

Redefinition of Rubisco Carboxylase Reaction Reveals Origin of Water for Hydration and New Roles for Active-Site Residues

Babu Kannappan and Jill E. Gready*

Computational Proteomics Group, John Curtin School of Medical Research, Australian National University, Canberra, ACT 0200, Australia

Received May 9, 2008; E-mail: jill.gready@anu.edu.au

Abstract: Crystallographic, mutagenesis, kinetic, and computational studies on Rubisco over three decades have revealed much about its catalytic mechanism and the role played by several active-site residues. However, key questions remain unanswered. Specific details of the carboxylase and oxygenase mechanisms, required to underpin the rational re-engineering of Rubisco, are still speculative. Here we address critical gaps in knowledge with a definitive comprehensive computational investigation of the mechanism of carboxylase activity at the Rubisco active site. Density functional theory calculations (B3LYP/6-31G(d,p)) were performed on active-site fragment models of a size up to 77 atoms, not previously possible computationally. All amino acid residues suspected to play roles in the acid–base chemistry in the multistep reaction, and interacting directly with the central Mg^{2+} atom and the reactive moiety of substrate and intermediates, were included. The results provide a firm basis for us to propose a novel mechanism for the entire sequence of reactions in the carboxylase catalysis and to define precise roles for the active-site residues, singly and in concert. In this mechanism, the carbamylated LYS201 plays a more limited role than previously proposed but is crucial for initiating the reaction by acting as a base in the enolization. We suggest a wider role for HIS294, with involvement in the carboxylation, hydration, and C2–C3 bond-scission steps, consistent with the suggestion of Harpel et al. (1998) but contrary to the consensus view of Cleland et al. (1998). In contrast to the common assumption that the water molecule for the hydration step comes from within the active site, we propose that the Mg-coordinated water is not dissociated at the start of the gas-addition reaction but rather remains coordinated and is used for the hydration of the C3 carbon atom. New roles are also proposed for LYS175, GLU204, and HIS294. The mechanism suggests roles in the gas-addition step for residues in three spatially distinct regions of the active site, HIS294 and LYS334 in the C-terminal domain of the large subunit (LSU), but also hitherto unsuspected roles for a cluster of three residues (ASN123, GLU60, and TYR20) in the N-terminal domain of the partner LSU of the dimer containing the active site. Our new mechanism is supported by existing experimental data, provides new convincing interpretations of previously puzzling data, and allows new insights into mutational strategies for improving Rubisco activity.

1. Introduction

Ribulose1,5-bisphosphate carboxylase-oxygenase (EC 4.1.1.39), named Rubisco,¹ plays a critical role in the global carbon cycle providing the main route for fixation of atmospheric CO_2 into the biosphere^{2–4} to the tune of about 400 billion tonnes per year. As part of the Calvin cycle, Rubisco catalyzes a sequence of reactions during photosynthesis that results in the fixation of CO_2 ⁵ into energy-rich carbon compounds that fuel all life. Despite this critical role, Rubisco has the paradoxical distinction of being a very inefficient enzyme, with a catalytic rate of only

a few CO_2 fixations per second, and this sets fundamental limits on the efficiency of photosynthesis. The rapid fall in atmospheric CO_2 partial pressure (pCO_2) over geologic time scales since Rubisco first appeared⁶ has compounded the effects of this deficiency. Furthermore, Rubisco fails to differentiate efficiently between its natural substrate CO_2 and atmospheric oxygen, the concentration of which has risen rapidly over the same evolutionary time scale, and catalyzes an analogous sequence of reactions fixing O_2 . As shown in Figure 1, both of these gaseous molecules are small and have similar patterns of electrostatic potential.

The apparent failure of the evolution machinery to retune the catalytic efficiency and substrate selectivity of the enzyme under changing environmental conditions is widely attributed to the complexity of the reaction.^{1,5} Rubisco catalyzes a series of steps, with the chemistry taking place within the same active-

(1) Wildman, S. G. *Photosynth. Res.* **2002**, 73, 243–250.

(2) Mann, C. C. *Science* **1999**, 283, 314–316.

(3) Parry, M. A. J.; Andralojc, P. J.; Mitchell, R. A. C.; Madgwick, P. J.; Keys, A. J. *J. Exp. Bot.* **2003**, 54, 1321–1333.

(4) Spreitzer, R. J.; Salvucci, M. E. *Annu. Rev. Plant Biol.* **2002**, 53, 449–475.

(5) Andrews, T. J.; Lorimer, G. H. In *The Biochemistry of Plants: A Comprehensive Treatise*; Hatch, M. D., Broadman, N. K., Eds.; Academic Press: New York, 1987; Vol. 10, pp 131–218.

(6) Beerling, D. J.; Lake, J. A.; Berner, R. A.; Hickey, L. J.; Taylor, D. W.; Royer, D. L. *Geochim. Cosmochim. Acta* **2002**, 66, 3757.

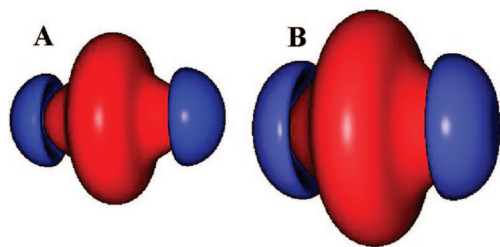


Figure 1. Electrostatic potential iso-value surfaces for (A) CO₂ (surface value ± 0.02 au) and (B) O₂ (surface value ± 0.002 au) computed at HF (CO₂) and UHF (O₂) levels using the 6-31G(d,p) basis set. Plots are on the same scale. Negative potential is plotted in blue, and positive potential in red.

site region and with the same atoms of reactant, intermediates, and products all similarly coordinated to the active-site Mg²⁺ complex. It is thought that this has restricted the “residue space” that Rubisco can sample evolutionarily by mutation while maintaining sufficient activity in all steps to constitute a viable enzyme.⁷ The consequence of this bottleneck is that although Rubisco catalyzes carboxylation and oxygenation of D-ribulose 1,5-bisphosphate (RuBP) in the ratio of 9:1 (green plants), forward processing of the oxygenation products results in the loss of organic carbon. The net outcome is that the oxygenation reaction reduces the photosynthetic efficiency of Rubisco by up to 50%.⁸ To sustain high rates of carbon assimilation, Nature’s solution in most photosynthetic organisms is to synthesize large amounts of Rubisco, making it the most abundant protein in plants⁹—constituting 30–50% of leaf proteins³—and, indeed, on earth. Alternatively, more complex mechanisms have been evolved, such as in C₄ grasses which synthesize large amounts of phosphoenolpyruvate carboxylase as part of a CO₂-concentrating mechanism to increase the pCO₂ at the active site and, hence, effectively reduce the extent of the wasteful oxygenation reaction.¹⁰

