. Kálmán Tory and Gusztáv Schay for fruitful discussions. This work was supported by the Hungarian Scientific Research Fund (OTKA) through grants K111862, K116305 and FK124230. J. O. and D. K. M. thank the financial support of Project no. 2018-1.2.1-NKP-2018-00005 of the National Research, Development and Innovation Fund of Hungary. J.O. thanks the financial support of KU Leuven – Budapest University of Technology and Economics joint research funding (CELSA/19/017). The authors are grateful for the KIFÜ-NIIF Institute for granting computational time on the Hungarian HPC Infrastructure.

References

(1) Machnicka, M. A.; Milanowska, K.; Oglou, O. O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K. M.; Helm, M.; Bujnicki, J. M.; Grosjean, H. MODOMICS: A Database of RNA Modification Pathways - 2013 Update. *Nucleic Acids Res.* **2013**, *41* (D1), 262–267.

(2) Kellner, S.; Neumann, J.; Rosenkranz, D.; Lebedeva, S.; Ketting, R. F.; Zischler, H.; Schneider, D.; Helm, M. Profiling of RNA Modifications by Multiplexed Stable Isotope Labelling. *Chem. Commun.* **2014**, *50* (26), 3516–3518.

(3) Davis, F. F.; Allen, F. W. Ribonucleic Acids from Yeast Which Contain a Fifth Nucleotide. *J. Biol. Chem.* **1957**, *227* (2), 907–915.

(4) Ge, J.; Yu, Y. RNA Pseudouridylation : New Insights into an Old Modification. *Trends Biochem. Sci.* **2013**, *38* (4), 210–218.

(5) Grosjean, H.; Sprinzl, M.; Steinberg, S. Posttranscriptionally Modified Nucleosides in

Transfer RNA: Their Locations and Frequencies. *Biochimie* **1995**, *77*, 139–141.

(6) Sprinzl, M.; Horn, C.; Brown, M.; Ioudovitch, A.; Steinberg, S. Compilation of tRNA Sequences and Sequences of tRNA Genes. *Nucleic Acids Res.* **1998**, *26* (1), 148–153.

(7) Branlant, C.; Krol, A.; Machatt, M. A.; Pouyet, J.; Ebel, J.; Edwards, K.; Kössel, H. Primary and Secondary Structures of Escherichia Coli MRE 600 23S Ribosomal RNA. Comparison with Models of Secondary Structure for Maize Chloroplast 23S rRNA and for Large Portions of Mouse and Human 16S Mitochondrial rRNAs. *Nucleic Acids Res.* **1981**, *9* (17), 4303–4324.

(8) Wu, G.; Yu, A. T.; Kantartzis, A.; Yu, Y. Functions and Mechanisms of Spliceosomal Small Nuclear RNA Pseudouridylation. *WIREs RNA* **2011**, *2*, 571–581.

(9) Carlile, T. M.; Rojas-duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine Profiling Reveals Regulated mRNA Pseudouridylation in Yeast and Human Cells. *Nature* **2014**, *515* (7525), 143–146.

(10) Liang, X.-H.; Liu, Q.; Fournier, M. J. Loss of rRNA Modifications in the Decoding Center of the Ribosome Impairs Translation and Strongly Delays Pre-rRNA Processing. *RNA* **2009**, *15* (9), 1716–1728.

(11) Wu, G.; Adachi, H.; Ge, J.; Stephenson, D.; Query, C. C.; Yu, Y.-T. Pseudouridines in U2 snRNA Stimulate the ATPase Activity of Prp5 during Spliceosome Assembly. *EMBO J.* **2016**, *35*, 654–667.

(12) Wu, G.; Xiao, M.; Yang, C.; Yu, Y.-T. U2 snRNA Is Inducibly Pseudouridylated at Novel Sites by Pus7p and SnR81 RNP. *EMBO J.* **2011**, *30* (1), 79–89.

(13) Zhao, X.; Yu, Y.-T. Incorporation of 5-Fluorouracil into U2 snRNA Blocks Pseudouridylation and pre-mRNA Splicing in Vivo. *Nucleic Acid Res.* **2007**, *35* (2), 550–558.

(14) Yu, Y.; Meier, U. T. RNA-Guided Isomerization of Uridine to Pseudouridine — Pseudouridylation. *RNA Biol.* **2014**, *11* (12), 1483–1494.

