Structural basis for hydrolytic splicing of a circularly permuted group II intron

Xiaobin Ling[1,2,5,§,*]; Yuqi Yao[1,§]; Dmitrij Golovenko[2,§]; Xiang Zhang[3,§]; Kangkang Song[4]; Chen Xu[4]; Jianhua Gan[1]; Zhenguo Chen[3,*]; Wenwen Fang[2,*]; Jinbiao Ma[1,*].

[1]*State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, School of Life Sciences, Fudan University, Shanghai 200438, China.*

[2]*RNA Therapeutics Institute, UMass Chan Medical School, Worcester 01655, USA.*

[3]*MOE/NHC/CAMS Key Laboratory of Medical Molecular Virology, Shanghai Institute of Infectious Disease and Biosecurity, The Fifth People's Hospital of Shanghai, Institutes of Biomedical Sciences, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China.*

[4]*Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA.*

[5]*Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA*

§*These authors contributed equally.*

**Respondence should be addressed to X.L. (lingx1@mskcc.org) Z.C.(zhenguochen@fudan.edu.cn), W.F.(wenewn.fang@umassmed.edu), J.M. (majb@fudan.edu.cn).*

Abstract

Group II introns are self-splicing ribozymes that are considered the ancestors of the eukaryotic spliceosome. Unlike canonical group II introns that self-splice to generate linear exons and lariats, circularly permuted (CP) group II introns identified in various bacterial phyla perform back-splicing, resulting in the production of circular RNAs and branched products via branching pathway. Furthermore, CP introns may switch to a hydrolysis pathway resulting in distinct products to differentially regulate gene expression. In this study, we present biochemical data and high-resolution cryogenic electron microscopy (cryo-EM) structures of a CP group II intron from *Comamonas testosteroni* KF-1 (*Cte* 1), allowing mechanistic dissection of the branching pathway switch to the hydrolysis pathway and enabling reconstruction of both both steps of the hydrolysis pathway. The structures reveal that CP undergoes the hydrolysis pathway upon mutations of the branch point or splice sites (SS) due to rearrangements in the active site. Here, the branching nucleotide in domain D6 is retracted from its catalytically competent conformation, giving way to the nucleophilic water molecule to attack the 5′ splice site. Furthermore, we visualized the intermediates of the second splicing step, which reveal the movement of domain D6 out of the way for the 3′-splicing to occur, closely resembling the second-step structures of the branching pathway. Finally, our structures provide direct evidence for domains D1–D3 acting as a scaffold in group II introns.Together, these findings visualize the complete hydrolysis pathway and offer a new strategy to engineer CP introns for circular RNA production, with potential applications in both basic research and therapeutic development.

Key words: CP group II intron; branching pathway; hydrolysis pathway; back-splicing; circular RNA.

Introduction

Group II introns are mobile ribozymes that can self-splice from RNA precursors and reverse splice into DNA, thereby propagating themselves in bacterial and organellar genomes[1,2]. They are considered the ancestors of spliceosome introns and retrotransposons in eukaryotes, thus providing a model for understanding the mechanism and regulation of these related processes[3]. A typical splicing reaction of group II introns and spliceosomes involves two sequential transesterification reactions. The initial step involves a nucleophilic attack by the 2′-OH group of an internal adenosine (the "branch point") at the 5′ splice junction, resulting in the formation of a 2′-5′ phosphodiester branch within the intron. In the second step, the 3′-OH of the cleaved 5′ exon attacks the 3′ splice site, thus joining the two exons with a 3′-5′ phosphodiester bond and releasing the intron as a lariat. For most group II introns, the first step is rate-limiting[4,5].

An important alternative to the "branching" pathway of splicing is the hydrolytic splicing, where a water molecule acts as a nucleophile instead of the 2′-OH of the branch point adenosine, resulting in lariat-free products [6,7]. Hydrolytic splicing occur in vivo [7,8], and was characterized *in vitro* using constructs with mutations at or around the branch point[9,10]. The branching and hydrolytic pathways were proposed to underlie the regulation of gene expression. First, because hydrolytic splicing is irreversible, its linear intron product cannot reverse-splice into DNA for further genomic propagation. Second, the linear intron may allow translation of intron-encoded proteins (IEPs), facilitating subsequent splicing, reverse-splicing[5], and/or other processes.

Group II introns are categorized into three main classes—IIA, IIB, and IIC—based on the sequence of their IEPs and RNA secondary structural features. Group IIA and IIB introns readily forms branched lariats, whereas group IIC introns predominantly splice via the hydrolysis pathway[2]. Despite their differences, they all share a conserved arrangement of six structural domains (D1–6). The largest domain D1 enables exon binding sites (EBS) recognition by base pairing with intron binding sites (IBS) on the exon. D2 and D3 provide additional structural stability through interactions with D5 and D6 and notably, a conserved intervening junction (J2/3) that is part of the catalytic core. D4 often encodes an IEP that facilitates self-splicing and reverse-splicing. D5 docks into D1 and contains the most conserved nucleotides that form the active sites. D6 provides the branch point adenosine for the first transesterification step[2,6].

Substantial progress in the structural understanding of group II intron splicing and reverse-splicing revealed a conserved catalytic center among different classes of group II introns, which is used in both steps of splicing[11-20]. The active site for the pre-first step of splicing is almost identical between the branching and hydrolysis pathway[11,19], raising the question of how the choice between the two pathways is made. Several artificial and natural mutations have been reported to convert the branching pathway to the hydrolysis pathway[7-9], but a structural basis for this conversion is lacking. Interestingly, when the D1–5 of the hydrolytic group IIC intron from *O. i* is fused to D6 from *A.v.* group II intron, the branching pathway is activated, suggesting that D6 provides the critical structural features determining the choice of the pathway[21]. While recent cryo-EM studies have provided many details of splicing through the branching pathway, structural analysis of the hydrolysis pathway is limited to early crystallographic studies, in which D6 conformations were not resolved [11,12].

A family of naturally occurring, circularly permuted (CP) group II intron was recently identified in diverse bacterial genomes. These CP group II introns contain rearranged structural domains and produce circular RNAs (circRNAs) via back-splicing[22]. We have reported high-resolution cryo-EM structures of the 3′ truncated CP group II intron from *Comamonas testosteroni* KF-1 (*Cte* 1), showing that splicing of this intron shares the conserved catalytic mechanism with group II introns through the branching pathway, facilitated by an auxiliary sequence (AUX). Here, we determined structures of *Cte* 1 mutants, which utilize the hydrolysis pathway, capturing the steps of this pathway. Our work reveals how the intron employs unconventional D6 docking into D1 to route the splicing via the hydrolysis pathway. D1–3 domains rearrange to accomplish the second step of the hydrolysis pathway. Finally, the two hydrolysis products reveal the structural rearrangements leading to the circular RNA produced.

Results

Conversion from the branching to hydrolysis pathway in *Cte* 1 mutants

To compare the CP introns undergoing the branching and hydrolytic pathways, we first determined the full-length wild-type structure of group II intron from *Cte 1* undergoing the branching (Fig. 1a-b). Cryo-EM classification revealed the conformation representing the steps of pre-splicing (Pre, 2.75 Å; Fig. 1c and Extended

133 Data Fig. 2). The overall conformation are similar to that in our previous cryo-EM
134 structures of *Cte* 1, which lacked the D4 domain[23], which is thought to interact with
135 the ORF encoding the IEP and to facilitate the interaction between IEP and the
136 ribozyme[17-19]. Stabilized by D4, the neighboring subdomains P1 and P2 dock at D2
137 and D3 via long range interactions, as predicted previously[22], stabilizing the overall
138 conformation of the Pre state (Fig. 1c and Extended Data 6). Furthermore, G809–
139 C813 of D4 base-pair with C50–C54 of D5 (Extended Data Fig. 6a). Our structures
140 therefore reveal that D4 contributes to the stabilization of CP group II intron in the
141 absence of IEP.
142
143 To address the mechanism of the hydrolysis pathway, we identified *Cte* 1 mutants that
144 partially or completely employ the hydrolysis pathway for the first step of splicing.
145 The the branchpoint mutation A127G and the near-5′ splice mutant G221U switch the
146 enzyme to the hydrolysis pathway without producing notable branching intermediates,
147 whereas the 5′-splice-site (5′-SS) mutant G220C yields a mixture of the hydrolysis
148 and branching products (Fig. 1b and Extended Data Fig. 1a–c).
149
150 We next determined cryo-EM structures of these constructs at ~2.6-2.8 Å resolution,
151 sufficient to visualize the structural details underlying active-site differences (Fig. 1d-
152 f and Extended data Fig. 3-5). To capture the early Pre-splicing states, we determined
153 structures of G220C and G221U mutants in the presence of $CaCl_2$ to inhibit the
154 splicing activity(Fig. 1e-f and Extended data Fig. 1d and 4-5). The A127G mutant
155 also captured the first step of hydrolysis in the Pre state even though in the presence
156 of $MgCl_2$ (Fig. 1b and Extended data Fig. 3).
157
158 **Active-center rearrangements after D6 conformational change leading to the**
159 **hydrolysis pathway**
160
161 For all structural superposition we used D5 domain for aligning structures. Notably,
162 there is a significant shift in the D6 conformation (Extended data Fig. 7b). The
163 A127G result in a displacement of 7.8 Å, while the G221U mutation causes a
164 displacement of approximately 3.8 Å. The displacement difference between A127G
165 and G221U is 6.5 Å.
166
167 In all three mutants, the active sites adopt different conformations, but they have two
168 common features. First, the G220 loop retains its sharply kinked conformation with
169 phosphate 220 at the tip coordinating either one or two cations, resembling the
170 conformation of the pre-attack position in the WT structure (Fig. 2a). Second, the

branching nucleotides 127, guanosine in the A127G mutant or A127 in the other mutants, are found in different positions away from G220 (Fig. 2). Unlike in the WT Pre structure, where the 2′-OH 127 is placed 3.3 Å from phosphorus 220, the respective distances range from 4 Å in the G220C mutant to 7.5 Å in G221U and to ~12.5 Å in A127G. In these conformations, the G220 loop remains an attractive nucleophilic target, whereas the nucleophilic 2′-OH group is detached, providing space for a smaller nucleophile – such as water – to attack the G220 phosphate group. Indeed, the separation of 2′-OH in WT Pre and our mutant structures correlates with the extent of the hydrolysis pathway, as the most closely positioned hydroxyl in WT undergoes the branching pathway, the slightly shifted G220C demonstrates both the branching and hydrolysis intermediates, while the remaining most separated mutants undergo the hydrolysis pathway (Fig. 1b and Fig. 2).

In all mutants, the rearrangement of the active site occurs due to the differences in the nucleotide identities next to the branch-point that disrupt the interactions present in the WT Pre structures. In the G220C mutant, C220 is unstacked from A127, consistent with the less favorable stacking of pyrimidines on purines than purines on purines[24]. This results in the ~2 Å shift of the A127 ribose, placing its hydroxyl group farther from the scissile phosphate of G220 than in the WT Pre structure (Fig. 2a). In the G221U mutant, U221 is also separated from A127, in contrast to WT Pre, where G221 participates in a triple-pairing with the D6 helix and stacks on A127. Here, A127 is shifted farther from its position in WT Pre, placing its ribose ~7 Å away from the G220 phosphate (Fig. 2a). In A127G, the mutated residue G127 acquires a new base-pairing capability, which brings it into the D6 helix to Watson-Crick pair with C100 (Fig. 2a). This base pair is further stabilized by being sandwiched between Watson-Crick pairs G126-C538 and G99-C128. Due to the large rearrangement of the branching position, the D6 helix in the A127G mutant is less well ordered than in other structures, indicating D6 mobility as described in Supplementary Discussion.

δ-δ′ interaction regulate the conversion from branching to hydrolysis pathway

Through density analysis, we captured the formation of a δ-δ′ interaction between G220 and G539 in both the G220C and G221U mutants. This interaction is crucial for correctly recognizing the 5′-SS during the branching pathway. During the conformational shift from branching to hydrolysis, however, notable differences were observed.

In the G221U mutant, $G220_A$ does not exhibit strong residual branching density associated with G539; instead, $G220_B$ dissociates from G539 and inserts into the minor groove of D6. In contrast, in the G220C mutant, two strong density signals are present. One, $C220_A$, base-pairs with G539 during branching, while the other, $C220_B$, points downward, moving away from G539 and creating sufficient space for the branch point A127 to adopt a new conformation. In the A127G mutant, G220 stabilizes its conformation by stacking with A127 and G126, thus distancing itself from G539 (Fig. 2b).

Overall, in all three mutants, G220 or C220 disrupts the δ-δ′ interaction between G220 and G539 that is formed during branching and instead establishes new structural stabilizing interactions. These changes facilitate the transition from the branching to the hydrolysis pathway.

Conformational changes of the triple helix and metal cluster

The triple helix provides a stable platform for organizing the hydrolytic active site. Upstream of the 5′-SS, stability is mediated by base pairing with the EBS; downstream, metal cluster coordination—particularly at K2—supports the conformation necessary for nucleophilic attack[12]. Using PDB 4FAQ as a reference, we defined K2 as the density between the phosphate groups of G224 and A695 (Extended data Fig. 8). In the wild type, no extra downstream density is observed. In contrast, all three hydrolysis-prone mutants exhibit additional density near the 5′-SS, though distance measurements exclude assignment to K^+ or Mg^{2+} (data not show).

The wild-type and 4FAQ structures are highly similar, with the main difference being a downward shift of K2 to G224, corresponding to G5 in 4FAQ. In the wild type, K2 stabilizes the 5′-SS terminal base through interaction with A695 rather than the 2-nt bulge. Apart from this, the architecture remains nearly identical (Extended data Fig. 8a). Comparison to 4FAQ highlights the role of domain D6. In the wild type, G221 engages D6 via π-π′ interactions, akin to U2 in 4FAQ, reducing the reliance on K2 for stabilization. The absence of D6 in 4FAQ shifts K2 upward to stabilize multiple downstream bases, introducing a conformational tilt.

We next analyzed triple helix and metal cluster rearrangements in the mutants. For A127G mutant, G221 adopts two conformations: an outward-flipped (branching-like) state ($G221_A$)and an inward-stacked state with A695 and U82 ($G221_B$). K2 density is weak or absent; stacking compensates for the loss of metal-mediated stabilization. For

G221U mutant, J2/3 distortion disrupts stacking among A697-G696, G83, and C289-G288/C377, destabilizing the branching configuration and favoring hydrolysis. For G220C mutant, structural features of both the wild type and 4FAQ are preserved, maintaining competence for both pathways (Extended data Fig. 8b-d).