In view of its importance, Rubisco has been a prime target for genetic engineering aimed at improving the photosynthetic efficiency in plants with a view to increasing crop productivity¹¹ and in other photosynthetic organisms for applications in biomass or biofuel production.¹² Intensive experimental effort has been expended to provide a molecular description of the complete reaction mechanism, including the roles of active-site residues and the origins and fates of protons in the different steps, and the origin of the water molecule for the hydration step. However, while it is clear that the chemical definition of the gas-addition reactions and the subsequent hydration step, which irreversibly traps the CO₂ and O₂ adducts, is essential to underpin rational re-engineering, these steps are precisely those least tractable to experimental definition. X-ray crystallographic studies of complexes of Rubisco with different phosphorylated sugar ligands or metal ions have provided snapshots of several

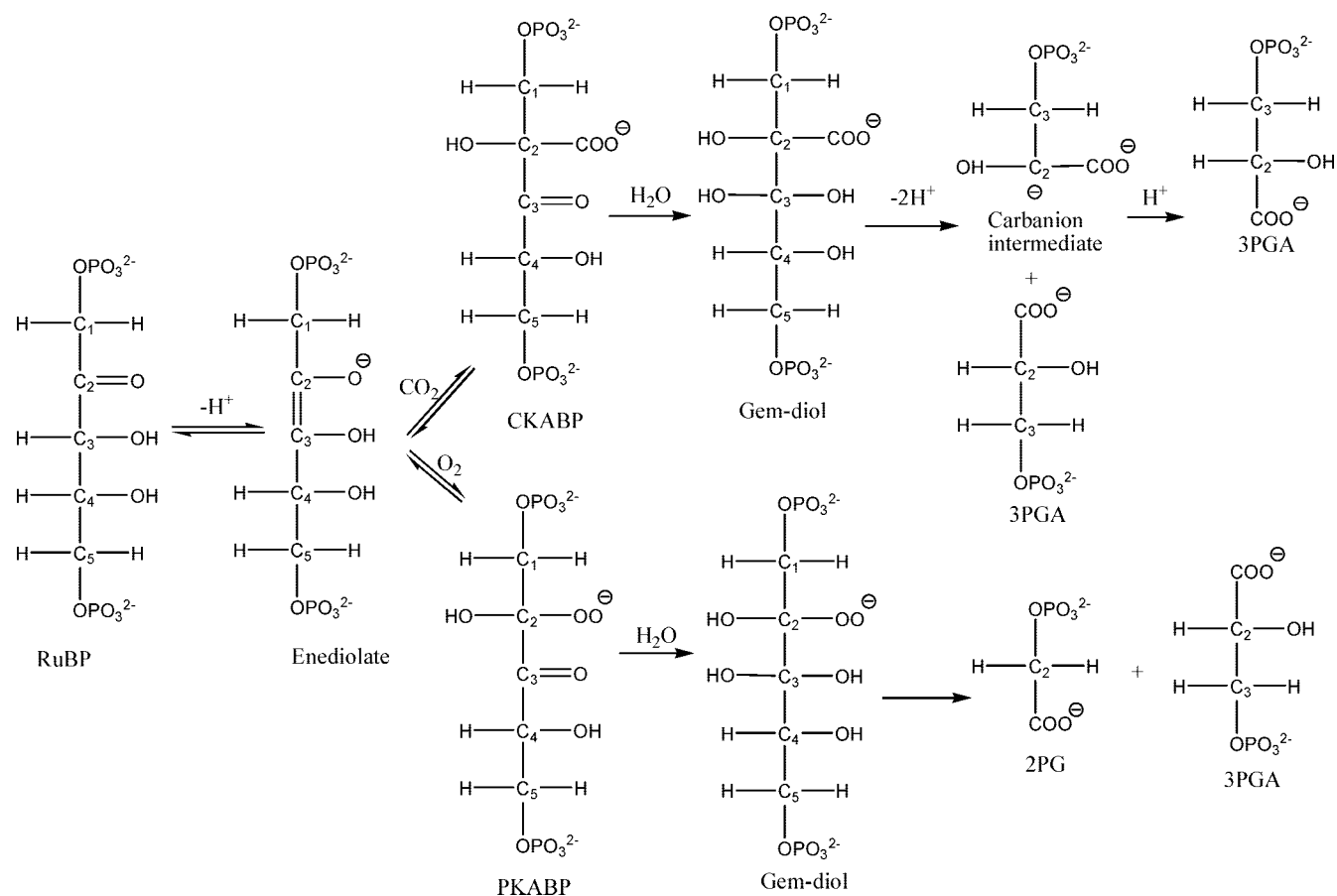
of the reaction steps catalyzed by Rubisco.^{13–17} Together with the results of mutational studies,^{18–30} these have helped to reveal the sequence of reactions that take place at the Rubisco active site (see Scheme 1) and have provided insights into its overall structure and the subtle changes in its conformation at different reaction stages. Also, putative roles for various active-site residues at different steps in the reaction mechanism have been deduced.

The holoenzyme of form I Rubisco, the form present in higher plants, algae, and cyanobacteria, is a hexadecamer of eight small and eight large subunits (LSUs in four dimers; see Figure 2) with one active site per LSU;¹³ the amino acid sequence numbers used in this article correspond to the numbering in spinach Rubisco. The enzyme is activated^{31,32} by carbamylation of a crucial residue LYS201 (KCX201), which then binds to a Mg²⁺ ion. The metal ion also coordinates with ASP203, GLU204, and three more water molecules to form an octahedral coordination shell. Subsequently, RuBP displaces two water molecules from the coordination sphere, binding the Mg²⁺ ion through the O2 and O3 oxygen atoms of RuBP. The enzyme remains in an open conformation with the active site accessible to solvent along the direction by which RuBP reaches the active site. Once RuBP is bound or, possibly, after the first step of enolization, loop 6 of the C-terminal α/β -barrel (TIM-barrel) domain extends over the active site occluding it from solvent. Figure 2 depicts the active site and key residues, in closed conformation. The solvent-excluded active site of Rubisco (i.e., it does not reopen at intermediate steps) catalyzes the sequence of five reaction steps shown in Scheme 1: enolization of RuBP, carboxylation/oxygenation, hydration, C2–C3 bond cleavage, and the stereospecific protonation of C2 (only for carboxylated product).¹⁴

Despite the now quite detailed definition of Rubisco’s molecular structure and mechanism, numerous site-directed and random mutation studies have reported uniformly unsuccessful results in engineering a better Rubisco, i.e., one with improved selectivity or kinetic activity (efficiency).^{3,4} Selectivity, $S_{C/O}$, also called the specificity factor, τ , is the ratio of the efficiencies

- (7) King, W. A.; Gready, J. E.; Andrews, T. J. *Biochemistry* **1998**, *37*, 15414–15422.
- (8) Lu, G. G.; Lindqvist, Y.; Schneider, G. *Proteins: Struct., Funct., Genet.* **1992**, *12*, 117–127.
- (9) Ellis, R. J. *Trends Biochem. Sci.* **1979**, *4*, 241–244.
- (10) Hatch, M. D. *Photosynth. Res.* **2002**, *73*, 251–256.
- (11) Parry, M. A. J.; Madgwick, P. J.; Carvalho, J. F. C.; Andralojc, P. J. *J. Agric. Sci.* **2007**, *145*, 31–43.
- (12) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484–489.