(15) Kiss, T.; Fayet-Lebaron, E.; Jády, B. E. Box H/ACA Small Ribonucleoproteins. *Mol. Cell* **2010**, *37*, 597–606.

(16) Hamma, T.; Ferré-D ’amaré, A. R. The Box H/ACA Ribonucleoprotein Complex: Interplay of RNA and Protein Structures in Post-Transcriptional RNA Modification. *J. Biol. Chem.* **2010**, *285* (2), 805–809.

(17) Meier, U. T. The Many Facets of H/ACA Ribonucleoproteins. *Chromosoma* **2005**, *114*, 1–14.

(18) Shay, J. W.; Wright, W. E. Telomeres and Telomerase: Three Decades of Progress. *Nat. Rev. Genet.* **2019**, *20* (5), 299–309.

(19) Kirwan, M.; Dokal, I. Dyskeratosis Congenita, Stem Cells and Telomeres. *Biochim. Biophys. Acta* **2009**, *1792*, 371–379.

(20) Balogh, E.; Chandler, J. C.; Varga, M.; Tahoun, M.; Menyhárd, D. K.; Schay, G.; Goncalves, T.; Hamar, R.; Légrádi, R.; Szekeres, Á.; Gribouval, O.; Kleta, R.; Stanescu, H.; Bockenhauer, D.; Kerti, A.; Williams, H.; Kinsler, V.; Di, W. L.; Curtis, D.; Kolatsi-Joannou, M.; Hammid, H.; Szőcs, A.; Perczel, K.; Maka, E.; Toldi, G.; Sava, F.; Arrondel, C.; Kardos, M.; Fintha, A.; Hossain, A.; D'Arco, F.; Kaliakatsos, M.; Koeglmeier, J.; Mifsud, W.; Moosajee, M.; Faro, A.; Jávorszky, E.; Rudas, G.; Saied, M. H.; Marzouk, S.; Kelen, K.; Götze, J.; Reusz, G.; Tulassay, T.; Dragon, F.; Mollet, G.; Motameny, S.; Thiele, H.; Dorval, G.; Nürnberg, P.; Perczel, A.; Szabó, A. J.; Long, D. A.; Tomita, K.; Antignac, C.; Waters, A. M.; Tory, K. Pseudouridylation Defect Due to DKC1 and NOP10 Mutations Causes Nephrotic Syndrome with Cataracts, Hearing Impairment, and Enterocolitis. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (26), 15137–15147.

(21) Penzo, M.; Montanaro, L. Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function. *Biomolecules* **2018**, *8* (2), 38.

(22) Newby, M. I.; Greenbaum, N. L. A Conserved Pseudouridine Modification in Eukaryotic U2 snRNA Induces a Change in Branch-Site Architecture. *RNA* **2001**, *7*, 833–845.

(23) Davis, D. R. Stabilization of RNA Stacking by Pseudouridine. *Nucleic Acids Res.* **1995**, *23* (24), 5020–5026.

(24) Hudson, G. a; Bloomingdale, R. J.; Znosko, B. M. Thermodynamic Contribution and Nearest-Neighbor Parameters of Pseudouridine-Adenosine Base Pairs in Oligoribonucleotides. *RNA* **2013**, *19* (11), 1474–1482.

(25) Kierzek, E.; Malgowska, M.; Lisowiec, J.; Turner, D. H.; Gdaniec, Z.; Kierzek, R. The Contribution of Pseudouridine to Stabilities and Structure of RNAs. *Nucleic Acids Res.* **2014**, *42* (5), 3492–3501.

(26) Arnez, J. G.; Steitz, T. A. Crystal Structure of Unmodified tRNAGln Complexed with Glutaminyl-tRNA Synthetase and ATP Suggests a Possible Role for Pseudo-Uridines in Stabilization of RNA Structure. *Biochemistry* **1994**, *33* (24), 7560–7567.

(27) Newby, M. I.; Greenbaum, N. L. Investigation of Overhauser Effects between Pseudouridine and Water Protons in RNA Helices. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12697–12702.

(28) Hall, K. B.; Mclaughlin, L. W. Properties of Pseudouridine N1 Imino Protons Located in the Major Groove of an A-Form RNA Duplex. *Nucleic Acids Res.* **1992**, *20* (8), 1883–1889.