Visualizing the second step of the hydrolysis pathway

Due to the unique inverted structure of the D5-D6 domain in the CP group II intron, hydrolytic pathway must generate two linear RNA products: D1-D2-D3 (D1-3) and P1-P2-D5-D6-Exon (P1-E) (Fig. 1a)[22].Our cryo-EM analysis of the A127G mutant incubated with $MgCl_2$ yielded a 3. 1Å map corresponding to the D1-3 structure (Fig. 1b and 3c, Extended Data Fig. 3). The conformation of these domains is very similar to those in the pre-reaction full-length intron, consistent with the idea that the large D1 and smaller D2-3 domains form a scaffold for the splicing reaction[2]. Accordingly, the second step of the hydrolysis pathway was proposed to also depend on the D1 scaffolding[2], likely required to organize the splice site for freeing the exon from the P1-D6 domains. Because the second products of the hydrolysis pathway (P1-E after the first step or P1-D6 after the second step) are much smaller than the D1-containig molecules, we did not identify them in the cryo-EM dataset. Our biochemical assays suggest that the conversion of P1-E to P1-D6 in the A127G mutant is fast and coincident with the formation of D1-3 (Fig. 1b), consistent with previous studies[2,12]. It is therefore likely that D1-3 in our dataset arises as the result of the complete hydrolysis splicing and subsequent intron disassembly, consistent with the requirement of D1-3 for both steps (Fig. 1b, 3c-d).

To visualize the fast second step of splicing, we combined artificially transcribed products of the first step of the hydrolysis reaction P1-E and D1-4 in the presence of inhibitory $CaCl_2$ to prevent the fast conversion to final products. In the hydrolysis intermediate (hyI) structure corresponding to the 2nd step of the hydrolysis pathway, the 3′–5′ phosphodiester bond has formed by U219 and C135, the circular exon was retained in the structure (Fig. 3e, 4a, Extended data Fig. 9). This conformation confirmed the process of P1-Exon truly could docking back into D1–4 structure.

Compared with the pre-branching product (BP^{pre}) conformation, where the formation of a covalent 3′-5′ phosphodiester bond (Fig. 3f), the catalytic active sites of BP^{pre} and hyI are generally similar (Fig. 4a). To investigate the roles of C538 and G539 in AUX during the transition from the 1st step to the 2nd step of the hydrolysis pathway, we compared the conformational changes between the Pre and hyI states for three

mutants (Fig. 4b). In the branching pathway, C538 undergoes a ~90° downward flip from the Pre to the BP^{pre} state, and the G-G interaction between G539 and G220 is replaced by a Watson-Crick base pair between G539 and C134. In the hydrolysis pathway, except for the A127G mutant, C538 in the G220C and G221U mutants undergoes a ~180° flip backward. In the A127G mutant, C538 only shows slight conformational movement. In the hydrolysis reaction, G539 does not interact with G220; instead, A127 assists C220 of G220C mutant, G127 helps G539 of A127G mutant, ultimately stabilizing the Pre structure. The G221U mutant do not provide any assisting forces for G539 or G220. Finally, G539 and C134 form a non-classical Watson-Crick base pair with three hydrogen bonds, compared to the weaker interaction with only two hydrogen bonds in other mutants.

During the transition from the 1st to the 2nd step of the hydrolysis pathway in the Group IIC intron from *O. i*, G288 of J2/3 undergoes a ~180° flip for the first time founded[12]. For *Cte* 1's J2/3, the G696 base pairs with G61, and A697 base pairs with A60, forming specific G-G and A-A interactions (Fig. 4c). However, A695 undergoes a similarly significant ~180° flip, accompanied by a conformational shift in U82 of the 2-nt bulge and a movement of G224, which were previously unmentioned.

The conformational changes from 2nd step to hydrolysis product

Since we only obtained the hyI conformation from our cryo-EM samples of P1-E and D1-4 under $CaCl_2$ conditions, we wondered if reintroducing P1-E and D1-4 into $MgCl_2$ conditions would activate the second step of the reaction and allow us to capture the conformation of the hydrolysis products. Analysis of the resulting circular RNAs revealed that the products from both the branching pathway and the direct second-step reaction of the hydrolysis pathway are identical, confirming that the correct reaction proceeds under $MgCl_2$ conditions (Fig. 3b, Extended data Fig. 11d-e). We subsequently analyzed cryo-EM sample of P1-E and D1-4 under $MgCl_2$ conditions, resulting in the identification of two hydrolysis product conformations, which we named hyP1 and hyP2 (Fig. 3b, g-h, Extended data Fig. 10). By comparing the conformations of hyI and hyP1, we can clearly observe a significant conformational shift in the AUX-EBS (Extended data Fig. 12a), and A695 flips back to a position similar to that in the Pre conformation. During this process, the conformational changes of U82 and G224 are minimal (Extended data Fig. 12b).

New hydrolysis product captured based on undock and dock model

However, we also observed a new D6 conformation in hyP2 that resembles the Pre state, positioned across D1 (Fig. 3h). Comparison between hyP1 and hyP2 structures reveals that the AUX-EBS region remains largely unchanged, with the exception of a substantial upward and downward flipping of D6 and minor conformational shifts in J2/3 (Extended Data Fig. 12c). We propose two possible hypotheses for the generation of this conformation. First, hyP2 may not represent the typical product of the second step of hydrolysis, as achieving this state from the hyI conformation would require an energetically unfavorable reversal of the D6 conformation. Instead, hyP2 may remain at the catalytic center, enabling continued cleavage reactions mediated by water-derived hydroxyl groups, thereby separating P1-D6 and D1-4, with the released P1-D6 and P1-E competing for binding to D1-4 to form the hyP2 conformation. Alternatively, if hyP2 represents an intermediate state, structural instability could cause P1-D6 to undock from the hyP1 conformation. This undocking would then allow competition between P1-D6 and P1-E for docking back into D1-4, leading to the formation of hyP2. (Fig. 3a, h, 5b).

The density of the water molecules instead of 2′-OH of branch point

Although previous studies have modeled the positions of water molecules[12,25], we hypothesized that the true catalytic water responsible for nucleophilic attack is located precisely at the original 2′-OH position of branch point A127. This site would most likely correspond to the catalytic water density. Beyond prior limitations such as the absence of the D6 domain or insufficient map quality, earlier studies have not identified a water molecule replacing the 2′-OH density at the branch site. In our G221U mutant structure, we observed for the first time that after the 2′-OH of A127 moves away from the active center, its base engages with a water molecule, anchoring it at the original 2′-OH position of branching. This water molecule is thus poised to provide the nucleophile for attack on the 5′-splice site (Fig. 5a).

Discussion

Plausible mechanism of CP group II intron to produce circular RNA in vivo

Based on the structural information we have gathered, we propose a plausible mechanism for the generation of circular RNA by CP group II introns *in vivo*. First, the CP group II intron is transcribed from the genome by RNA polymerase and then folds into a three-dimensional structure (the secondary structure diagram in the Extended data Fig.13 was generated using the RNArtist software based on the already

resolved tertiary structure) capable of performing back splicing functions. In the wild type, it transitions from the Pre conformation to the BI state through a change in the D6 conformation, completing two-step transesterification reactions that releases circular RNA. This process occurs via the back-splicing branch pathway (Fig 5b, indicated by the blue arrows in the Extended data Fig. 13).

However, in the presence of natural branch point A mutations or under stress conditions that generate branch point A, splice site mutations, etc., the intron can follow the hydrolysis pathway. In this case, the mutants regulate the transition from the branch pathway to either the incomplete hydrolysis pathway or the complete hydrolysis pathway through shifts in the D6 conformation, along with changes in the AUX and 5′ SS conformations (Fig. 2, 5b and Extended data Fig. 7). In the hydrolysis pathway, once the 1st step is completed, the P1-Exon may undocks from the D1-4 structure and then docks back into D1-4 to initiate the second step of the hydrolysis reaction, generating circular RNA and hydrolysis product1 (Fig 5b, indicated by the orange arrows in the Extended data Fig. 13). Although the undocking and docking of the P1-Exon after the first step of splicing is an energetically unfavorable process, it is a necessary event due to the requirement for circular RNA production by the CP group II intron or the spliceosome. The P1-Exon is inherently inverted, making this process inevitable. This undocking and docking of the P1-Exon during the generation of circular RNA via the hydrolysis pathway is a novel observation that adds to our understanding of the previously defined mechanism of linear exon splicing through the hydrolysis pathway, which had not been mentioned before. Since hydrolysis intermediates and hydrolysis products conformations without covalent bond formation, the P1-exon and P1-D6 of these products may be in a dynamic process of docking and undocking. As a result, there may also be a conformation of hydrolysis product2 in which D6 lies across D1. This conformation has never been observed before (Fig 5b, indicated by the orange arrows in the Extended data Fig. 13).

D6, AUX and 5′ SS required for hydrolysis pathway transition

The C538 of AUX in A127G mutant flips down to insert into a small groove formed by C100-G101-C102 and G126-G127 in D6, creating a triple interaction with G126 and G101 in D6. This results in the mutated G127 base pairing with C100, ultimately acting as a wedge that stabilizes the conformational changes (Extended data Fig. 7). Moreover, G220 in the 5′ splice site (5′ SS) interacts with G539 in AUX, forming the important δ-δ′ interaction in the wild type, while it is involved in D6, base pairing

with C102, which leads to a weak interaction between G221 and U103 (Extended data Fig. 7).

In the G221U structure, C538 of AUX does not insert into the small groove of D6 (Extended Data Fig. 7), that means the AUX conformation change is not necessary for branch pathway converse to complete hydrolysis pathway. The δ-δ' interaction between G220 and G539 is also weakened, leading to a noticeable shift 3.8 Å of the D6 conformation (Extended data Figs. 7). The overall catalytic center of G221U mutant is quite similar to the wild type, with the most significant change being the noticeable displacement of the D6 conformation. As the G221U mutant is also a complete hydrolysis mutant, a displacement of more than 3.8 Å in the D6 conformation is a necessary condition for the transition from the branching pathway to the hydrolysis pathway.

The 5′ SS mutant G220C splices via both branching and hydrolysis pathways, with comparable intermediate products from each pathway (Fig. 1b and Extended Data Fig. 1c). G539 instead interacts with A127, while G221 forms two separate triple interactions with C102-U124 and G101-C125 (Fig. 2, Extended Data Fig. 7). Although the conformation of the 5′ SS undergoes significant changes, the D6 conformation does not experience substantial displacement (Extended Data Fig. 7b). Thus, the conformational change of the G539 of AUX and 5′ SS is not necessary for the transition from the branching pathway to the hydrolysis pathway.

In summary, significant shifts in the D6 conformation are necessary for the transition from the branching pathway to the hydrolysis pathway. For different mutants, the C538 of AUX, as seen in the A127G, or the conformational changes in the 5′ SS, as observed in the G221U and G220C mutants, can regulate this transition process.

The possible of CP group II intron to produce circular RNA by trans splicing

Regarding the significance of this energetically unfavorable undock and dock process, we hypothesize that the D1-4 structure acts as a stable scaffold in the cellular environment. The undocked P1-Exon could potentially originate from other group II introns but can dock onto the non-self D1-4 scaffold to facilitate the generation of heterologous circular RNA. Thus, while this process may not be economical, it plays an important role in increasing the diversity of heterologous circular exons within species. (indicated by the gray arrows in the Extended data Fig. 13).

At a first glance, the reconstructed 2nd step of the hydrolytic splicing involving P1–Exon and D1–4 to produce RNA circles resembles trans-splicing[26,27], where exons from two separate molecules are spliced together. This type of splicing occurs naturally in spliceosome[28], group I intron[29], or group II intron[30] based splicing, and can be engineered for therapeutics[31]. However, our strategy is distinct for several reasons. First, a typical trans-splicing undergoes two steps as in conventional splicing, whereas our strategy only involves the second step of splicing, bypassing the first step. Second, in a typical trans-splicing the ligated exons come from two separate molecules, whereas in our strategy the exons reside in one molecule in an inverted manner to produce circles. Nevertheless, both typical trans-splicing and our finding highlight the modular nature of group II intron, for which we now provide a structural basis. In CP group II intron, D1–4 acts primarily as a stable "scaffold module", to which the "catalytic module" P1–D6 can dock back, bringing in the inverted exon and catalyzing circRNA formation. From an evolutionary standpoint, trans-splicing is thought to facilitate exon shuffling and gene evolution. Based on our study of the CP group II intron, we speculate that intron modules from different sources/origins may be able to form a functional ribozyme in trans[21], and this mix-and-match might facilitate the evolution of group II intron themselves, eventually leading to the modern spliceosome. It would be interesting to search for such orphan group II intron modules in diverse genomes. Indeed, disrupted group II introns have been reported[32-37], with extreme examples in which the intron is split into at least three pieces[33,38-40]. From an engineering perspective, our finding provides a streamlined strategy to produce circular RNA, with a constant scaffold module and variable catalytic-exon module.

Hydrolysis pathway implications

First, the splicing of group IIC introns[12], reverse splicing of HYER[25] (Extended data Fig. 11a-c), and back splicing of CP group II introns have all elucidated the hydrolysis pathway structure, indicating that hydrolysis pathways occur in various splicing processes. It has also been suggested that the hydrolysis pathway may be ancestral to the branch pathway[8]. Given that group II introns could also be ancestral to spliceosomes[41], different splicing forms have been reported to have an evolutionary relationship with spliceosome structure and mechanisms[42]. Although it is currently believed that the major spliceosome in humans does not utilize a hydrolysis pathway[43], the second step of the hydrolysis pathway in CP group II introns resembles the second step of trans-splicing, which has been extensively documented in eukaryotes and humans. We have reason to believe that there may be other spliceosomes in humans capable of following hydrolysis pathways to generate

473 circular RNA or linear RNA, which have yet to be discovered and identified. Our
474 current findings underscore the complexity and evolutionary significance of splicing
475 mechanisms, highlighting the need for further investigation into the functional
476 diversity of intron splicing pathways.
477
478 **Where do the back splicing product ultimately end up?**
479
480 However, many questions remain unanswered. For instance, what is the ultimate fate
481 of the generated stable product conformation? Does it, like other group II introns,
482 undergo a retro-homing process back to the genome[44]? If so, what mechanisms govern
483 this process? If these back-splicing products cannot be re-back into the genome, will
484 they remain in the cell indefinitely, or will they be degraded by siRNA or miRNA[45]?
485 These questions are currently unresolved. Regarding whether the mechanisms of other
486 group II introns or the spliceosome in generating circular RNA are of broader
487 relevance, we will need collective efforts to produce more functional and structural
488 data on circular RNA before making any conclusions. After all, circular RNA is still a
489 relatively new field with many aspects awaiting further research.
490
491 **Materials and Methods**
492
493 **RNA preparation.** The *Cte* group II intron sequence was derived from NCBI
494 (GenBank: NZ_AAUJ02000001.1). The wild type and mutated RNAs were prepared
495 as described previously[46]. In brief, the DNA templates in pUC-57 plasmid were
496 amplified with the regular forward primer and reverse primer (supplementary excel).
497 In vitro transcription of all RNAs in this article were carried out with 0.3 μM DNA
498 template, 5 mM NTPs and 1 U/μL RNase inhibitor in 1× transcription buffer
499 containing 50 mM Tris-HCl (pH 7.9), 0.01% TritonX-100, 20 mM $MgCl_2$, 2 mM
500 spermidine, 10 mM DTT, and incubated at 37 °C for 1 h[47]. The transcription products
501 were mixed with 2× denaturing gel loading buffer containing 95% formamide,
502 0.025% SDS, 10 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue,
503 and loaded on an 4% 19:1 acrylamide:bis, 8 M urea polyacrylamide gel. The gel was
504 run at 15 W for 3 h, then visualized briefly with a 254-nm UV lamp, held far from the
505 gel to minimize RNA damage[48]. Then the RNA was eluted from the gel overnight in
506 elution buffer containing 30 mM sodium acetate (pH 5.2) and 5 mM EDTA on an
507 active rotator at 4 °C. The resulting gel slurry was then filtered through 0.45 μm
508 filters (Minisart® Syringe Filter, Sartorius). The resulting RNA was and precipitated
509 with equal volume of isopropanol to remove urea and salts, then washed by 75%
510 ethanol for 2-3 times. Air-dry the pellet and resuspend with RNase-free water. The