- (13) Andersson, I.; Taylor, T. C. *Arch. Biochem. Biophys.* **2003**, *414*, 130–40.
- (14) Taylor, T. C.; Andersson, I. *J. Mol. Biol.* **1997**, *265*, 432–44.
- (15) Taylor, T. C.; Andersson, I. *Biochemistry* **1997**, *36*, 4041–6.
- (16) Andersson, I. *J. Mol. Biol.* **1996**, *259*, 160–74.
- (17) Zhang, K. Y.; Cascio, D.; Eisenberg, D. *Protein Sci.* **1994**, *3*, 64–9.
- (18) Harpel, M. R.; Larimer, F. W.; Hartman, F. C. *Biochemistry* **2002**, *41*, 1390–7.
- (19) Harpel, M. R.; Larimer, F. W.; Hartman, F. C. *Protein Sci.* **1998**, *7*, 730–8.
- (20) Harpel, M. R.; Hartman, F. C. *Biochemistry* **1996**, *35*, 13865–70.
- (21) Larimer, F. W.; Harpel, M. R.; Hartman, F. C. *J. Biol. Chem.* **1994**, *269*, 11114–20.
- (22) Harpel, M. R.; Hartman, F. C. *Biochemistry* **1994**, *33*, 5553–61.
- (23) Harpel, M. R.; Larimer, F. W.; Hartman, F. C. *J. Biol. Chem.* **1991**, *266*, 24734–40.
- (24) Lee, E. H.; Harpel, M. R.; Chen, Y. R.; Hartman, F. C. *J. Biol. Chem.* **1993**, *268*, 26583–91.
- (25) Harpel, M. R.; Hartman, F. C. *J. Biol. Chem.* **1992**, *267*, 6475–8.
- (26) Soper, T. S.; Mural, R. J.; Larimer, F. W.; Lee, E. H.; Machanoff, R.; Hartman, F. C. *Protein Eng.* **1988**, *2*, 39–44.
- (27) Larson, E. M.; Larimer, F. W.; Hartman, F. C. *Biochemistry* **1995**, *34*, 4531–7.
- (28) Gutteridge, S.; Rhoades, D. F.; Herrmann, C. *J. Biol. Chem.* **1993**, *268*, 7818–24.
- (29) Hartman, F. C.; Soper, T. S.; Niyogi, S. K.; Mural, R. J.; Foote, R. S.; Mitra, S.; Lee, E. H.; Machanoff, R.; Larimer, F. W. *J. Biol. Chem.* **1987**, *262*, 3496–501.
- (30) Hartman, F. C.; Lee, E. H. *J. Biol. Chem.* **1989**, *264*, 11784–9.
- (31) Lorimer, G. H.; Mizziorko, H. M. *Biochemistry* **1980**, *19*, 5321–8.
- (32) Pierce, J.; Reddy, G. S. *Arch. Biochem. Biophys.* **1986**, *245*, 483–93.

Scheme 1. Sequence of Reactions Catalyzed by Rubisco^a

^a RuBP: D-ribulose 1,5-bisphosphate, CKABP: 2-carboxy-3-keto-D-arabinitol 1,5-bisphosphate, PKABP: 2-peroxo-3-keto-D-arabinitol 1,5-bisphosphate, 3PGA: 3-phospho-D-glycerate, 2PG: 2-phosphoglycolate.

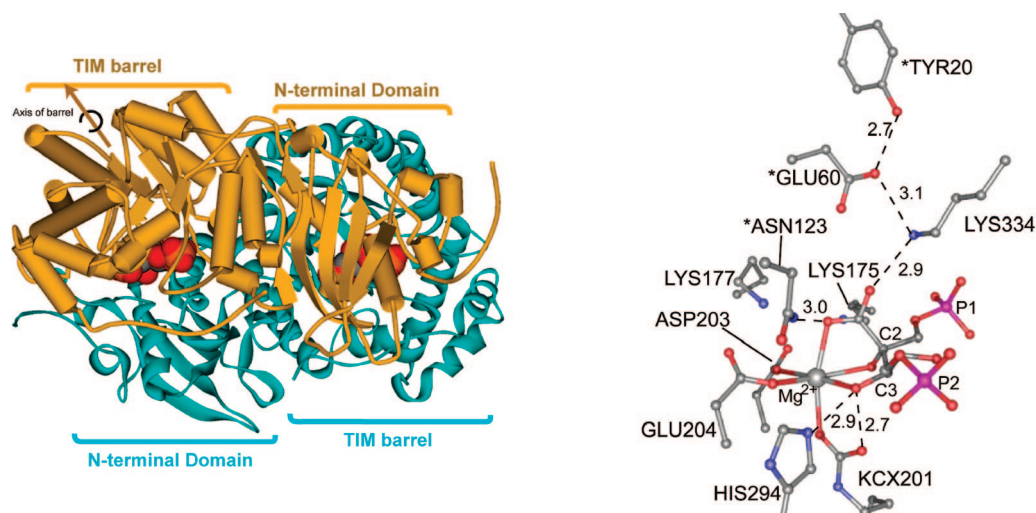


Figure 2. Active site of Rubisco plotted from the pdb structure 8ruc,¹⁶ (complex of Rubisco with Mg^{2+} and 2-carboxy-D-arabinitol 1,5-bisphosphate (2CABP)). The figure on left shows the dimer of large subunits (LSUs); one LSU (cyan) is displayed in ribbon form, and the other (mustard) as a secondary-structure cartoon. The active site is formed by the C-terminal TIM barrel of one LSU and the N-terminal domain of the partner LSU. The two 2CABP molecules, analogues of the carboxylated reaction intermediate, bound to the two active sites in a dimer are shown as vdW models. The Mg^{2+} ion at the active site coordinates to side chain groups of residues from the TIM barrel and oxygen atoms of 2CABP, while several residues from the N-terminal domain covering the active site interact with 2CABP. The figure on right shows the coordination sphere with all key residues and some important H bond distances. Residues with an asterisk in the label are from the N-terminal domain of the partner LSU of the dimer.

of fixations of CO_2 compared with O_2 under given assay conditions. Rubiscos from phylogenetically diverse organisms have substantially different values for the kinetic constants. Although the specificity factor, τ , and the maximum catalytic

rate of carboxylation per active site, k_{cat} , are generally observed to have an inverse relationship, Zhu et al.³³ have calculated that replacing C3-plant Rubiscos (τ of ~ 80) with those of high τ , such as the Rubisco in red algae (τ of ~ 140 – 160), could

increase carbon assimilation by more than 25%. However, Rubisco from the red algal species *Galdieria partita* failed to assemble in tobacco,³⁴ and as little is known about Rubisco assembly in red algae, creation of transgenic plants with active red-algal Rubiscos that would allow assessment of proposed better functionality of these high-specificity Rubiscos in higher plants is currently infeasible. Hence, the primary focus for improving Rubisco's efficiency is still the mutational approach. As indicated by our assessment of an evolutionary bottleneck above, and pointed out by others,³⁵ engineered enhancement of Rubisco specificity may require coordinated mutations at multiple sites.