(29) Meroueh, M.; Grohar, P. J.; Qiu, J.; SantaLucia Jr, J.; Scaringe, S. A.; Chow, C. S. Unique Structural and Stabilizing Roles for the Individual Pseudouridine Residues in the 1920 Region of Escherichia Coli 23S rRNA. *Nucleic Acids Res.* **2000**, *28* (10), 2075–2083.

(30) Charette, M.; Gray, M. W. Pseudouridine in RNA : What , Where , How , and Why. *IUBMB Life* **2000**, *49*, 341–351.

(31) Karikó, K.; Buckstein, M.; Ni, H.; Weissman, D. Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* **2005**, *23* (2), 165–175.

(32) Karikó, K.; Muramatsu, H.; Welsh, F. A.; Ludwig, J.; Kato, H.; Akira, S.; Weissman, D. Incorporation of Pseudouridine into mRNA Yields Superior Nonimmunogenic Vector with Increased Translational Capacity and Biological Stability. *Mol. Ther.* **2008**, *16* (11), 1833–1840.

(33) Anderson, B. R.; Muramatsu, H.; Nallagatla, S. R.; Bevilacqua, P. C.; Sansing, L. H.; Weissman, D.; Karikó, K. Incorporation of Pseudouridine into mRNA Enhances Translation by Diminishing PKR Activation. *Nucleic Acids Res.* **2010**, *38* (17), 5884–5892.

(34) Eyler, D. E.; Franco, M. K.; Batool, Z.; Wu, M. Z.; Dubuke, M. L.; Dobosz-Bartoszek, M.; Jones, J. D.; Polikanov, Y. S.; Roy, B.; Koutmou, K. S. Pseudouridinylation of mRNA Coding Sequences Alters Translation. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (46), 23068–23074.

(35) Adachi, H.; Yu, Y.-T. Pseudouridine-Mediated Stop Codon Readthrough in S. Cerevisiae Is Sequence Context–Independent. *RNA* **2020**, *26* (9), 1247–1256.

(36) Huang, C.; Wu, G.; Yu, Y. T. Inducing Nonsense Suppression by Targeted Pseudouridylation. *Nat. Protoc.* **2012**, *7* (4), 789–800.

(37) Karijolich, J.; Yu, Y. Converting Nonsense Codons into Sense Codons by Targeted Pseudouridylation. *Nature* **2011**, *474* (7351), 395–398.

(38) Benhabiles, H.; Jia, J.; Lejeune, F. Strategies to Correct Nonsense Mutations. In *Nonsense Mutation Correction in Human Diseases*; Elsevier, 2016; pp 107–165.

(39) Kuttan, A.; Bass, B. L. Mechanistic Insights into Editing-Site Specificity of ADARs. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (48), E3295.

(40) Riedmann, E. M.; Schopoff, S.; Hartner, J. C.; Jantsch, M. F. Specificity of ADAR-Mediated RNA Editing in Newly Identified Targets. *RNA* **2008**, *14* (6), 1110–1118.

(41) Katrekar, D.; Chen, G.; Meluzzi, D.; Ganesh, A.; Worlikar, A.; Shih, Y. R.; Varghese, S.; Mali, P. In Vivo RNA Editing of Point Mutations via RNA-Guided Adenosine Deaminases. *Nat. Methods* **2019**, *16* (3), 239–242.

(42) Cox, D. B. T.; Gootenberg, J. S.; Abudayyeh, O. O.; Franklin, B.; Kellner, M. J.; Joung, J.; Zhang, F. RNA Editing with CRISPR-Cas13. *Science* **2017**, *358* (6366), 1019–1027.

(43) Abudayyeh, O. O.; Gootenberg, J. S.; Franklin, B.; Koob, J.; Kellner, M. J.; Ladha, A.; Joung, J.; Kirchgatterer, P.; Cox, D. B. T.; Zhang, F. A Cytosine Deaminase for Programmable

Single-Base RNA Editing. *Science* **2019**, *365* (6451), 382–386.

(44) Rupaimoole, R.; Slack, F. J. MicroRNA Therapeutics: Towards a New Era for the Management of Cancer and Other Diseases. *Nat. Rev. Drug Discov.* **2017**, *16* (3), 203–221.