511 RNA was quantified using NanoDrop spectrophotometer (Thermo Scientific) and
512 kept at -80 °C before use.
513
514 **Splicing reactions.** Splicing reactions of 0.05 μM wild-type and mutated *Cte* RNAs
515 were carried out in splicing buffer 40 mM HEPES (pH 7.5), 240 mM KCl at 23 °C.
516 The RNAs were heated to 80 °C for 1 min and cooled to 60 °C (0.5 °C/s) in the
517 splicing buffer. After annealing, a final concentration of 20 mM or 100 mM $MgCl_2$
518 was added to start the splicing assay, samples were then incubated at 23 °C for 30
519 min.
520
521 For cryo-EM sample preparation, the splicing reaction was activated in a total volume
522 of 0.5-1 ml as described above. The mixture was then rapidly chilled on ice and
523 concentrated to approximately 15 μM by concentrator columns with 100 kDa cutoff
524 (Ultrafiltration Centrifugal Tube, Millipore) at 4 °C.
525
526 For subsequent biochemical mechanism investigations, the splicing assay was
527 quenched by adding an equal volume of denaturing loading buffer containing 100 mM
528 EDTA, 8 M urea, 0.1% SDS, 0.025% xylene cyanol and 0.025% bromophenol blue.
529 Then the electrophoreses were performed on denaturing 8M urea and 6% 19:1
530 acrylamide:bis gel, running at 12 W for 2 h. The gels were stained for 10 min in
531 SyBrGold (Invitrogen) and visualized by Typhoon FLA 9500 (GE).
532
533 **Cryo-EM sample preparation.** A total of 3 μl of the resulting splicing products
534 (about 15 μM) was applied onto glow-discharged (10 s) 200-mesh R1.2/1.3-2nmC
535 Quantifoil Cu grids. The grids were blotted for 1.5 s in 100% humidity at 4 °C with
536 no blotting offset and rapidly frozen in liquid ethane using a Vitrobot Mark IV
537 (Thermo Fisher).
538
539 **Cryo-EM data collection and image processing**
540 Cryo-EM data were acquired using a Titan Krios microscope and Krios G4 (Thermo
541 Fisher) operating at 300 kV and 200 kV. The Titan Krios microscope was equipped
542 with a K3 summit direct detector (Gatan) and a GIF quantum energy filter (Gatan) set
543 to a slit width of 20 eV. The Titan Krios microscope was equipped with a Falcon 4i
544 summit direct detector (Thermo Fisher). Automated data collection was performed
545 using the EPU software[49] and SerialEM software[50] with the beam-image shift
546 method[51].
547

548 Movies of the Cte-fl_$MgCl_2$, A127G_$MgCl_2$ were captured in super-resolution mode
549 at a nominal magnification of 100,000× for Cte-fl_$MgCl_2$ and 105,000× for
550 A127G_$MgCl_2$ resulting in a physical pixel size of 0.959Å for Cte-fl_$MgCl_2$ and
551 0.832 Å for A127G_$MgCl_2$. Movies of the U82G_$CaCl_2$, G220C_$CaCl_2$,
552 G221U_$CaCl_2$, trans-splicing_$CaCl_2$ and trans-splicing_$MgCl_2$ were separately
553 recorded at a magnification of 120,000×, corresponding to a physical pixel size of 1.2
554 Å. Movies of the U133G_$CaCl_2$, were recorded at a magnification of 100,000×,
555 corresponding to a physical pixel size of 0.82 Å. Each movie covered a defocus range
556 from -1.2 μm to -2.2 μm. Movie stacks were dose-fractionated into 40 frames, with a
557 total exposure dose of approximately 50-51 e-/$Å^2$.
558
559 In cryo-EM data processing, a standardized procedure was employed throughout the
560 analysis within cryoSPARC v4[52]. The initial steps involved patch motion correction
561 and patch contrast transfer function (CTF) estimation to rectify stage shift, beam-
562 induced motion, and imaging distortions. Rigorous criteria were applied during the
563 selection of high-quality images, taking into account factors such as ice thickness,
564 defocus range, and estimated resolution.
565
566 The particle processing pipeline began with initial particle picking using blob picking
567 and template picking, followed by two rounds of reference-free 2D classification.
568 Subsequently, a subset of high-quality particles underwent ab-initio reconstruction,
569 yielding initial references. All viable particles then underwent multi-round 3D
570 classification without mask. Classes containing well-defined D6 domains were re-
571 extracted without binning. Through non-uniform refinement, local motion correction,
572 and global refinement without mask and local refinement with mask, high-resolution
573 maps were obtained. The reported resolutions above are based on the gold-standard
574 Fourier shell correlation (FSC) 0.143 criterion. Subsequently, local resolution
575 estimation was conducted. The final refined maps underwent separate sharpening
576 using either DeepEMhancer[53] or EMReady[54]. The resulting sharpened maps, featuring
577 enhanced density, were employed for subsequent structural analyses. All the
578 visualization and evaluation of 3D density maps were performed with UCSF
579 Chimera[55] and ChimeraX[56]. Detailed cryoEM data processing workflows are outlined
580 as indicated in Fig. S2, S4, S5, S6, S8 and S9.
581
582 **Cryo-EM model building and refinement**. We de novo reconstructed the Cte-fl_Pre
583 model by the information of IBS-EBS according to the conserved secondary structure
584 as the starting point of model. The other has no high resolution of map manually
585 adjusted and rebuilt with Coot[57] according the Alphafold3 prediction model. These

models were refined with Phenix.real_space_refine[58], yielding an averaged model-map correlation coefficient (CCmask) in range from 0.84 to 0.93, respectively. Final models were validated by MolProbity[59]. Secondary structure diagrams were drawn with RNArtist then manually regulated at Adobe Illustrator.

Acknowledgements

Cryo-EM data were collected at the Center of Cryo-Electron Microscopy at Life Medicine in Fudan University and the Cryo-EM Core Facility at UMass Chan Medical School. The data processed at Life Medicine High Performance Computing Center in Fudan University and HPC cluster in UMass Chan Medical School. Thank to Dr. Hongyun Shen from the Center of Cryo-Electron Microscopy at Life Medicine in Fudan University for helping to collect 300 kV cryo-EM data and Kangkang Song from the Cryo-EM core facility at UMass Chan Medical School for helping to collect 200 kV and 300 kV cryo-EM data. This work was supported by Natural Science Foundation of China (NSFC 31971130) to J.M and by Startup grant of W.F.

Author contributions

X.L. conceived the project; J.M., W.F. and Z.C. supervised the project; X.L. and Y.Y. prepared cryo-EM samples and performed electrophoresis and analyzed gel electrophoresis results; X.L., X.Z., K.S., C.X. and Y.Y. collected cryo-EM data; X.L. Y.Y. and X.Z. processed cryo-EM data. X.L, G.D., J.G. and J.M. built, refined and validated atomic coordinate models. X.L. and Y.Y. wrote manuscript and draw figures. Other authors contributed to the preparation of the manuscript.

Competing interests

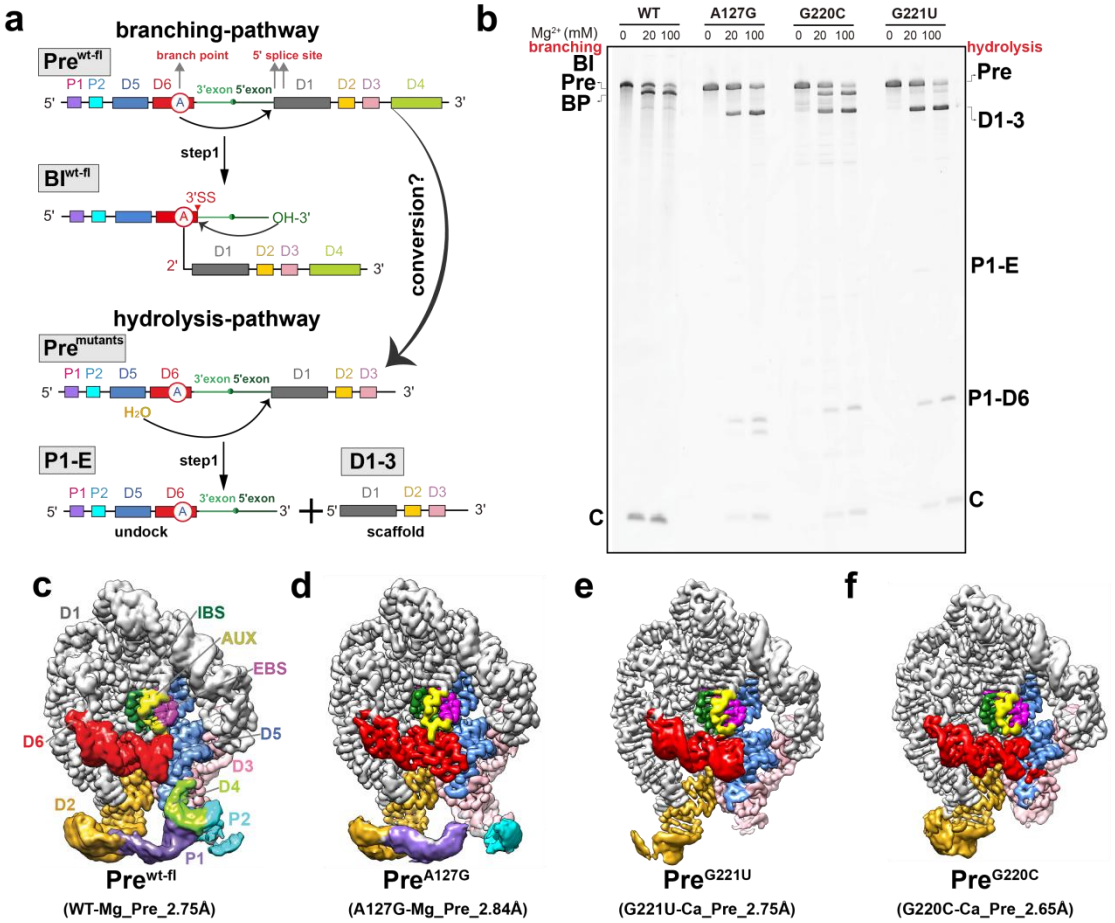
The authors declare no competing interests.

Data availability

The cryo-EM maps and associated atomic coordinate models of Cte-fl_$MgCl_2$_Pre, Cte-fl_$MgCl_2$_BP^{pre}, A127G_$MgCl_2$_Pre, A127G_$MgCl_2$_D1-3, G220C_$CaCl_2$_Pre, G221U_$CaCl_2$_Pre, hydrolysis-intermediate_$CaCl_2$, hydrolysis-product1_$MgCl_2$ and hydrolysis-product2_$MgCl_2$ have been deposited in the wwPDB OneDep System under EMD accession codes EMD-74859, EMD-74867, EMD-74857, EMD-74868, EMD-74869, EMD-74871, EMD-74872, EMD-74877, EMD-74874 and PDB ID

624 codes 9ZV6, 9ZVB, 9ZV5, 9ZVC, 9ZVD, 9ZVF, 9ZVG, 9ZVK, 9ZVI respectively.

625 **Figures:**



626

627 **Figure 1. Cryo-EM structures of hydrolysis-pathway by *Cte* mutants in back-**
628 **splicing. a.** Cartoon of the 1st step of branch-pathway and hydrolysis-pathway of *Cte*
629 group II intron back-splicing. **b.** The Syber Gold-stained 6% urea–polyacrylamide gel
630 electrophoresis (PAGE) gels showing different mutants back-splicing activity by
631 branch pathway or hydrolysis pathway at 20 or 100 mM MgCl2 concentration. WT
632 and A127G samples used to cryo-EM. **c-f.** The Pre states of cryo-EM map and
633 molecular model of the *Cte* 1_fl and branch point mutant A127G, 5′ SS mutant
634 G220C and G221U.

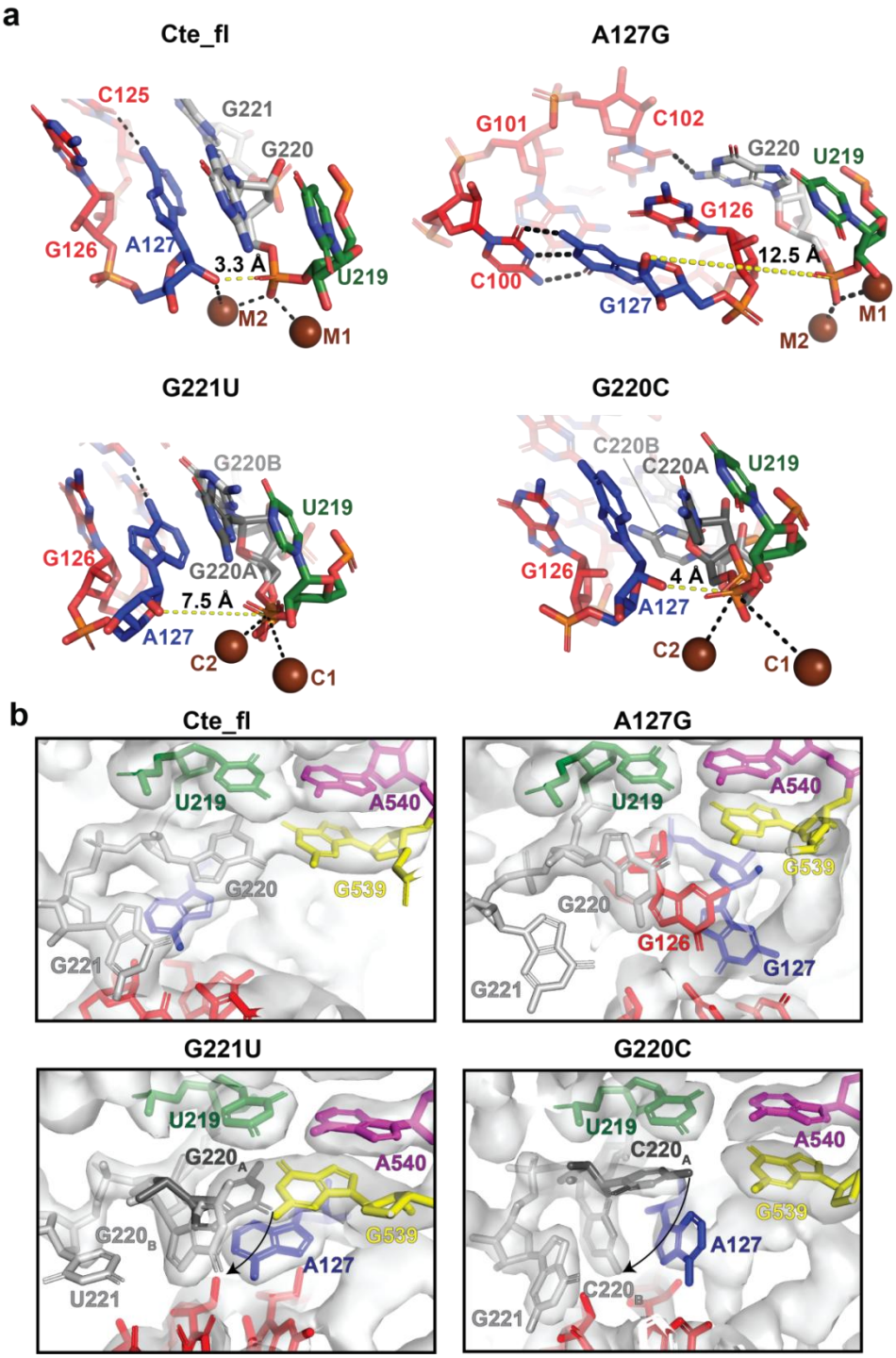


Figure 2: The catalytical center of hydrolysis pathway mutants in Pre states. a-d. The catalytical center of *Cte* full length (Cte-fl), mutants A127G, G221U and G220C from branch pathway conversed to complete hydrolysis pathway.