These disappointing results differ from successful rational re-engineering programs for many other enzymes where single mutations, or in some cases multiple mutations, deduced straightforwardly from structural, and mechanistic data obtained from experiment, sometimes aided by computation, have proven successful in modifying substrate specificity or catalytic efficiency in predicted desired directions. Hence, it is useful to re-examine the current state of knowledge of Rubisco chemistry which has underpinned redesign attempts so far, to better understand its peculiarities and identify possible incorrect deductions which may be responsible for this impasse. An obvious problem is in the small substrate molecules. Enzymes recognize their substrates through steric features and electrostatic and hydrophobic interactions, but in addition to having a similar shape both CO₂ and O₂ have strikingly similar electrostatic patterns (see Figure 1) although there is an order of magnitude difference in potential between the two molecules. This provides a major challenge to evolve selectivity into the molecular machinery that channels the correct (CO₂) gaseous substrate into the active site and assists in formation of the correct reactive encounter complex with enolized RuBP substrate. Thus, even though mutations of loop 6 and the C-terminal domain residues, which wall the gas diffusion channel, influence both reactivity and selectivity,²⁸ no single mutation in this region produced a better Rubisco. A second issue is the catalytic chemistry. As noted, there are gaps in the knowledge that experiment has—and can—provide on the precise details for all the steps, and although several studies, including our own,^{7,36,37} have addressed some of the experimentally intractable issues using computational methods,^{38–40} all of these studies have theoretical or computational limitations. Fortunately, we are now able to overcome these limitations and in the current work redefine the mechanism and energetics in Rubisco catalysis, including the roles of all key active-site residues, using a large chemical model of the active site which had hitherto been computationally infeasible to study with *ab initio* methods (HF and DFT). This model includes fragments of all the amino acid residues likely to have a direct role in the catalysis. As background to our work and

the results to be presented, we summarize relevant current understanding of the roles of active-site residues deduced from crystallographic, mutational, and computational studies in the recent literature; more comprehensive reviews of mutational studies of Rubisco are given by Kellogg and Juliano,⁴¹ Parry et al.,³ and Andersson.⁴²

1.1. Role of KCX201. Enolization is the most studied step in Rubisco catalysis. It is now accepted in the literature that KCX201 is the base that initiates enolization by abstracting the C3 proton.^{7,16,43} However, the immediate fate of the abstracted proton is unclear. The proton is eventually exchanged with solvent and is not observed in the products.⁴⁴ Cleland et al.⁴³ envisaged multiple roles of KCX201 as a base: abstraction of the C3 proton, transfer of a proton from O3 to O2, and deprotonation of the gem-diol triggering C2–C3 scission. Thus, even though there is consensus on the role of KCX201 as the initial base, its role in subsequent reaction steps is in question.

1.2. Role of LYS175. The K175G mutant lacks carboxylase and enolization activity but is not impaired in the forward processing of the reaction intermediate 2-carboxy-3-ketoarabinitol 1,5-bisphosphate (2C3KABP), although the C2 fragment is turned into pyruvate instead of 3-phosphoglycerate (3PGA).^{29,45} Furthermore, replacement of LYS175 by a less basic aminoethylcysteine residue reduces k_{cat} 5-fold, demonstrating a correlation between catalytic turnover and the pK_a of the amino group at this position.⁴⁶ While these studies unambiguously define a key role for LYS175 in stereospecific protonation of the *aci*-acid intermediate, how LYS175 influences the enolization step is not clear from experiments. It has been suggested⁴³ that LYS175 is initially in its neutral state and receives a proton from O2 after enolization, thereby enabling KCX201 to abstract another proton for the reaction to proceed further. Alternatively, from their computational studies King et al.⁷ implicated LYS175 as a proton donor, finding that the protonated form favors formation of the enol. In their report of computational studies of later steps, Mauser et al.³⁶ speculate an additional role for LYS175 of abstracting back from O2 the proton donated during enolization, prior to the C2–C3 bond cleavage. Andersson et al.^{13,42} note that by H bonding with ASP203, LYS175 would prevent ASP203 from forming bidentate coordination with Mg²⁺, which would disrupt the charge-flow mechanism and the reaction itself.

1.3. Role of LYS334. Mutants of LYS334 lack carboxylation/oxygenation activity and do not bind 2-carboxyarabinitol 1,5-bisphosphate (2CABP) but do catalyze enolization of RuBP and forward processing of 2C3KABP.^{22,26,27,30} This suggests LYS334 stabilizes the adducts of carboxylation/oxygenation and the corresponding transition states. Gutteridge et al.²⁸ observed that the K334R mutant of *Synechococcus* Rubisco has dismal carboxylase activity, but that its oxygenase activity was reduced by only a factor of 2. This result demonstrates the potential for differentiating the chemistry for additions of the two gas substrates, although in this case not in the desired direction. Carboxylation may involve the protonation of CO₂ by LYS334, or at least the formation of a short H bond that would make

- (33) Zhu, X. G.; Portis, A. R.; Long, S. P. *Plant Cell Environ.* **2004**, *27*, 155–165.
- (34) Whitney, S. M.; Baldet, P.; Hudson, G. S.; Andrews, T. J. *Plant J.* **2001**, *26*, 535–547.
- (35) Ramage, R. T.; Read, B. A.; Tabita, F. R. *Arch. Biochem. Biophys.* **1998**, *349*, 81–8.
- (36) Mauser, H.; King, W. A.; Gready, J. E.; Andrews, T. J. *J. Am. Chem. Soc.* **2001**, *123*, 10821–10829.
- (37) Gready, J. E. In *Artificial Photosynthesis*; Collings, A., Critchley, C., Eds.; Wiley-VCH: Weinheim, 2005; pp 263–282.
- (38) Oliva, M.; Safont, V. S.; Andres, J.; Tapia, O. *J. Phys. Chem. A* **2001**, *105*, 4726–4736.
- (39) Oliva, M.; Safont, V. S.; Andres, J.; Tapia, O. *J. Phys. Chem. A* **2001**, *105*, 9243–9251.
- (40) Tapia, O.; Andres, J.; Safont, V. S. *J. Phys. Chem.* **1996**, *100*, 8543–50.