(45) Rintala-Dempsey, A. C.; Kothe, U. Eukaryotic Stand-Alone Pseudouridine Synthases–RNA Modifying Enzymes and Emerging Regulators of Gene Expression? *RNA Biol.* **2017**, *14* (9), 1185–1196.

(46) Ferré-D'Amaré, A. R. RNA-Modifying Enzymes. *Curr. Opin. Struct. Biol.* **2003**, *13*, 49–55.

(47) Boschi-Muller, S.; Motorin, Y. Chemistry Enters Nucleic Acids Biology: Enzymatic Mechanisms of RNA Modification. *Biochem.* **2013**, *78* (13), 1392–1404.

(48) Baker, D. L.; Youssef, O. A.; Chastkofsky, M. I. R.; Dy, D. A.; Terns, R. M.; Terns, M. P. RNA-Guided RNA Modification: Functional Organization of the Archaeal H/ACA RNP. *Genes Dev.* **2005**, *19*, 1238–1248.

(49) Charpentier, B.; Muller, S. B.; Branlant, C. Reconstitution of Archaeal H/ACA Small Ribonucleoprotein Complexes Active in Pseudouridylation. *Nucleic Acids Res.* **2005**, *33* (10), 3133–3144.

(50) Liang, B.; Li, H. Structures of Ribonucleoprotein Particle Modification Enzymes. *Q Rev Biophys.* **2011**, *44* (1), 95–122.

(51) Duan, J.; Li, L.; Lu, J.; Wang, W.; Ye, K. Structural Mechanism of Substrate RNA Recruitment in H/ACA RNA-Guided Pseudouridine Synthase. *Mol. Cell* **2009**, *34* (4), 427–439.

(52) Wang, P.; Yang, L.; Gao, Y. Q.; Zhao, X. S. Accurate Placement of Substrate RNA by Gar1 in H/ACA RNA-Guided Pseudouridylation. *Nucleic Acids Res.* **2015**, *43* (10), 7207–7216.

(53) Yang, X.; Duan, J.; Li, S.; Wang, P.; Ma, S.; Ye, K.; Zhao, X. S. Kinetic and Thermodynamic Characterization of the Reaction Pathway of Box H/ACA RNA-Guided Pseudouridine Formation. *Nucleic Acids Res.* **2012**, *40* (21), 10925–10936.

(54) Liang, B.; Zhou, J.; Kahen, E.; Terns, R. M.; Terns, M. P.; Li, H. Structure of a Functional Ribonucleoprotein Pseudouridine Synthase Bound to a Substrate RNA. *Nat Struct Mol Biol* . **2009**, *16* (7), 740–746.

(55) Hamma, T.; Ferré-D'Amaré, A. R. Pseudouridine Synthases. *Chem. Biol.* **2006**, *13*, 1125–1135.

(56) Mccleverty, C. J.; Hornsby, M.; Spraggon, G.; Kreusch, A. Crystal Structure of Human Pus10, A Novel Pseudouridine Synthase. *J. Mol. Biol.* **2007**, *373*, 1243–1254.

(57) Spenkuch, F.; Motorin, Y.; Helm, M. Pseudouridine : Still Mysterious , but Never a Fake (Uridine)! *RNA Biol.* **2014**, *11* (12), 1540–1554.

(58) Gu, X.; Liu, Y.; Santi, D. V. The Mechanism of Pseudouridine Synthase I as Deduced from Its Interaction with 5-Fluorouracil-tRNA. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (25), 14270–14275.

(59) Huang, L.; Pookanjanatavip, M.; Gu, X.; Santi, D. V. A Conserved Aspartate of tRNA Pseudouridine Synthase Is Essential for Activity and a Probable Nucleophilic Catalyst. *Biochemistry* **1998**, *2960* (97), 344–351.

(60) Miracco, E. J.; Mueller, E. G. The Products of 5-Fluorouridine by the Action of the Pseudouridine Synthase TruB Disfavor One Mechanism and Suggest Another. *J. Am. Chem. Soc.* **2011**, *133* (31), 11826–11829.

(61) Veerareddygari, G. R.; Singh, S. K.; Mueller, E. G.; Reddy Veerareddygari, G.; Singh, S. K.; Mueller, E. G. The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J. Am. Chem. Soc.* **2016**, *138* (25), 7852–7855.