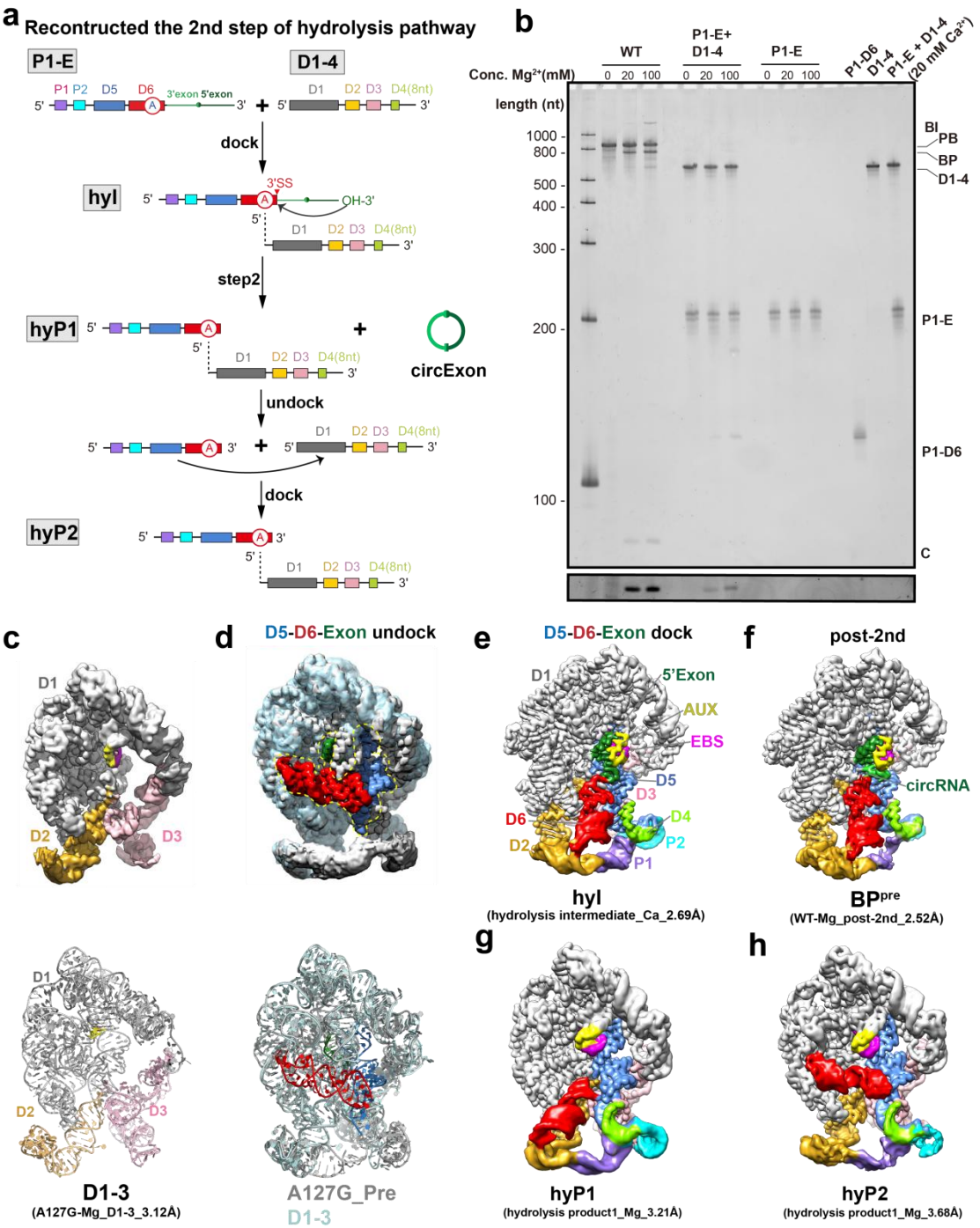


Figure 3: The 2nd step of hydrolysis pathway. a. The cartoon of the 2nd step of CP group II intron to produce circular RNA by hydrolysis pathway. **b.** Schematic diagram of the truncations P1-E and D1-3. The Syber Gold-stained 6% urea–polyacrylamide gel electrophoresis (PAGE) gels showing P1-E docked with D1-4 to produce circular exon like Cte-1 wild type. And use P1-E combined with D1-4 as cryo-EM sample. **c.** The cryo-EM map and model of D1-3. **d.** The aligned cryo-EM maps and models shos the D5-D6-Exon undocked from 1st step structure. **e-f.** Cryo-EM map and molecular model of the Cte-1 hydrolysis intermediate (hyI) and branching intermediate (BI) states.

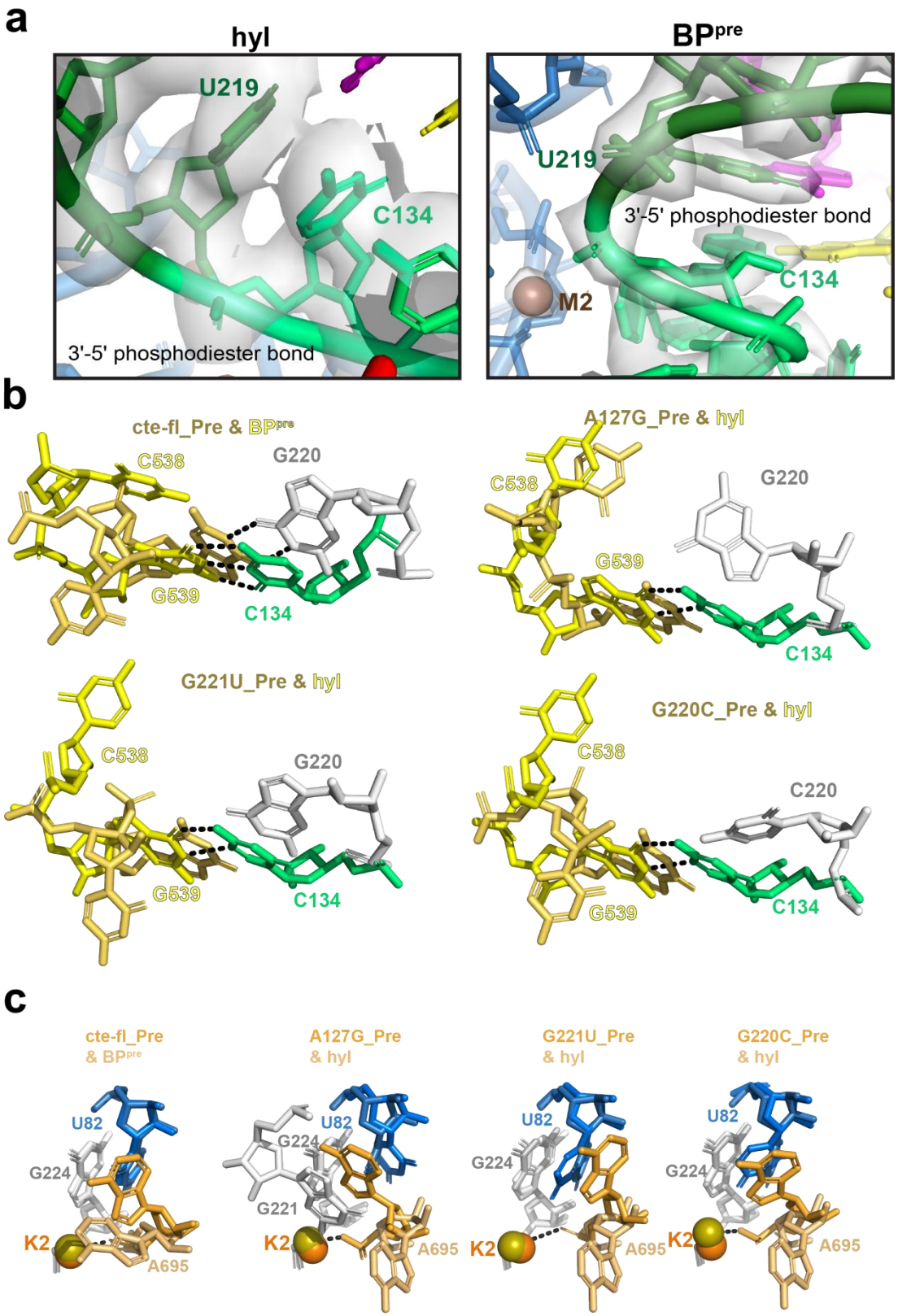


Figure 4: The conformational changes from 1st step to 2nd step by hydrolysis pathway. a. The zoom of circular RNA before forming 3′-5′ phosphodiester bond in BI state and formed in hyI conformation. **b.** The C538 and G539 regulate the process from 1st step to 2nd step of hydrolysis pathway in CP group II intron. **c.** The A695 of J2/3 conformation flips down from 1st step to 2nd step of hydrolysis pathway.

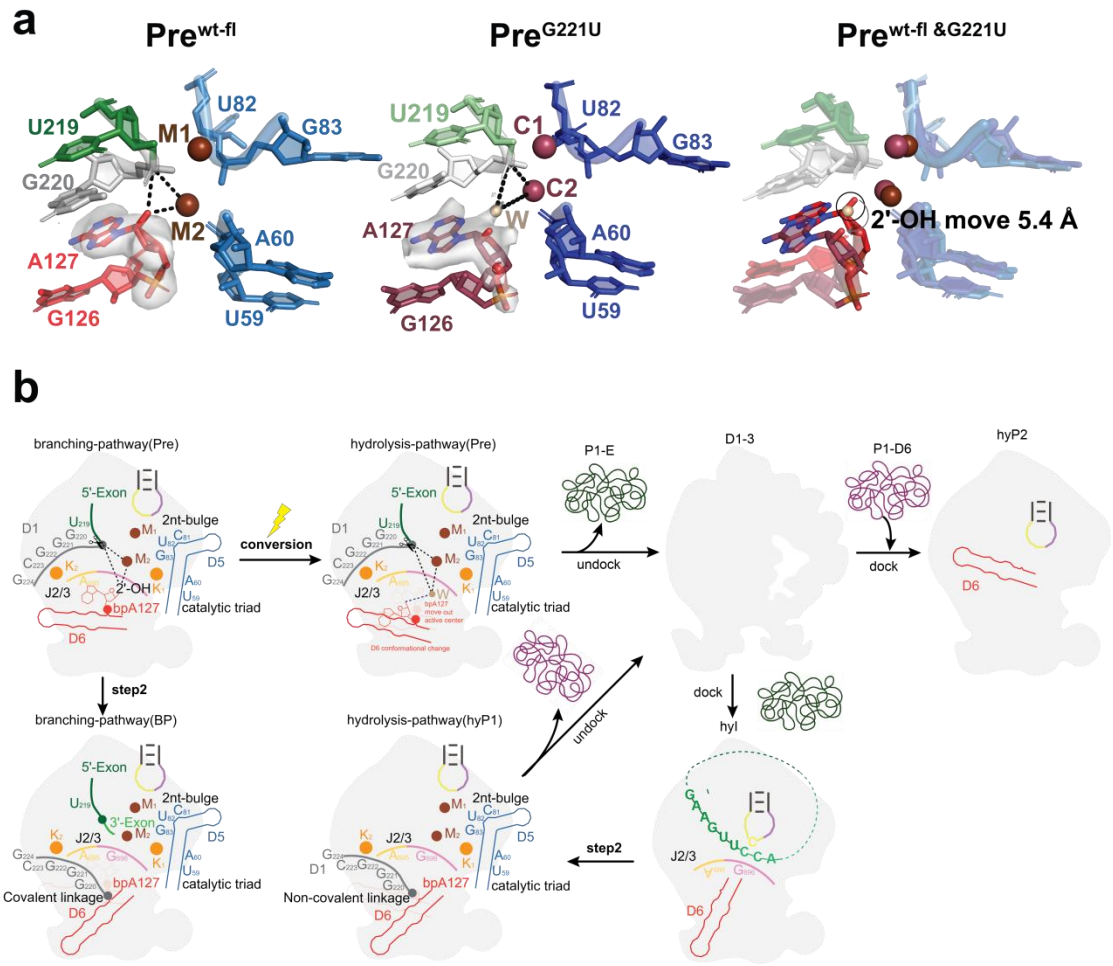
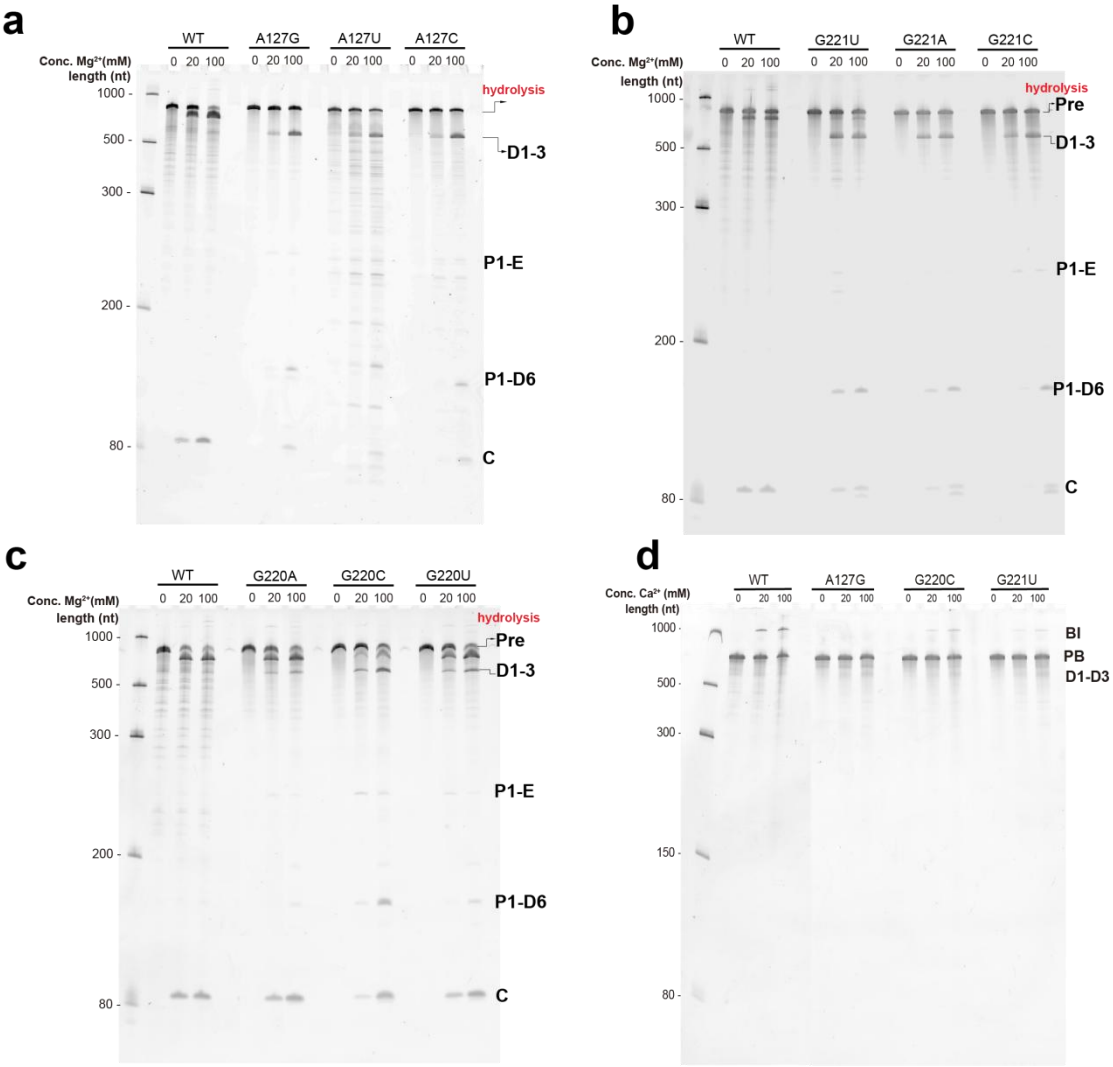