- (41) Kellogg, E. A.; Juliano, N. D. *Am. J. Bot.* **1997**, *84*, 413–428.
- (42) Andersson, I. *J. Exp. Bot.* **2008**, *59*, 1555–68.
- (43) Cleland, W. W.; Andrews, T. J.; Gutteridge, S.; Hartman, F. C.; Lorimer, G. H. *Chem. Rev.* **1998**, *98*, 549–561.
- (44) Fiedler, F.; Mullhofer, G.; Trebst, A.; Rose, I. A. *Eur. J. Biochem.* **1967**, *1*, 395–9.
- (45) Lorimer, G. H.; Hartman, F. C. *J. Biol. Chem.* **1988**, *263*, 6468–71.
- (46) Smith, H. B.; Hartman, F. C. *J. Biol. Chem.* **1988**, *263*, 4921–5.

C(CO₂) more electrophilic. Siegbahn et al.⁴⁷ have shown that the presence of a particular protonated HIS residue near the O₂ molecule in glucose oxidase greatly enhances its electron affinity. In a similar fashion, LYS or ARG residues at position 334 in Rubisco might facilitate oxygenation.

1.4. Role of HIS294. Taylor and Andersson,¹⁴ and later King et al.⁷ and Mauser et al.³⁶ based on their computational studies, suggested that HIS294 acts as a proton acceptor from O₃. Taylor and Andersson also proposed HIS294 as a likely candidate, together with KCX201, for the base that deprotonates the gem-diol to initiate the C2–C3 bond cleavage.¹⁵ As H294N and H294Q mutants show very poor enolization and carboxylation activities, but much stronger binding to 2CABP, Harpel et al.¹⁹ suggested an indirect role for HIS294 in modulating the basicity of KCX201 by H bonding, or by being a part of a proton-relay network that connects KCX201 with solvent. Additionally, as the H294N mutant predominantly catalyzes the decarboxylation of the 2C3KABP intermediate rather than its hydration, they suggested that HIS294 deprotonates the gem-diol after hydration but still kept open the possibility of HIS294 playing a role in hydration itself.

Apart from the residues discussed above, several other active-site residues are critical for the catalysis. Even residues beyond the first coordination shell of the active site are known to influence the activity and specificity of Rubisco, further confirming the intricate relation between Rubisco structure and function and reinforcing the difficulty in designing a better Rubisco.⁴⁸

In summary, while some roles for KCX201 and LYS334 are known with confidence, there are many uncertainties in roles played by other key residues, which when mutated affect various partial reactions. Scheme 1 summarizes the mechanism advanced by experimentalists for both the carboxylation and oxygenation reaction sequences at the Rubisco active site: a consensus view of residue roles in each step is given by Cleland et al.⁴³ Each step involves acid–base chemistry of the enzyme with the substrate or its intermediates, but the crux of the current uncertainty is which residue is acting at which step, and how, and the effective ionization state and H bonding partners of the substrate or its intermediates. We note other computational work by Tapia, Andres and co-workers^{38–40} on the carboxylase and oxygenase reactions of RuBP in which they have adopted a totally different “enzyme-free” approach to modeling the overall Rubisco reaction shown in Scheme 1. Their reaction-step models are entirely intramolecular with no enzyme-assisted acid–base chemistry involved. In their mechanism all bond breaking and formation occurs by homolytic bond cleavages involving pairing and unpairing of electrons, with the enzyme role restricted to molding the substrate and subsequent intermediates into appropriate conformations. We evaluate these results in comparison with our results in the Discussion section.

In the work we present here we have explored the enzyme-assisted description of acid–base catalysis in Rubisco using a consistent set of high-level *ab initio* calculations and a staged set of models, including an active-site fragment model large enough to include all key residues. Our goal is to provide precise and reliable mechanistic predictions of the chemistry at each step.

2. Computational Methods and Models

2.1. Model Construction and Geometry Optimization. We have explored the use of several active-site fragment models (FMs) of increasing complexity to identify the minimum size model that can realistically represent the active site and produce reliable energetics using *ab initio* quantum chemical methods. Starting geometries representing the first stationary point (the ene-diolate) of the reaction sequence at the active site were constructed for the FMs using the coordinates of crystal structure 1rxo¹⁴ (activated spinach Rubisco complexed with Ca²⁺ and RuBP). All atoms except those required in the FM were deleted and the atomic valences saturated by adding hydrogen atoms. The central coordinating metal ion was replaced with Mg²⁺. Further changes in the ligands were carried out by changing the internal coordinates and addition/removal of atoms using the program Gaussview.⁴⁹ For FMs that include LYS334 in the model, the starting structure for the β -keto intermediate was constructed from the crystal structure 8ruc¹⁶ (activated spinach Rubisco complexed with Mg²⁺ and 2CABP). Calculations were carried out using the Gaussian 03⁵⁰ program at the DFT level using the B3LYP functional. The basis set 6-31G(d,p) was used for all models except **FM15**, for which the 6-31G(d) basis was used. During optimization of the stationary points of the structures built from the crystal-structure coordinates, the heavy atom of each residue farthest from Mg²⁺ in the model was initially constrained. This was followed by a complete unconstrained optimization. Geometries for all other stationary points were constructed by editing the z-matrices of previous/next stationary points in the reaction sequence. All stationary points were located by a final unconstrained optimization. Intrinsic reaction coordinate (IRC) calculations were performed for some first-order saddle points in both forward and reverse directions. In such cases, the calculation was continued only until the geometry closely resembled the expected stationary point to minimize computational cost. The resultant geometries were later optimized completely to obtain minima connecting the saddle point in the IRC. Some Rubisco active-site fragment computational studies reported in the literature were performed with constrained optimization of the reaction species, freezing one or more atoms of each residue fragment.^{7,38,39} However, we have noted substantial changes in the conformation and orientation of the substrate and intermediate atoms during the sequence of reactions, suggesting the possibility for different reaction species to have different interactions, or strengths of interaction, with the active-site residues and intramolecularly. Constrained minimization of the geometries of reaction species cannot account for such changes in interaction during the course of reaction. Despite the well-known difficulties in optimizations of floppy truncated fragments with dangling atoms, we achieved complete optimizations for all the reaction species discussed here. Careful screening of the geometries during the optimization helps in avoiding structures with unphysical interactions of dangling residue tails that could lead to artifacts in the energetics. This involves comparison of some relevant inter-residue distances in optimized geometries with the corresponding distances from the crystal structure¹⁶ to determine the extent of deviation of positions of coordination-sphere residue atoms in the optimized structure relative to their positions in the active site in the crystal complex of 2CABP and Rubisco. Figure S2 in the Supporting Information shows a plot of structural deviation for all the geometries optimized with **FM20** and the cutoff values used to screen out invalid structures. Interatomic distances (11) were used to compare the optimized structures with the active site from crystal structure 8RUC. For all the structures included in the potential energy surface (PES) the deviation from the crystal structure is less than 0.6 Å

(47) Prabhakar, R.; Siegbahn, P. E. M.; Minaev, B. F.; Agren, H. *J. Phys. Chem. B* **2002**, *106*, 3742–3750.