(62) Hoang, C.; Chen, J.; Vizthum, C. A.; Kandel, J. M.; Hamilton, C. S.; Mueller, E. G.; Ferré-D'Amaré, A. R. Crystal Structure of Pseudouridine Synthase RluA: Indirect Sequence Readout through Protein-Induced RNA Structure. *Mol. Cell* **2006**, *24* (4), 535–545.

(63) Hoang, C.; Ferré-D'Amaré, A. R. Cocrystal Structure of a tRNA Ψ55 Pseudouridine Synthase: Nucleotide Flipping by an RNA-Modifying Enzyme. *Cell* **2001**, *107* (7), 929–939.

(64) Kiss, D. J.; Oláh, J.; Tóth, G.; Menyhárd, D. K.; Ferenczy, G. G. Quantum Chemical Calculations Support Pseudouridine Synthase Reaction through a Glycal Intermediate and Provide Details of the Mechanism. *Theor. Chem. Acc.* **2018**, *137* (12), 162.

(65) Wright, J. R.; Keffer-Wilkes, L. C.; Dobing, S. R.; Kothe, U. Pre-Steady-State Kinetic Analysis of the Three Escherichia Coli Pseudouridine Synthases TruB, TruA, and RluA Reveals Uniformly Slow Catalysis. *RNA* **2011**, *17*, 2074–2084.

(66) Safra, M.; Nir, R.; Farouq, D.; Slutzkin, I. V.; Schwartz, S. TRUB1 Is the Predominant Pseudouridine Synthase Acting on Mammalian mRNA via a Predictable and Conserved Code. *Genome Res.* **2017**, *27* (3), 393–406.

(67) Wrzesinski, J.; Nurse, K.; Bakin, A.; Lane, B. G.; Ofengand, J. A Dual-Specificity Pseudouridine Synthase: An Escherichia Coli Synthase Purified and Cloned on the Basis of Its Specificity for Psi 746 in 23S RNA Is Also Specific for Psi 32 in tRNA(Phe). *RNA* **1995**, *1* (4), 437–448.

(68) Hur, S.; Stroud, R. M. How U38, 39, and 40 of Many tRNAs Become the Targets for Pseudouridylation by TruA. *Mol. Cell* **2007**, *26* (2), 189–203.

(69) Biswas, P. K.; Chakraborty, S. Targeted DNA Oxidation and Trajectory of Radical DNA Using DFT Based QM/MM Dynamics. *Nucleic Acids Res.* **2019**, *47* (6), 2757–2765.

(70) Kästner, J.; Sherwood, P. The Ribosome Catalyzes Peptide Bond Formation by Providing

High Ionic Strength. *Mol. Phys.* **2010**, *108* (3–4), 293–306.

(71) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344–4350.

(72) Kanaan, N.; Marti, S.; Moliner, V.; Kohen, A. QM/MM Study of Thymidylate Synthase: Enzymatic Motions and the Temperature Dependence of the Rate Limiting Step. *J. Phys. Chem. A* **2009**, *113* (10), 2176–2182.

(73) Zhou, J.; Liang, B.; Li, H. Functional and Structural Impact of Target Uridine Substitutions on the H/ACA Ribonucleoprotein Particle Pseudouridine Synthase. *Biochemistry* **2010**, *49* (29), 6276–6281.

(74) Koo, B. K.; Park, C. J.; Fernandez, C. F.; Chim, N.; Ding, Y.; Chanfreau, G.; Feigon, J. Structure of H/ACA RNP Protein Nhp2p Reveals Cis/Trans Isomerization of a Conserved Proline at the RNA and Nop10 Binding Interface. *J. Mol. Biol.* **2011**, *411* (5), 927–942.

(75) Li, S.; Duan, J.; Li, D.; Yang, B.; Dong, M.; Ye, K. Reconstitution and Structural Analysis of the Yeast Box H/ACA RNA-Guided Pseudouridine Synthase. *Genes Dev.* **2011**, *25* (22), 2409–2421.

(76) Schrödinger Release 2020-4: Prime. Schrödinger LLC, New York, NY, 2020.

(77) Pronk, S.; Rd, S.; Ll, P.; Schulz, R.; Larsson, P.; Bjelkmar, P. R.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P. R.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854.