Figure 5: The mechanism of hydrolysis pathway. a. The water density of G221U compared with 2′-OH of A127 of wild type in Pre states. **b.** The mechanism of hydrolysis pathway.

661 **Supplement Figures**

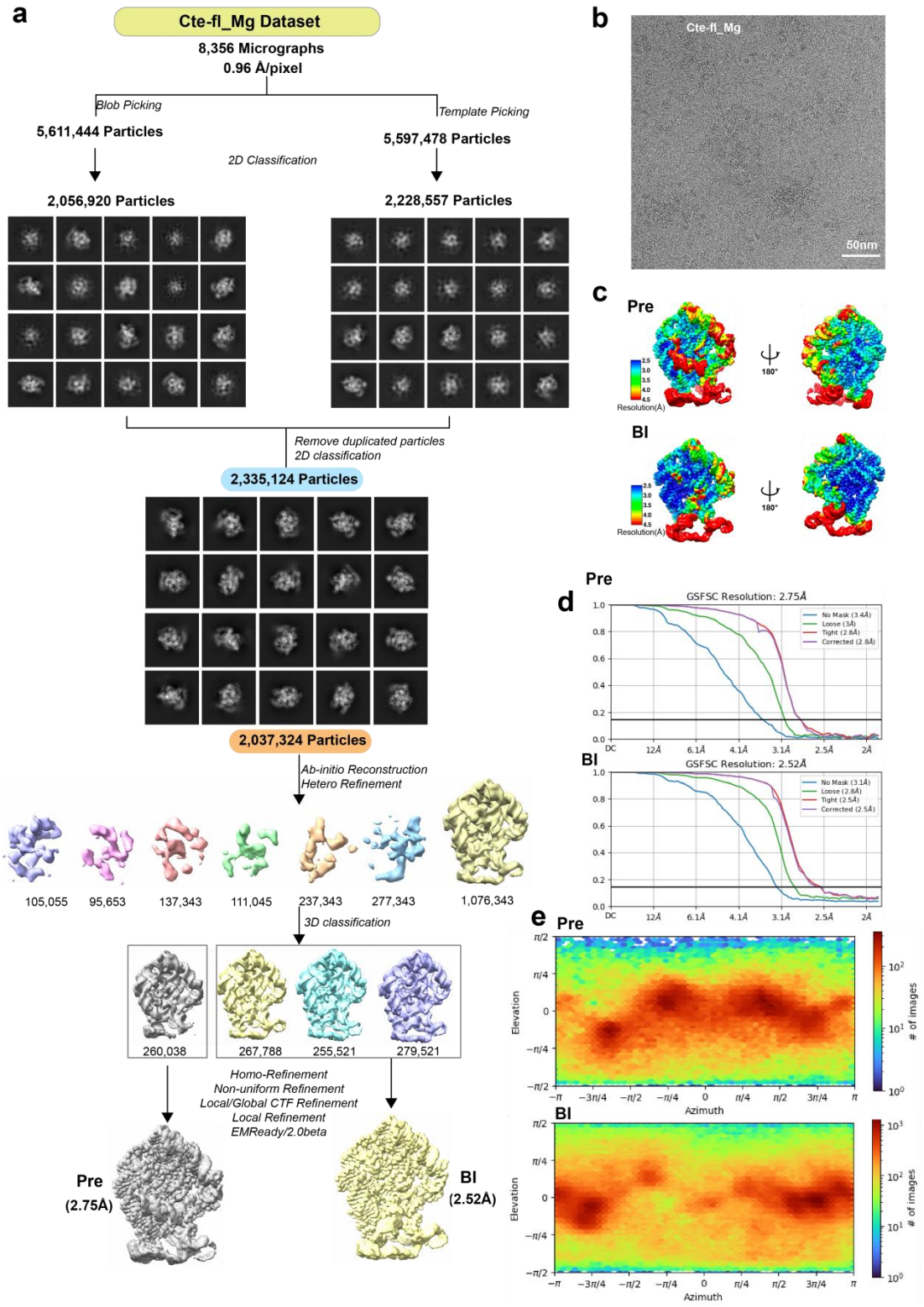
662



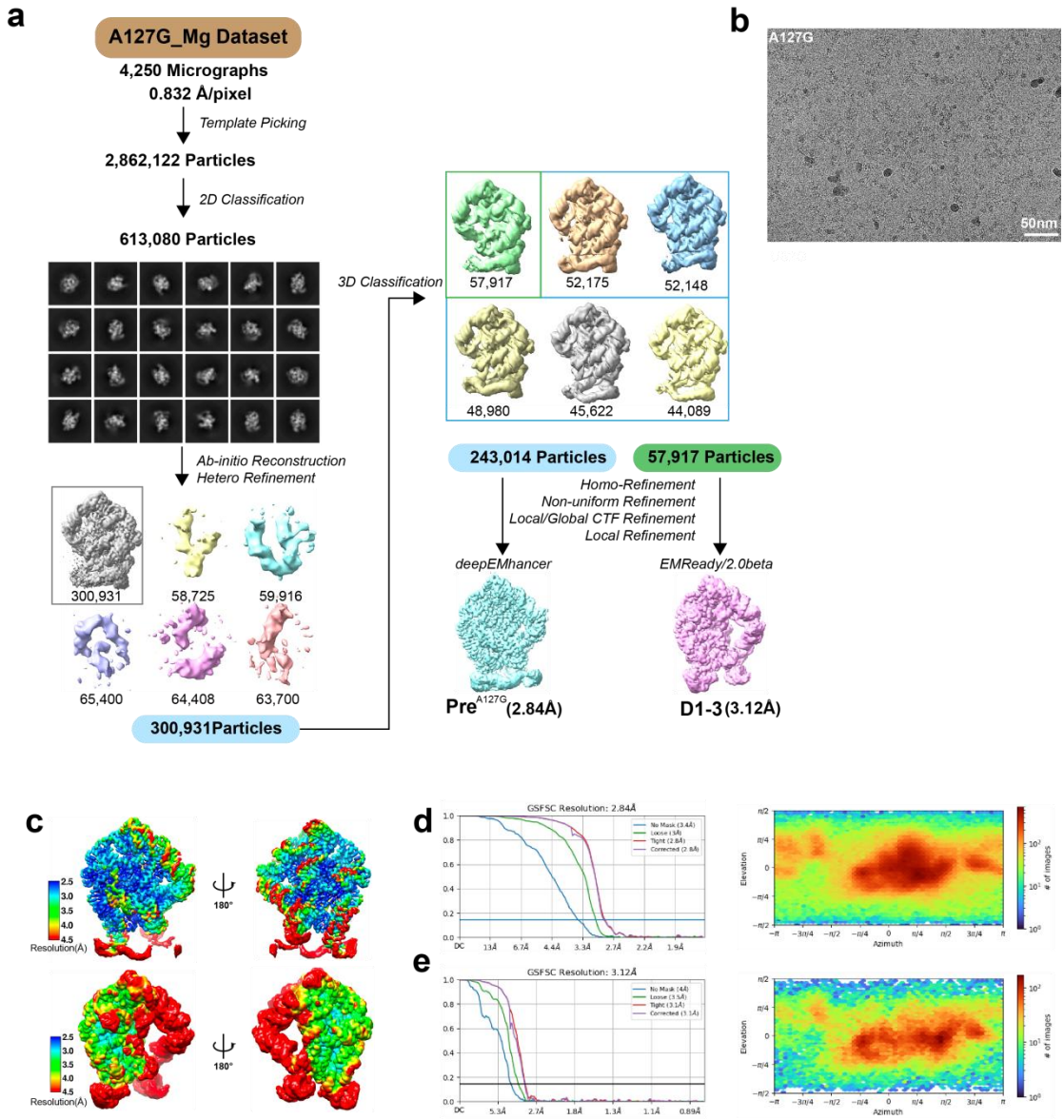
663

664 **Extended data Fig. 1: Biochemical analyses of *Cte* 1 wild type and mutants.** The
665 *Cte* 1 hydrolysis mutants back-splicing activity analysised at different Mg^{2+} (**a-c**) and
666 Ca^{2+} (**d**) concentration. The G220C and G221U samples at 20 mM $CaCl_2$ condition
667 were for cryo-EM. The experiments shown in these panels were repeated more than
668 three times with similar results.