(48) Du, Y. C.; Peddi, S. R.; Spreitzer, R. J. *J. Biol. Chem.* **2003**, *278*, 49401–5.

(49) Dennington, R., II; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W. L.; Gilliland, R. *GaussView*, version 3.0; Semichem Inc.: Shawnee Mission, KS, 2003.

(50) Frisch, M. J. et al. *Gaussian 03*, revision C.02; Gaussian Inc.: Wallingford, CT, 2004.

for 75% of the reported distances (see pages S3 and S4 in the Supporting Information for details). This is a good indication that all the fragment-model optimized structures fit well within the enzyme active site.

Computational modeling of enzyme reactions requires consideration of a sufficient region of the active site and its enzyme environment at a sufficiently high QM model to represent the chemistry reliably. What constitutes “sufficient” and what QM model is appropriate depends on both the enzyme and active-site structure and chemistry. In the case of Rubisco, the catalytic mechanism consists of not one but a series of five reaction steps and with a relatively large number of active-site residues directly involved in these reactions. This unusually large reactive “core” needs to be represented by a relatively high *ab initio* QM level. It is now common to complement the core QM model with a lower-level treatment representing the perturbation of the core by the surrounding enzyme. There are two types of methodologies to do this: an implicit model where the enzyme environment surrounding the active-site fragment is substituted by a dielectric continuum and an explicit model where a large part of the enzyme is modeled at the MM level (hybrid QM/MM methods). Although the latter is a more accurate representation of the enzyme system, it is difficult to apply in studies of reaction mechanisms and especially so for a system as complex as Rubisco, as our trial work has shown.³⁷ As the active site of Rubisco is at the interface of two LSUs, a minimalist model would require 2 LSUs and a small subunit constituting ~20 000 atoms, excluding solvent. This in itself implies a prohibitive computational cost as our previous work has shown that it is necessary to apply a QM/MM treatment within an MD simulation,⁵¹ or some other method for conformational sampling;⁵² enzyme structures merely travel between local minima uncoupled to the reaction coordinate if stationary states are simply subject to geometry optimization, introducing random errors into energy differences. Also, as there is a large number of charged residues taking an active part in one or more reaction steps and it is essential to take into account alternative possible positions of protons and resultant H bond interaction networks when considering the larger enzyme environment, it is currently computationally impossible to perform a hybrid QM/MM calculation on Rubisco using a Hamiltonian for the QM region adequate to compute reliable energetics. Ramos and Fernandes⁵³ addressed this problem in studying enzyme reactions in a recent review with benchmarks of different computational strategies. Using dielectric continuum and QM/MM calculations they showed that if the size of the fragment model is large enough to shield the reaction center from the environment, then the energetics are not significantly affected by the inclusion of enzyme environment in the calculations. To validate our fragment model, in this work we performed single-point energy calculations using a polarized continuum model, IEFPCM as implemented in Gaussian 03,⁵⁴ on the final geometries to test the effect of incorporation of the enzyme environment into the energetics. We used an empirical dielectric constant (ϵ) of 4, as there is consensus in the literature that this value gives a generally good agreement with experimental results and accounts for the average effect of both the protein and buried water molecules.⁵⁵

2.2. Structural Basis for Fragment Models. Fragment models (shown in Figure 3) were constructed systematically starting with a minimalist model (**FM1**), similar to that used in our earlier study.⁷ **FM1** contains the central Mg^{2+} ion, a two-carbon (2C) mimic for the substrate and $\text{H}_2\text{N}-\text{COO}^-$ in place of the carbamylated lysine (KCX201). The remaining three coordination positions, including

for ASP203 and GLU204, were filled with water molecules. Later, to better model the effect of coordination of the carboxyl oxygen atoms of ASP203 and GLU204, the two water molecules in **FM1** were replaced with bidentate ligands ethan-1,2-dione (**FM2**) and propan-1,3-dione (**FM4**). Based on our assessment from earlier studies^{7,36} that there is substantial neutralization of the ASP203 and GLU204 negative charges by the nearby positively charged LYS175 and LYS177 residues, neutral ligands were used in these coordination positions. However, during the course of this study a substantial increase in computational power became available, allowing us to increase the complexity of the model to include fragments of the key residues, ASP203, GLU204, LYS175, and LYS177, and also a four-carbon (4C) mimic for the substrate. This **FM15** model has CH_3COO^- ions for ASP203 and GLU204 and CH_3NH_3^+ ions for LYS175 and LYS177. Finally, to allow definition of the mechanistic roles of HIS294 and LYS334, fragment model **FM16** was generated. Models **FM1a** and **FM10** were used to simulate a different protonation state of the enediol intermediate of substrate, while **FM17** and **FM18** were included to assess the effect of LYS334 on the energetics of CO_2 addition.

As these background studies showed **FM16** was suitable for representing the redefined roles for KCX201 and HIS294, it was used to trace the entire reaction path of the carboxylase activity at the Rubisco active site. Results for all the other models are restricted to the energetics of the CO_2 addition step. However, due to the absence of charge delocalization on **LYS334** in the enediol geometry for **FM16**, the energies of the carboxylated adduct and subsequent intermediates are artificially low relative to enediol. Furthermore, for the **FM16** model transition states for the carboxylation and C2-protonation steps could not be located. To overcome this defect in **FM16**, an even larger model, **FM20**, was generated. **FM20** includes an $-\text{OPO}(\text{OH})_2$ group attached to the C1 carbon of the substrate in addition to the atoms included in **FM16** (see Figure 3 and Table 1). This additional functional group mimics the P1 phosphate (P_1) of the substrate, but the charge has been neutralized with two protons to prevent artifactual energetics for **FM20**. This choice is justified by the fact that in the enzyme the P1 phosphate group interacts with several positively polarized groups that form the phosphate binding site. The entire reaction path obtained initially using **FM16** was retraced using **FM20**.

To simplify understanding of the electronic implications of the FMs, they can broadly be classified into three template models, A, B, and C, as shown in Figure 4, each representing a different protonation state or form of the metal complex with respect to KCX201 or the substrate. Table 1 summarizes the different characteristics of the FMs, shown schematically in Figure 3.