(78) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study Using Proline and Hydroxyproline Sidechain Dynamics. *Proteins* **2014**, *82*, 195–215.

(79) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405–4412.

(80) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871.

(81) Heiss, N. S.; Girod, A.; Salowsky, R.; Wiemann, S.; Pepperkok, R.; Poustka, A. Dyskerin Localizes to the Nucleolus and Its Mislocalization Is Unlikely to Play a Role in the Pathogenesis of Dyskeratosis Congenita. *Hum. Mol. Genet.* **1999**, *8* (13), 2515–2524.

(82) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118*

(8), 4177–4338.

(83) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537–W541.

(84) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7* (4), 931–948.

(85) Ewald, P. P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369* (3), 253–287.

(86) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma. *J. Comput. Phys.* **1974**, *14* (2), 148–158.

(87) Case, D. A. A.; Ben-Shalom, I. Y.; Brozell, S. R. R.; Cerutti, D. S. S.; T.E. Cheatham, I.; Cruzeiro, V. W. D.; Darden, T. A. A.; Duke, R. E. E.; Ghoreishi, D.; Gilson, M. K. K.; Gohlke, H.; Goetz, A. W. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S. S.; LeGra, S.; Kollman, P. A. A.; Cheatham, T. E.; Cruzeiro, I. V. W. D.; Darden, T. A. A.; Duke, R. E. E.; Ghoreishi, D.; Gilson, M. K. K.; Gohlke, H.; Goetz, A. W. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzmann, T.; Lee, T. S. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. A. AMBER 2018, University of California, San Francisco. 2018.

(88) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

(89) Majumder, M.; Bosmeny, M. S.; Gupta, R. Structure-Function Relationships of Archaeal Cbf5 during in Vivo RNA-Guided Pseudouridylation. *RNA* **2016**, *22* (10), 1604–1619.

(90) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710.

(91) Wu, H.; Feigon, J. H/ACA Small Nucleolar RNA Pseudouridylation Pockets Bind Substrate RNA to Form Three-Way Junctions That Position the Target U for Modification. *Proc. Natl. Acad. Sci.* **2007**, *104* (16), 6655–6660.

(92) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. Uracil-DNA Glycosylase-DNA Substrate and Product Structures:

Conformational Strain Promotes Catalytic Efficiency by Coupled Stereoelectronic Effects. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (10), 5083–5088.

(93) Das, R.; Vázquez-Montelongo, E. A.; Andrés Cisneros, G.; Wu, J. I. Ground State Destabilization in Uracil DNA Glycosylase: Let's Not Forget "Tautomeric Strain" in Substrates. *J. Am. Chem. Soc.* **2019**, *141* (35), 13739–13743.

(94) Pan, H.; Agarwalla, S.; Moustakas, D. T.; Finer-Moore, J.; Stroud, R. M. Structure of tRNA Pseudouridine Synthase TruB and Its RNA Complex: RNA Recognition through a Combination of Rigid Docking and Induced Fit. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (22), 12648–12653.

(95) Phannachet, K.; Huang, R. H. Conformational Change of Pseudouridine 55 Synthase upon Its Association with RNA Substrate. *Nucleic Acids Res.* **2004**, *32* (4), 1422–1429.

(96) Dinner, A. R.; Blackburn, G. M.; Karplus, M. Uracil-DNA Glycosylase Acts by Substrate Autocatalysis. *Nature* **2001**, *413* (October), 752–755.

(97) Drohat, A. C.; Maiti, A. Mechanisms for Enzymatic Cleavage of the N-Glycosidic Bond in DNA. *Org. Biomol. Chem.* **2014**, *12* (42), 8367–8378.

(98) Werner, R. M.; Stivers, J. T. Kinetic Isotope Effect Studies of the Reaction Catalyzed by Uracil DNA Glycosylase: Evidence for an Oxocarbenium Ion - Uracil Anion Intermediate. *Biochemistry* **2000**, *39* (46), 14054–14064.

(99) Gao, J. Catalysis by Enzyme Conformational Change as Illustrated by Orotidine 5'-Monophosphate Decarboxylase. *Curr. Opin. Struct. Biol.* **2003**, *13* (2), 184–192.