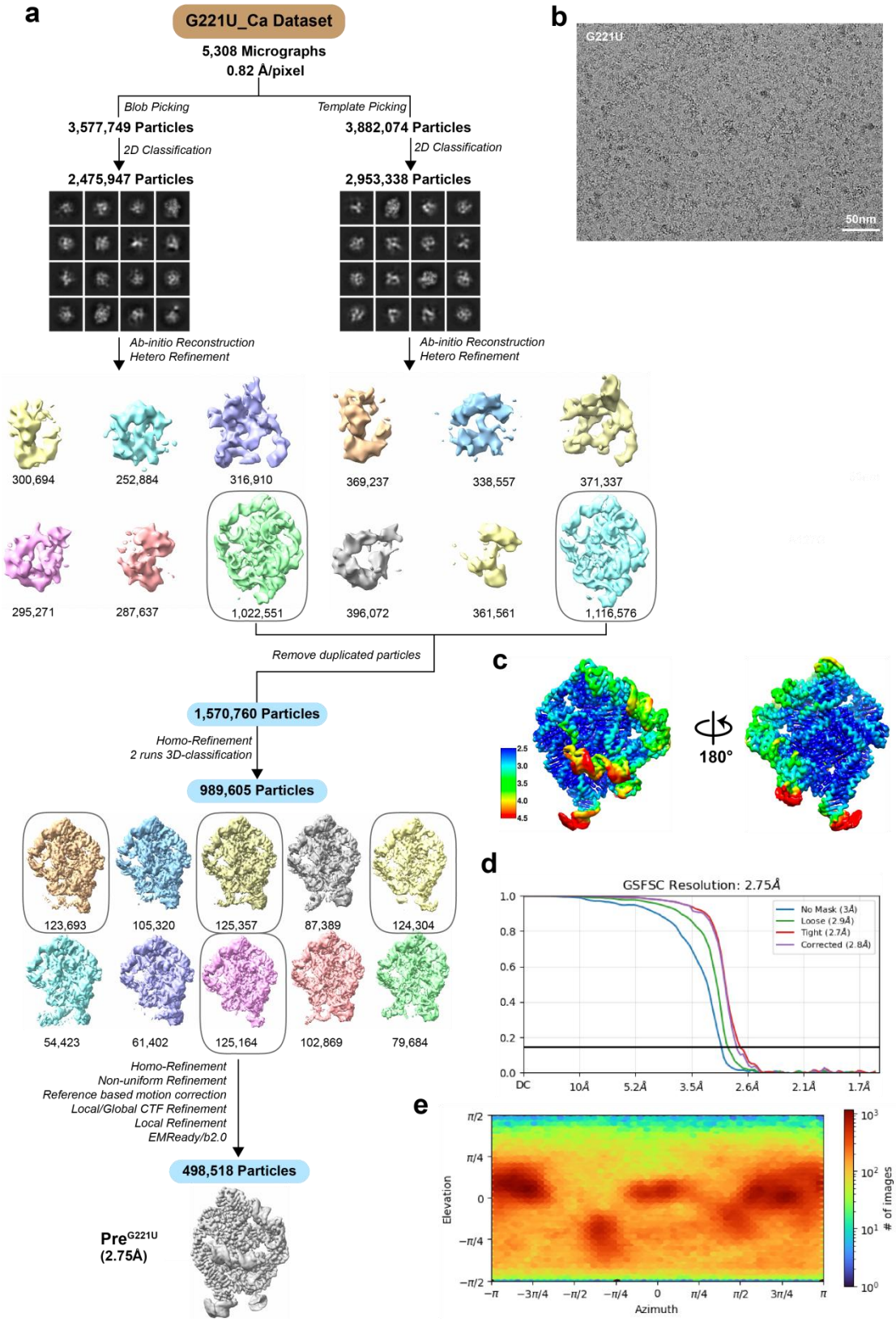
669



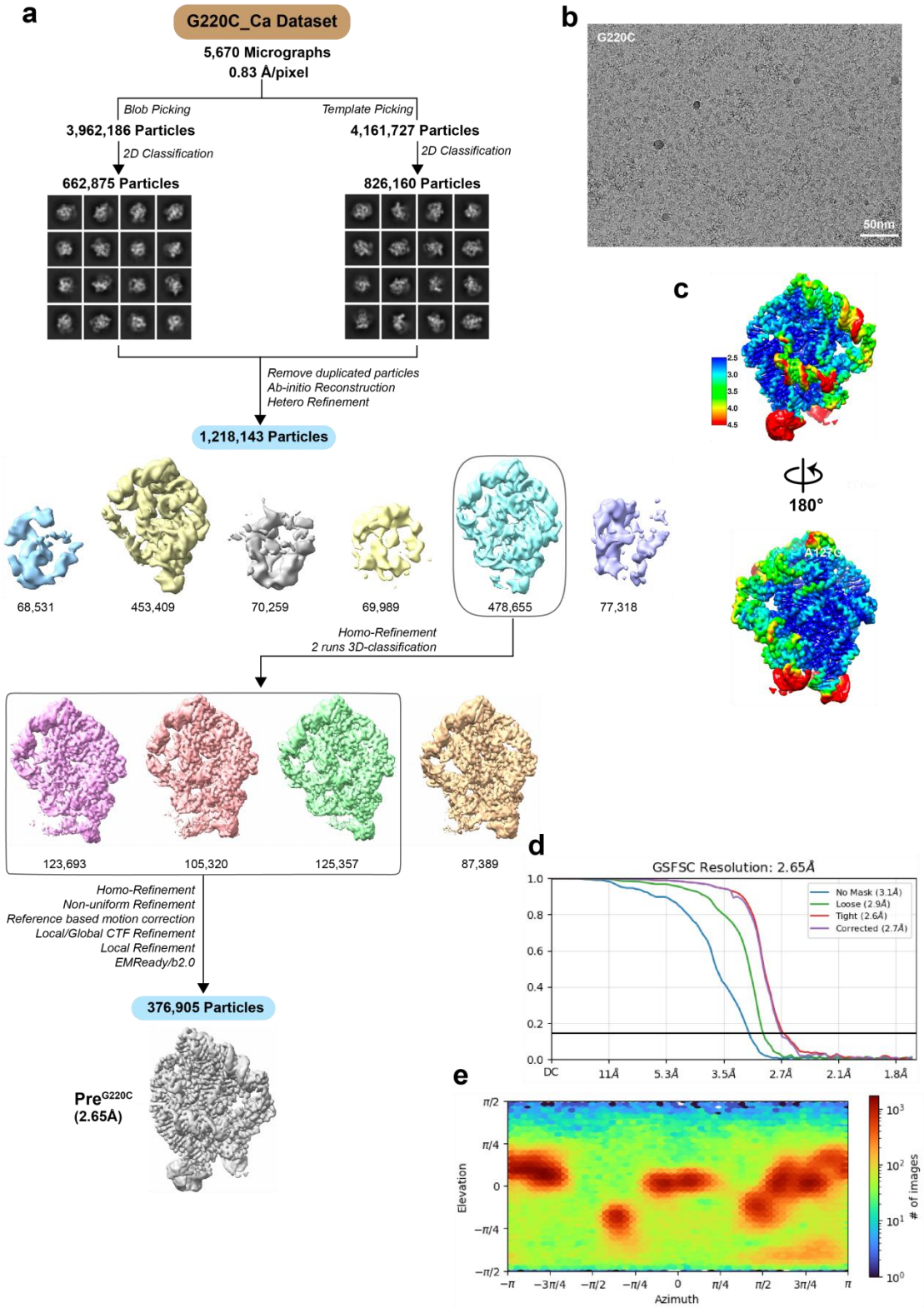
Extended data Fig. 2: Cryo-EM workflow of the wild type *Cte* 1 full length in $MgCl_2$ condition. a. Cryo-EM data processing workflow of the wild type *Cte* 1 full length at the $MgCl_2$ condition. **b-e.** The particle distribution of full micrograph (**b**) and local resolution map (**c**) and FSC curves with the gold standard threshold of 0.143 (**d**) and eular angular distribution (**e**) for *Cte* 1 full length Pre and BI conformation maps.



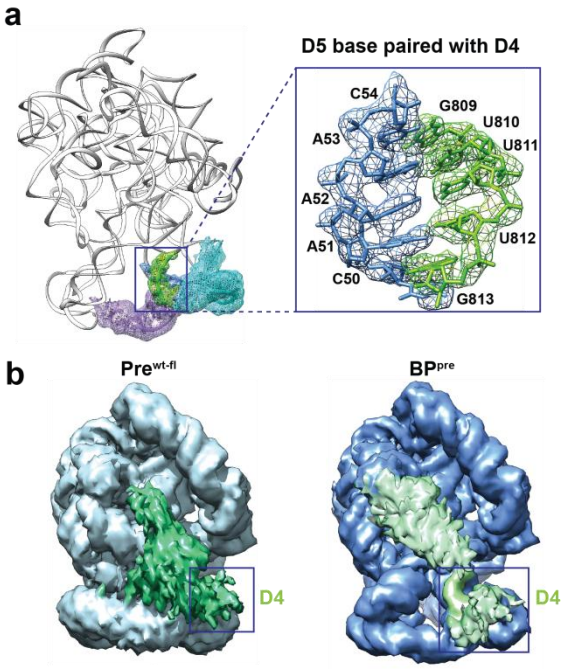
Extended data Fig. 3: Cryo-EM workflow of the branch point A127G mutant in $MgCl_2$ condition. a. Cryo-EM data processing workflow of the branch point A127G mutant sample at the $MgCl_2$ condition. **b-e.** The particle distribution of full micrograph (b) and local resolution map (c) and FSC curves with the gold standard threshold of 0.143 (d) and eular angular distribution (e) for conformation maps.



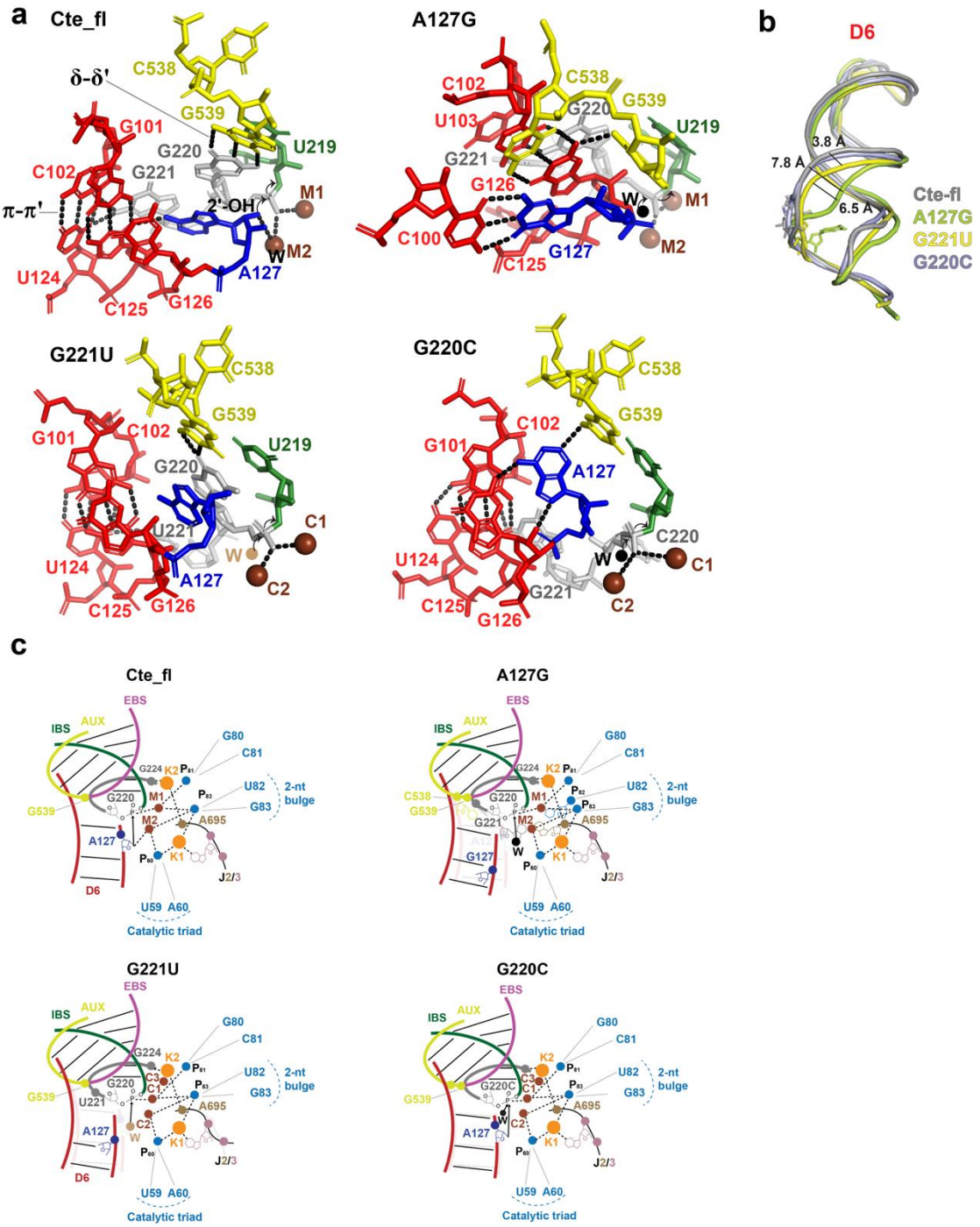
Extended data Fig. 4: Cryo-EM workflow of the 5′ SS mutant G221U in $CaCl_2$ conditions. a. Cryo-EM data processing workflow of the 5′ SS mutant G221U at the $CaCl_2$ condition. **b-e.** The particle distribution of full micrograph (**b**) and local resolution map (**c**) and FSC curves with the gold standard threshold of 0.143 (**d**) and eular angular distribution (**e**) for conformation maps.



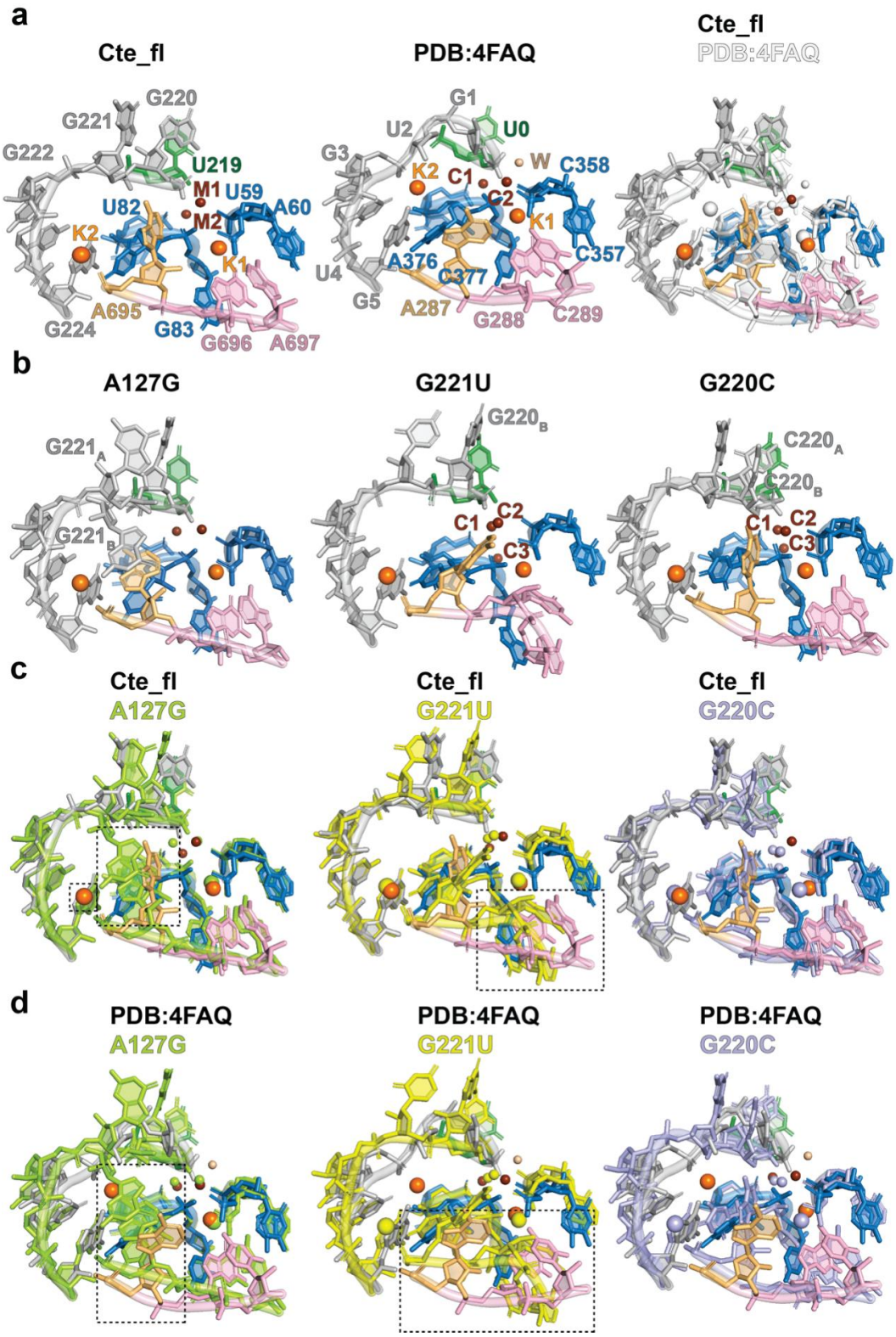
Extended data Fig. 5: Cryo-EM workflow of the 5′ SS mutant G220C in $CaCl_2$ conditions. a. Cryo-EM data processing workflow of the 5′ SS mutant G220C at the $CaCl_2$ condition. **b-e.** The particle distribution of full micrograph (**b**) and local resolution map (**c**) and FSC curves with the gold standard threshold of 0.143 (**d**) and eular angular distribution (**e**) for conformation maps.