2.3. Key Mechanistic Aspects Considered.

2.3.1. Fate of Mg-Coordinated Water Molecule. Prior to binding of RuBP, three of the Mg-coordination sites are filled with water molecules. Two of these are displaced by RuBP upon binding. There is no direct evidence from experiments, or clues from computational calculations so far reported, as to what happens to the third water molecule during the course of reaction. However, it has been shown experimentally that CO_2 is not required for the enolization activity⁵⁶ and that neither CO_2 nor O_2 forms a Michaelis–Menten complex with Rubisco.⁵⁷ This suggests that CO_2 does not necessarily displace the third water molecule from the coordination shell prior to carboxylation. This has also been shown computationally by our previous work.³⁶ However, in the crystal structures of the Rubisco•2CABP complex, the C2-carboxylate group is coordinated to Mg^{2+} , which appears to have led to the general belief that the third water molecule from the coordination shell is displaced by the newly formed carboxylate group⁵⁸ either during the carboxylation reaction or immediately after the CO_2

(51) Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comput. Chem.* **2002**, *23*, 1314–1322.

(52) Cummins, P. L.; Rostov, I. V.; Gready, J. E. *J. Chem. Theory Comput.* **2007**, *3*, 1203–1211.

(53) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689–98.

(54) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(55) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Am. Chem. Soc.* **2007**, *129*, 1378–1385.

(56) Gutteridge, S. *Biochim. Biophys. Acta* **1990**, *1015*, 1–14.

(57) Hartman, F. C.; Harpel, M. R. *Annu. Rev. Biochem.* **1994**, *63*, 197–234.

(58) Schneider, G.; Lindqvist, Y.; Branden, C. I. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 119–143.

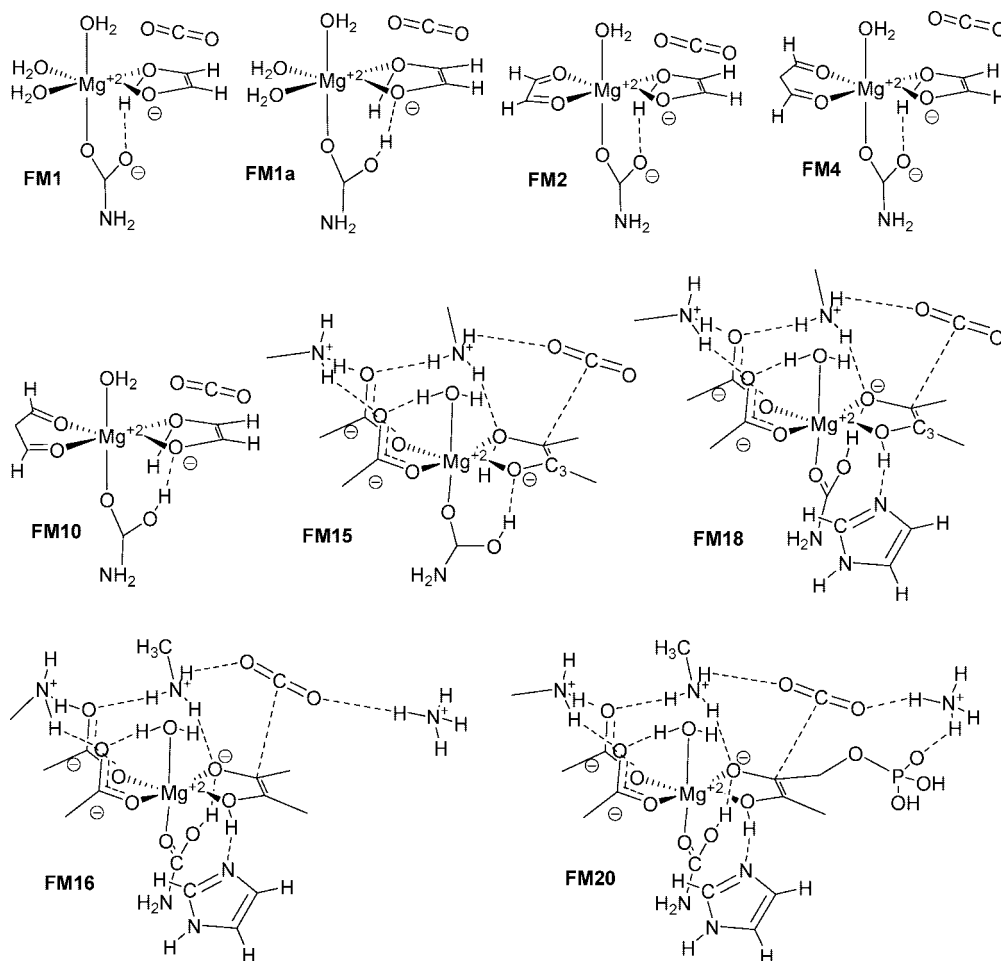


Figure 3. Schematic of different fragment models used, shown in increasing order of complexity. The structures correspond to the coordination sphere with enediolate bound CO₂. **FM17**, which differs from **FM16** only by the absence of a proton on the LYS334 mimic, is not shown here.

Table 1. Comparison of Details of the Fragment Models Used Showing Fragment Size, Net Charge on Fragment, Substrate-Molecule Truncation, And Mimics or Truncations^a for ASP203, GLU204, LYS175, LYS177, HIS294, and LYS334 Residues

	FM1	FM1a	FM2	FM4	FM10	FM15	FM18	FM17	FM16	FM20 ^e
charge	0	1	0	0	1	1	1	1	2	2
<i>N</i> _{atoms} ^b	26	27	26	29	30	57	66	70	71	77
<i>N</i> _{prot} ^c	1	2	1	1	2	2	2	2	2	2
substrate ^d	2C	2C	2C	2C	2C	4C	4C	4C	4C	4C
mimic _{203/204}	W	W	ED	PD	PD	Ac	Ac	Ac	Ac	Ac
mimic _{175/177}	—	—	—	—	—	MeA	MeA	MeA	MeA	MeA
mimic ₂₉₄	—	—	—	—	—	—	Imd	Imd	Imd	Imd
mimic ₃₃₄	—	—	—	—	—	—	—	NH ₃	NH ₄ ⁺	NH ₄ ⁺

^a Mimic_{residue number}: structural mimic for residues corresponding to ASP203, GLU204, LYS175, LYS177, HIS294, and LYS334 used in the fragment model; W: two water molecules coordinated to Mg; ED: ethan-1,2-dione in bidentate coordination to Mg; PD: propan-1,3-dione in bidentate coordination to Mg; Ac: two acetate ions coordinated to Mg; MeA: CH₃NH₃⁺ ion; Imd: imidazole. ^b *N*_{atoms}: total number of atoms. ^c *N*_{prot}: total number of protons on O2 and O3, either bonded are H bonded by a base. ^d 2C: two-carbon substrate model; 4C: four-carbon substrate model. ^e In addition to the tabulated structural features, **FM20** has an -OPO(OH)₂ group attached to the C1 carbon of the substrate.