(100) Ramamurthy, V.; Swann, S. L.; Paulson, J. L.; Spedaliere, C. J.; Mueller, E. G. Critical Aspartic Acid Residues in Pseudouridine Synthases. *J. Biol. Chem.* **2002**, *274* (32), 22225–22230.

(101) Hoang, C.; Hamilton, C. S.; G.Mueller, E.; Ferré-D'Amaré, A. R. Precursor Complex Structure of Pseudouridine Synthase TruB Suggests Coupling of Active Site Perturbations to an RNA-Sequestering Peripheral Protein Domain. *Protein Sci.* **2005**, *14* (8), 2201–2206.

(102) Spedaliere, C. J.; Mueller, E. G. Not All Pseudouridine Synthases Are Potently Inhibited by RNA Containing 5-Fluorouridine. *RNA* **2004**, *10* (2), 192–199.

(103) Reddy Veerareddygari, G.; Singh, S. K.; Mueller, E. G. The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J. Am. Chem. Soc.* **2016**, *138*, 7852–7855.

(104) Zhou, J.; Liang, B.; Li, H. Structural and Functional Evidence of High Specificity of Cbf5 for ACA Trinucleotide. *RNA* **2011**, *17* (2), 244–250.

(105) Zhou, J.; Lv, C.; Liang, B.; Chen, M.; Yang, W.; Li, H. Glycosidic Bond Conformation Preference Plays a Pivotal Role in Catalysis of RNA Pseudouridylation: A Combined

Simulation and Structural Study. *J. Mol. Biol.* **2010**, *401*, 690–695.

(106) Mohammed, O. F.; Pines, D.; Dreyer, J.; Pines, E.; Nibbering, E. T. J. Chemistry: Sequential Proton Transfer through Water Bridges in Acid-Base Reactions. *Science* **2005**, *310* (5745), 83–86.

(107) Ikeda, T.; Saito, K.; Hasegawa, R.; Ishikita, H. The Existence of an Isolated Hydronium Ion in the Interior of Proteins. *Angew. Chem. Int. Ed.* **2017**, *56* (31), 9151–9154.

(108) Intharathep, P.; Tongraar, A.; Sagarik, K. Ab Initio QM/MM Dynamics of H3O+ in Water. *J. Comput. Chem.* **2006**, *27* (14), 1723–1732.

(109) Stirling, A.; Pápai, I. H2CO3 Forms via HCO3- in Water. *J. Phys. Chem. B* **2010**, *114* (50), 16854–16859.

(110) Wang, B.; Cao, Z. How Water Molecules Modulate the Hydration of CO2 in Water Solution: Insight from the Cluster-Continuum Model Calculations. *J. Comput. Chem.* **2013**, *34* (5), 372–378.

(111) Kamalampeta, R.; Keffer-wilkes, L. C.; Kothe, U. tRNA Binding , Positioning , and Modification by the Pseudouridine Synthase Pus10. *J. Mol. Biol.* **2013**, *425* (20), 3863–3874.

(112) Bobo, C.; Saliou, J.-M.; Van Dorsselaer, A.; Atmanene, C.; Tillault, A.-S.; Charpentier, B.; Cianférani, S.; Manival, X.; Branlant, C. Combining Native MS Approaches to Decipher Archaeal Box H/ACA Ribonucleoprotein Particle Structure and Activity. *Proteomics* **2015**, *15* (16), 2851–2861.

(113) Dou, L.; Li, X.; Ding, H.; Xu, L.; Xiang, H. Is There Any Sequence Feature in the RNA Pseudouridine Modification Prediction Problem? *Mol. Ther. - Nucleic Acids* **2020**, *19*, 293–303.

(114) Kelly, E. K.; Czekay, D. P.; Kothe, U. Base-Pairing Interactions between Substrate RNA and H/ACA Guide RNA Modulate the Kinetics of Pseudouridylation, but Not the Affinity of Substrate Binding by H/ACA Small Nucleolar Ribonucleoproteins. *RNA* **2019**, *25* (10), 1393–1404.

(115) Zoysa, M. D. De; Wu, G.; Katz, R.; Yu, Y.-T. Guide-Substrate Base-Pairing Requirement for Box H/ACA RNA-Guided RNA Pseudouridylation. *RNA* **2018**, *24* (8), 1106–1117.

For Table of Contents Only