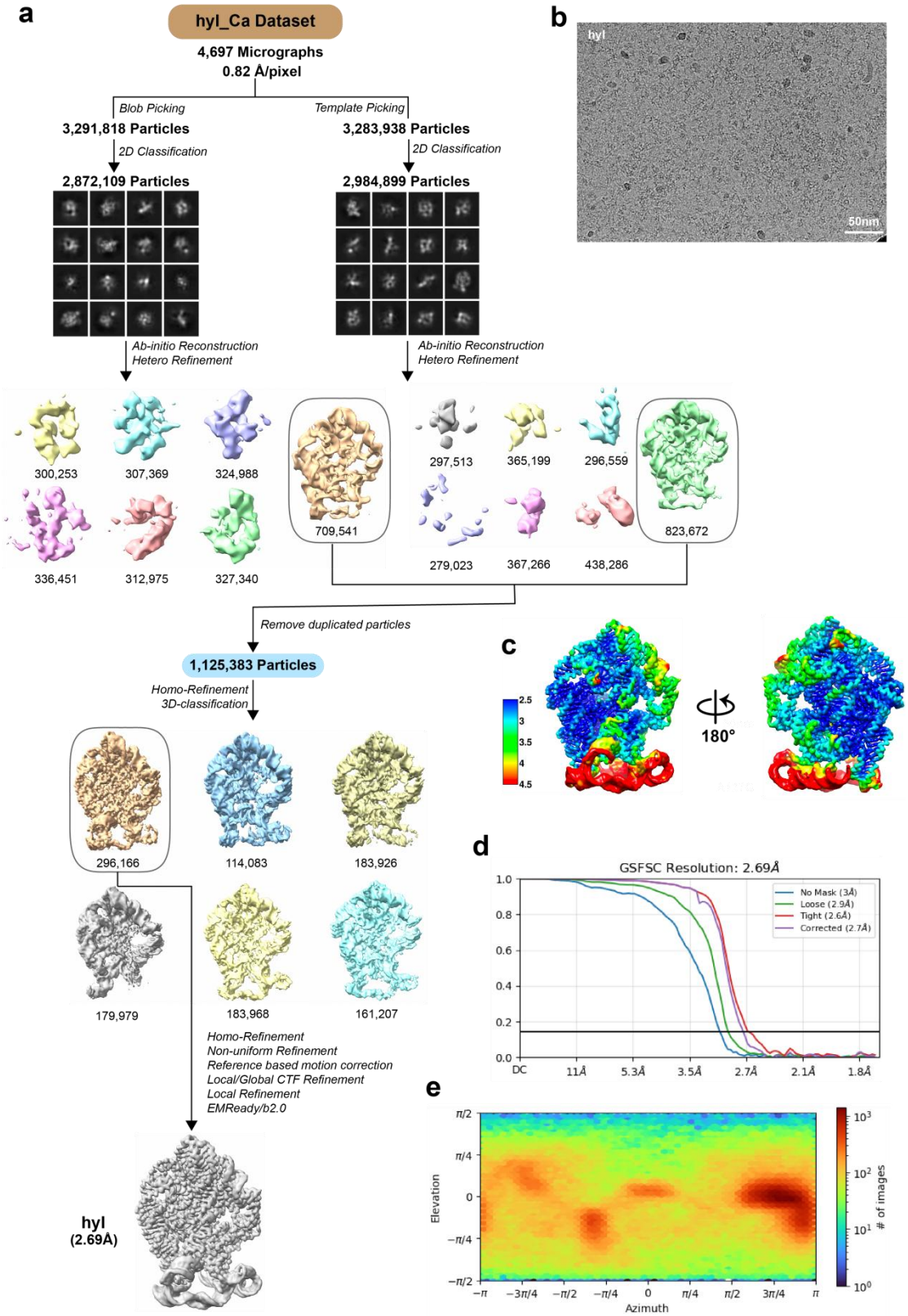
Extended data Fig. 6: The extra density of D4 domain in *Cte* 1 wild type. a. The residues of 809-813 of D4 stable by base pairing with residues of 50-54 of D5 domain. **b.** The low density show the D4 domain position of $Pre^{wt\text{-}fl}$ and BP^{post} states.



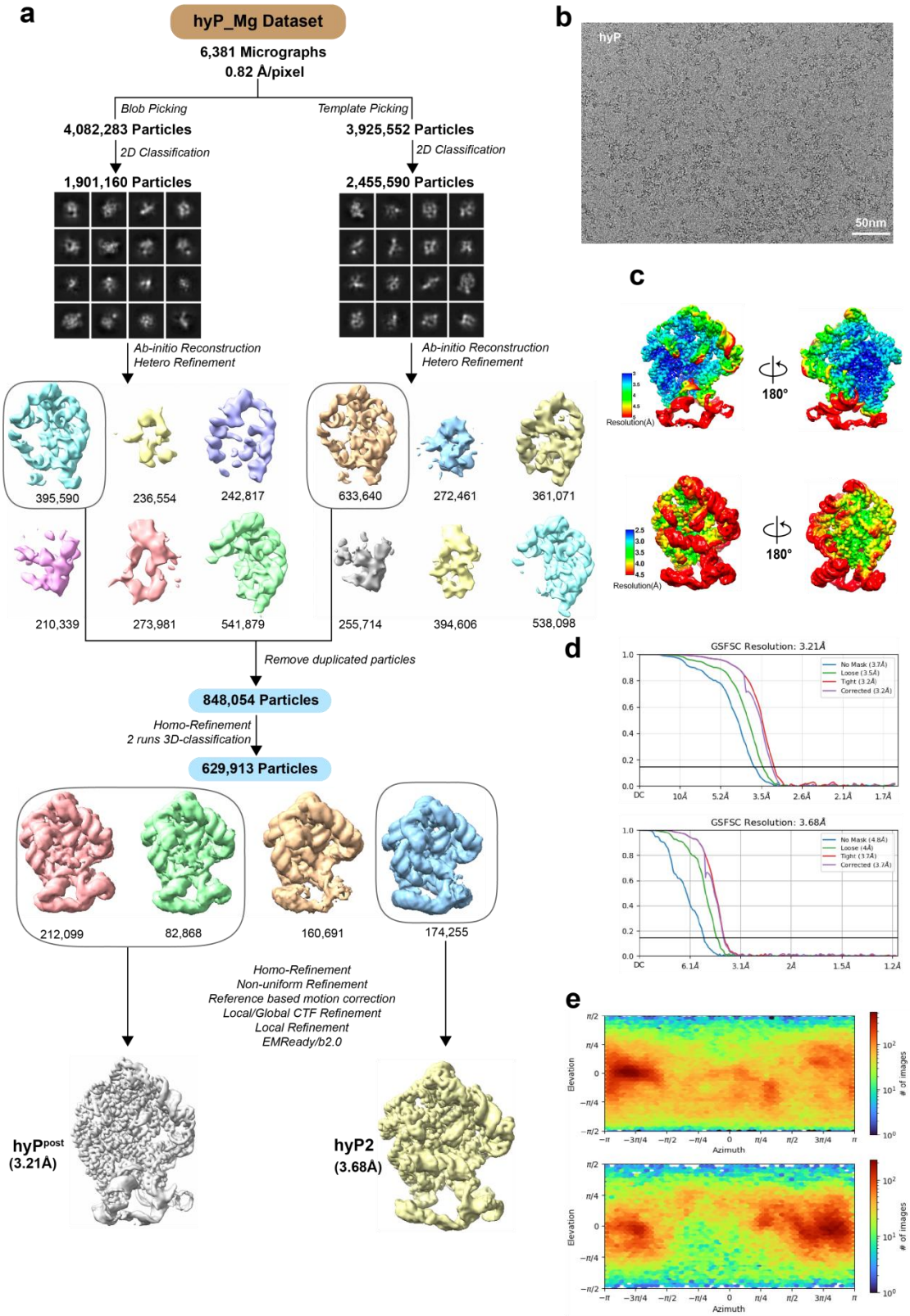
Extended data Fig. 7: The catalytical center comparation between wild type and hydrolysis mutants. a. The C538 of mutant A127G stick into D6 small groove to asistant the branch pathway change to hydrolysis pathway. **b.** Schematic diagram of the catalytic center of *Cte* full length and hydrolysis mutants.



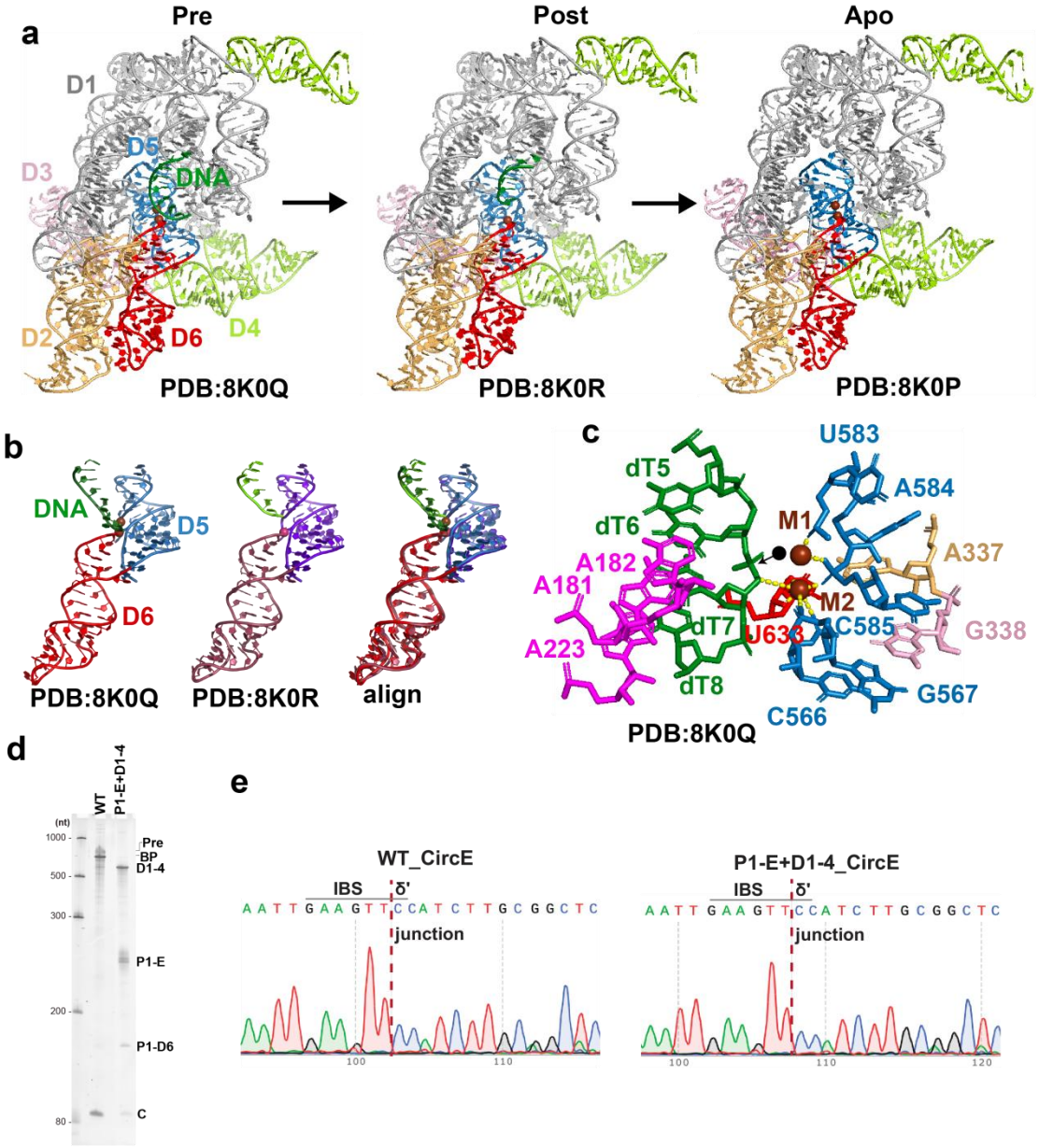
Extended data Fig. 8: The triple helix conformation and mental cluster of hydrolysis pathway mutants in Pre states. a. The triple helix conformation and mental cluster comparision of *Cte* full length (Cte-fl) and group IIC intron (PDB: 4FAQ) from *Oceanobacillus iheyensis*. **b.** The triple helix conformation and mental cluster of mutants A127G, G221U and G220C. **c-d.** The triple helix conformation and mental cluster comparision of Cte-fl or group IIC intron and mutants A127G, G221U and G220C.



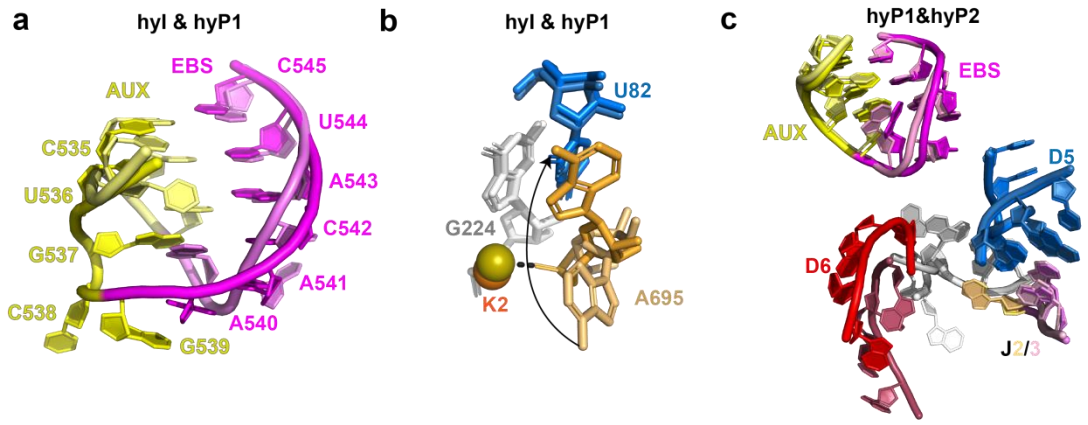
Extended data Fig. 9: Cryo-EM workflow of the hydrolysis intermediate in $CaCl_2$ conditions. a. Cryo-EM data processing workflow of the hydrolysis intermediate at the $CaCl_2$ condition. **b-e.** The particle distribution of full micrograph (**b**) and local resolution map (**c**) and FSC curves with the gold standard threshold of 0.143 (**d**) and eular angular distribution (**e**) for conformation maps.



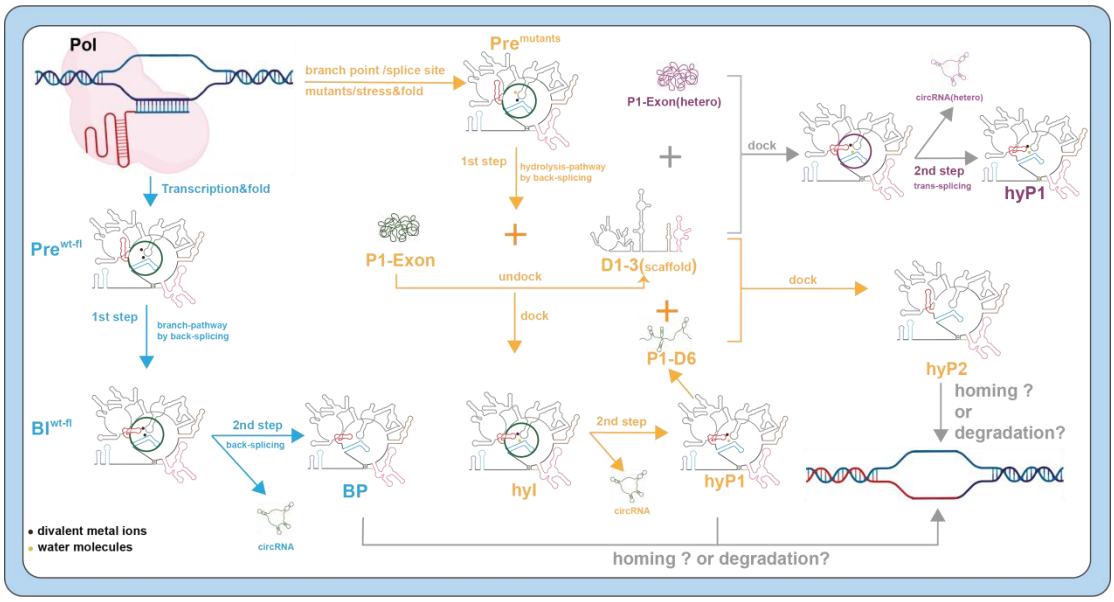
Extended data Fig. 10: Cryo-EM workflow of the hydrolysis product in $MgCl_2$ conditions. a. Cryo-EM data processing workflow of the hydrolysis product at the $MgCl_2$ condition. **b-e.** The particle distribution of full micrograph (**b**) and local resolution map (**c**) and FSC curves with the gold standard threshold of 0.143 (**d**) and eular angular distribution (**e**) for conformation maps.



Extended data Fig. 11: The catalytical center analysis of group IIC intron by hydrolysis pathway. a. The structures of group IIC intron HYER editing DNA substrate by hydrolysis pathway in reverse splicing. **b.** The zoom in of D5-D6-Exon domain of HYER from pre state to post state and alignment. **c.** The catalytical center of HYER pre state. The black circle represents the theoretical position of water molecules. **d-e.** The Syber Gold-stained 6% urea–polyacrylamide gel electrophoresis (PAGE) gels showing *Cte 1* wild type and hydrolysis truncation (P1-E+D1-4) produce circular Exon (CircE) in 100 mM MgCl2 conditions, the CircE recycled from this gel and sequencing after RT-PCR experiments. The sequencing result showing in **e**.



Extended data Fig. 12: From 2nd step to product of CP group II intron by hydrolysis pathway. a. The AUX and EBS confromation changes from hydrolysis intermediate (hyI) to hydrolysis product1 (hyP1) states. **b.** The A695 of J2/3 flip back from hyI to hyP1 state. **c.** The catalytical center alignment of hyP1 and hydrolysis product2 (hyP2) conformations.



Extended data Fig. 13: The plausible mechanism of CP group II intron to produce circular RNA in vivo.

751 **Table S1: Cryo-EM Data Collection, Refinement and Validation Statistics**

Cryo-EM data collection and processing	$Pre^{wt\text{-}fl}$	BP^{pre}	Pre^{A127G}	D1-3
Microscope	Titan Krios			
Voltage (kV)	300			
GIF Quantum energy filter slit	20			
Detector	Falcon 4i		Gatan K3	
Nominal Magnification	100,000 ×		105,000 ×	
Pixel size (Å)	0.959		0.83	
Symmetry imposed	C1			
Defocus range (μm)	-1.2– -2.2			
Electron exposure ($e^-/Å^2$)	50	50	51	51
Micrographs (acquired/used)	8356		4250	
Number of extracted particles	2,335,124		613,080	
Number of particles after 3D classifications	260,038	802,830	243,014	57,917
Map resolution at 0.143 FSC criterion (Å)	2.75	2.52	2.84	3.12
Local resolution range (Å)	2.5-4.5	2.5-4.5	2.5-4.5	2.5-4.5
Model refinement				
Atoms	15891	15950	15555	12684
Residues	737	740	720	587
CC_{mask}	0.93	0.93	0.84	0.87
$Resolution_{FSC\ map\ vs.\ model\ @\ 0.5}$ (Å)	2.75	2.52	2.84	3.12
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.006	0.006
Bond angles (°)	0.904	1.003	0.762	0.861
Validation				
Clash score	4.13	4.19	3.91	7.82
MolProbity score	2.35	2.35	2.33	2.58

752

Cryo-EM data collection and processing	Pre^{G220C}	Pre^{G221U}	hyI	hyP1	hyP2
Microscope	Titan Krios	Titan Actica			
Voltage (kV)	300	200			
Detector	Gatan K3				
Nominal Magnification	105,000 ×	100,000 ×			
Pixel size (Å)	0.83	0.82			
Symmetry imposed	C1				
Defocus range (μm)	-1.0– -1.8				
Electron exposure (e^-/Å^2)	50.15	51.21	50.12	50.46	
Micrographs (acquired/used)	5670	5308	4697	6381	
Number of extracted particles	1,218,143	1,570,760	1,125,383	848,054	
Number of particles after 3D classifications	376,905	498,518	296,166	294,967	174,255
Map resolution at 0.143 FSC criterion (Å)	3.62	3.23	2.69	3.21	3.68
Local resolution range (Å)	2.5-4.5	2.5-4.5	2.5-4.5	3-5	2.5-4.5
Model refinement					
Atoms	14510	14562	15860	15635	15729
Residues	673	675	735	726	730
CC_{mask}	0.92	0.91	0.92	0.91	0.89
$Resolution_{FSC\ map\ vs.\ model\ @\ 0.5}$ (Å)	2.65	2.75	2.69	3.21	3.68
R.m.s. deviations					
Bond lengths (Å)	0.007	0.006	0.008	0.008	0.008
Bond angles (°)	0.786	0.875	1.024	1.009	1.074
Validation					
Clash score	6.7	5.49	4.37	5.19	9.43
MolProbity score	2.52	2.45	2.36	2.43	2.65

753
754

Reference:

1 Lambowitz, A. M. & Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* **3**, a003616 (2011). https://doi.org:10.1101/cshperspect.a003616

2 Pyle, A. M. Group II Intron Self-Splicing. *Annu Rev Biophys* **45**, 183-205 (2016). https://doi.org:10.1146/annurev-biophys-062215-011149

3 Galej, W. P. Structural studies of the spliceosome: past, present and future perspectives. *Biochem Soc Trans* **46**, 1407-1422 (2018). https://doi.org:10.1042/bst20170240

4 Jacquier, A. & Jacquesson-Breuleux, N. Splice site selection and role of the lariat in a group II intron. *Journal of Molecular Biology* **219**, 415-428 (1991). https://doi.org:https://doi.org/10.1016/0022-2836(91)90183-7

5 Daniels, D. L., Michels, W. J., Jr. & Pyle, A. M. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *J Mol Biol* **256**, 31-49 (1996). https://doi.org:10.1006/jmbi.1996.0066

6 Zhao, C. & Pyle, A. M. Structural Insights into the Mechanism of Group II Intron Splicing. *Trends Biochem Sci* **42**, 470-482 (2017). https://doi.org:10.1016/j.tibs.2017.03.007

7 Podar, M., Chu, V. T., Pyle, A. M. & Perlman, P. S. Group II intron splicing in vivo by first-step hydrolysis. *Nature* **391**, 915-918 (1998). https://doi.org:10.1038/36142

8 Vogel, J. & Börner, T. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *The EMBO Journal* **21**, 3794-3803 (2002). https://doi.org:https://doi.org/10.1093/emboj/cdf359

9 van der Veen, R., Kwakman, J. H. & Grivell, L. A. Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *Embo j* **6**, 3827-3831 (1987). https://doi.org:10.1002/j.1460-2075.1987.tb02719.x

10 Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem* **263**, 3432-3439 (1988).

11 Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77-82 (2008). https://doi.org:10.1126/science.1153803

12 Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497-507 (2012). https://doi.org:10.1016/j.cell.2012.09.033

13 Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M. & Toor, N. Crystal structure of a group II intron in the pre-catalytic state. *Nat Struct Mol Biol* **19**, 555-557 (2012). https://doi.org:10.1038/nsmb.2270

14 Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193-197 (2014). https://doi.org:10.1038/nature13790

15 Qu, G. *et al.* Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol* **23**, 549-557 (2016). https://doi.org:10.1038/nsmb.3220

16 Costa, M., Walbott, H., Monachello, D., Westhof, E. & Michel, F. Crystal structures of a group II intron lariat primed for reverse splicing. *Science* **354** (2016). https://doi.org:10.1126/science.aaf9258

17 Haack, D. B. *et al.* Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. *Cell* **178**, 612-623.e612 (2019). https://doi.org:https://doi.org/10.1016/j.cell.2019.06.035

18 Chung, K., Xu, L., Chai, P., Peng, J., Devarkar, S. C. & Pyle, A. M. Structures of a mobile intron retroelement poised to attack its structured DNA target. *Science* **378**, 627-634 (2022). https://doi.org:10.1126/science.abq2844

19 Xu, L., Liu, T., Chung, K. & Pyle, A. M. Structural insights into intron catalysis and dynamics during splicing. *Nature* **624**, 682-688 (2023). https://doi.org:10.1038/s41586-023-06746-6

20 Marcia, M., Somarowthu, S. & Pyle, A. M. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* **4**, 14 (2013). https://doi.org:10.1186/1759-8753-4-14

21 Monachello, D., Michel, F. & Costa, M. Activating the branch-forming splicing pathway by reengineering the ribozyme component of a natural group II intron. *Rna* **22**, 443-455 (2016). https://doi.org:10.1261/rna.054643.115

22 Roth, A., Weinberg, Z., Vanderschuren, K., Murdock, M. H. & Breaker, R. R. Natural circularly permuted group II introns in bacteria produce RNA circles. *iScience* **24**, 103431 (2021). https://doi.org:10.1016/j.isci.2021.103431

23 Ling, X., Yao, Y. & Ma, J. Structures of a natural circularly permuted group II intron reveal mechanisms of branching and backsplicing. *Nat Struct Mol Biol* **32**, 1101-1110 (2025). https://doi.org:10.1038/s41594-025-01489-6

24 Abraham Punnoose, J. *et al.* High-throughput single-molecule quantification of individual base stacking energies in nucleic acids. *Nature Communications* **14**, 631 (2023). https://doi.org:10.1038/s41467-023-36373-8

25 Liu, Z.-X. *et al.* Hydrolytic endonucleolytic ribozyme (HYER) is programmable for sequence-specific DNA cleavage. *Science* **383**, eadh4859 (2024). https://doi.org:doi:10.1126/science.adh4859

26 De Lange, T., Michels, P. A., Veerman, H. J., Cornelissen, A. W. & Borst, P. Many trypanosome messenger RNAs share a common 5' terminal sequence. *Nucleic Acids Res* **12**, 3777-3790 (1984). https://doi.org:10.1093/nar/12.9.3777

27 Lasda, E. L. & Blumenthal, T. Trans-splicing. *WIREs RNA* **2**, 417-434 (2011). https://doi.org:https://doi.org/10.1002/wrna.71

28 Yang, Y. & Walsh, C. E. Spliceosome-Mediated RNA Trans-splicing. *Molecular Therapy* **12**, 1006-1012 (2005). https://doi.org:https://doi.org/10.1016/j.ymthe.2005.09.006

29 Müller, U. F. Design and Experimental Evolution of trans-Splicing Group I Intron Ribozymes. *Molecules* **22** (2017). https://doi.org:10.3390/molecules22010075

30 Belhocine, K., Mak, A. B. & Cousineau, B. Trans-splicing versatility of the LI.LtrB group II intron. *Rna* **14**, 1782-1790 (2008). https://doi.org:10.1261/rna.1083508

31 Hong, E. M., Ingemarsdotter, C. K. & Lever, A. M. L. Therapeutic applications of trans-splicing. *Br Med Bull* **136**, 4-20 (2020). https://doi.org:10.1093/bmb/ldaa028

32 Fukuzawa, H. *et al.* Coding sequences for chloroplast ribosomal protein S12 from the liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands. *FEBS Letters* **198**, 11-15 (1986).

https://doi.org:https://doi.org/10.1016/0014-5793(86)81174-7

33 Kück, U., Choquet, Y., Schneider, M., Dron, M. & Bennoun, P. Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-apoproteins in Chlamydomonas reinhardii: evidence for in vivo trans-splicing. *Embo j* **6**, 2185-2195 (1987). https://doi.org:10.1002/j.1460-2075.1987.tb02489.x

34 Knoop, V., Schuster, W., Wissinger, B. & Brennicke, A. Trans splicing integrates an exon of 22 nucleotides into the nad5 mRNA in higher plant mitochondria. *Embo j* **10**, 3483-3493 (1991). https://doi.org:10.1002/j.1460-2075.1991.tb04912.x

35 Wissinger, B., Schuster, W. & Brennicke, A. Trans splicing in Oenothera mitochondria: nad1 mRNAs are edited in exon and trans-splicing group II intron sequences. *Cell* **65**, 473-482 (1991). https://doi.org:10.1016/0092-8674(91)90465-b

36 Binder, S., Marchfelder, A., Brennicke, A. & Wissinger, B. RNA editing in trans-splicing intron sequences of nad2 mRNAs in Oenothera mitochondria. *J Biol Chem* **267**, 7615-7623 (1992).

37 Chapdelaine, Y. & Bonen, L. The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: A trans-splicing model for this gene-in-pieces. *Cell* **65**, 465-472 (1991). https://doi.org:https://doi.org/10.1016/0092-8674(91)90464-A

38 Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P. & Rochaix, J. D. Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C. reinhardtii chloroplast. *Cell* **52**, 903-913 (1988). https://doi.org:10.1016/0092-8674(88)90432-1

39 Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. & Rochaix, J. D. A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. *Cell* **65**, 135-143 (1991). https://doi.org:10.1016/0092-8674(91)90415-u

40 Knoop, V., Altwasser, M. & Brennicke, A. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet* **255**, 269-276 (1997). https://doi.org:10.1007/s004380050497

41 Zimmerly, S. & Semper, C. Evolution of group II introns. *Mobile DNA* **6**, 7 (2015). https://doi.org:10.1186/s13100-015-0037-5

42 Lambowitz, A. M. & Belfort, M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol Spectr* **3**, Mdna3-0050-2014 (2015). https://doi.org:10.1128/microbiolspec.MDNA3-0050-2014

43 Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3** (2011). https://doi.org:10.1101/cshperspect.a003707

44 Cousineau, B. *et al.* Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451-462 (1998). https://doi.org:10.1016/s0092-8674(00)81586-x

45 Lin, S. L., Miller, J. D. & Ying, S. Y. Intronic microRNA (miRNA). *J Biomed Biotechnol* **2006**, 26818 (2006). https://doi.org:10.1155/jbb/2006/26818

46 Woodson, S. A. & Cech, T. R. Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA. *Biochemistry* **30**, 2042-2050 (1991). https://doi.org:10.1021/bi00222a006

47 Inoue, T., Sullivan, F. X. & Cech, T. R. New reactions of the ribosomal RNA precursor

of Tetrahymena and the mechanism of self-splicing. *Journal of Molecular Biology* **189**, 143-165 (1986). https://doi.org:https://doi.org/10.1016/0022-2836(86)90387-6

48 Kladwang, W., Hum, J. & Das, R. Ultraviolet Shadowing of RNA Can Cause Significant Chemical Damage in Seconds. *Scientific Reports* **2**, 517 (2012). https://doi.org:10.1038/srep00517

49 Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat Protoc* **14**, 100-118 (2019). https://doi.org:10.1038/s41596-018-0084-8

50 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005). https://doi.org:10.1016/j.jsb.2005.07.007

51 Wu, C., Huang, X., Cheng, J., Zhu, D. & Zhang, X. High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. *J Struct Biol* **208**, 107396 (2019). https://doi.org:10.1016/j.jsb.2019.09.013

52 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017). https://doi.org:10.1038/nmeth.4169

53 Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S. & Vargas, J. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology* **4**, 874 (2021). https://doi.org:10.1038/s42003-021-02399-1

54 He, J., Li, T. & Huang, S.-Y. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nature Communications* **14**, 3217 (2023). https://doi.org:10.1038/s41467-023-39031-1

55 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004). https://doi.org:10.1002/jcc.20084

56 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021). https://doi.org:10.1002/pro.3943

57 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010). https://doi.org:10.1107/s0907444910007493

58 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018). https://doi.org:10.1107/s2059798318006551

59 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010). https://doi.org:10.1107/s0907444909042073