addition. In view of this assumption, the reaction pathway in our previous computational study contained a local minimum energy geometry for the β -keto intermediate from a model without a water molecule in the Mg-coordination sphere, but instead with a water molecule placed in the vicinity of the C3-carbon atom to provide a water molecule for hydration.³⁶ However, this created a “dis-

continuity” in our total modeling for the reaction, as the preceding local minimum in the reaction pathway (after the enolization step) was from a model with a Mg-coordinated water molecule and no water molecule near the C3-carbon atom.⁷

The current work removes this discontinuity in the reaction path. We have set aside previous unsupported assumptions in the literature regarding the fate of the Mg-coordinated water molecule and the source of the water molecule for hydration and explored *de novo* the simple possibility that the Mg-coordinated water molecule remains in the coordination sphere even after the gas-addition step and is available as a source for the hydration step. As an initial assessment of this possibility, the carboxylated adduct (β -keto intermediate) was optimized to two different structures using **FM16**: one with the water molecule remaining coordinated to Mg while H bonded to the newly formed C2-carboxylate group (m1), and the other with the water molecule replaced by the C2-carboxylate group directly coordinated to Mg (m1') (see Figure S1 of the Supporting Information). In m1', the displaced water molecule arches over the C2-carboxylate group of the substrate and also forms H bonds with GLU204 and LYS177. The energy of m1' relative to m1 (energy change for displacing the water molecule from Mg-coordination) is 4.5 kcal/mol. These calculations were repeated with structural modifications to mimic the four-carbon fragment of the analogue of the β -keto intermediate, 2CABP: this requires conversion of the C3-keto group to a hydroxyl group and adjustment of the FM charge. For this complex, the energy of m2' (water molecule displaced from Mg coordination) relative to m2 (water molecule coordinated to Mg) is 1.6 kcal/mol. Optimized geometries of all four structures are provided in Figure S1 of the Supporting Information.

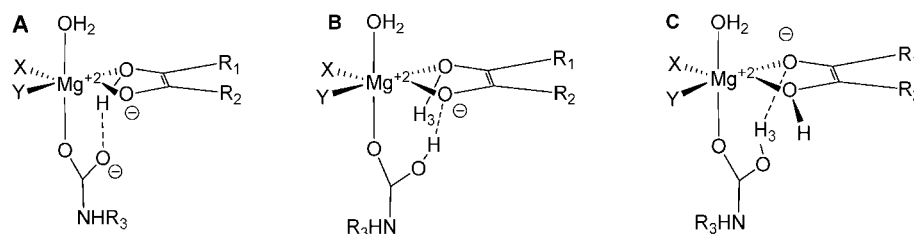


Figure 4. Three templates, **A**, **B**, and **C**, for the fragment models, representing different proton compositions and H bond configurations in the active site. Fragment models **FM1**, **FM2**, and **FM4** are based on template **A**; models **FM1a**, **FM10**, and **FM15** on template **B**; and models **FM16**, **FM17**, **FM18**, and **FM20** on template **C**. In template **A**, H3 is considered to have been transferred to the proton sink and, thus, has an additional negative charge.

The estimated rearrangement energy (i.e., 1.6 kcal/mol) is clearly small with a likely error of a few kcal/mol, so displacement of the water molecule as observed in the crystal structure of the Rubisco•2CABP inhibitor complex is compatible with these results. However, the results suggest this rearrangement is energetically less likely (4.5 kcal/mol) for the β -keto intermediate of the reaction (2C3KABP) and would need to be accompanied by a substantial local reorganization in the structure, including the breaking and formation of Mg-coordinate bonds. Thus, our suggestion that the chemical reaction may proceed beyond carboxylation without dissociation of the Mg-coordinated water molecule is energetically and kinetically plausible. There is also strong experimental evidence to support this proposition from the results of Pierce et al.⁵⁹ which showed that the binding of 2CABP takes place in two steps, a fast first step and a slower second step that takes place at the rate of 0.04 s^{-1} and accounts for the tighter binding of 2CABP to Rubisco.⁵⁹ While Pierce et al.⁵⁹ attributed this slow step to the “isomerization” of the enzyme complex, our mechanism suggests this slow process could involve displacement of the third Mg-coordinated water molecule from the coordination sphere by the C2-carboxylate group of 2CABP; this involves substantial local reorganization of atoms and would result in tighter binding of 2CABP.

Retaining the third water molecule in the Mg-coordination sphere after carboxylation leads to the interesting possibility that this water molecule acts as the water of hydration. Andersson¹⁶ has reported that there are no water molecules in the vicinity of the C3 center of 2CABP in the 1.6 Å resolution crystal structure of activated spinach Rubisco complexed with 2CABP. The nearest water molecule to C3 is 5.8 Å away, H bonded to HIS294; she suggested that this water could participate in the hydration of 2C3KABP, and in the absence of an alternate hypothesis, this suggestion has become accepted in the literature. On the other hand, the Mg-coordinated water molecule is only ~ 3.7 Å from C3 in the 2C3KABP complex and is, thus, ideally positioned for nucleophilic attack on the C3-carbonyl group. This hydration mechanism would also avoid the need for the water molecule to approach C3 in a direction *anti* to the C2-carboxylate group.

2.3.2. Protonation State of the Enediol(ate) Intermediate.

Whether the first intermediate is in the enediol or enediolate form is unclear. In the intramolecular mechanism advocated by Oliva et al.,³⁸ there is no charge separation for any reaction species throughout the pathway; thus, the RuBP tautomer is in the enediol form. According to the widely accepted enzyme-assisted catalysis mechanism involving acid–base chemistry, an enediolate is generated.⁴³ However, there are differing views on how the enediolate is generated and on the proton composition and H bond configuration of the active-site region after the enolization step. These depend on assumptions on the residues participating in enolization and subsequent reaction steps. Residue KCX201 is widely believed to play additional roles in proton movement after the initial RuBP tautomerization. In the proposed consensus mechanism of Cleland et al.,⁴³ KCX201 abstracts H3 and transfers it to a proton sink near

Table 2. Activation (E_a) and Reaction Energy (ΔE) in kcal/mol for the Carboxylation Reaction Computed at the B3LYP/6-31G(d,p) Level with 10 Different Fragment Models within Three Template Categories

	template A			template B			template C			
	FM1	FM2	FM4	FM1a	FM10	FM15 ^a	FM18	FM17	FM16	FM20
E_a^b	3.9	11.1	4.4	—	12.5	6.9	15.4	13.6	—	5.9
ΔE	1.1	7.0	0.3	0.0	12.4	−12.9	11.3	7.5	−31.6	−7.7

^a **FM15** computations at the B3LYP/6-31G(d) level. ^b For **FM1a** and **FM16**, the transition state could not be located.

the binding site of the P1 phosphate group of RuBP and then acts again as a base to shuttle the O3 proton to O2. This would leave the carboxylate group of KCX201 H bonded to H—O2 and a formal negative charge on O3. This scenario is represented by template **A** in Figure 4. Alternatively, in view of the lack of evidence to support this proton relay mechanism to a proton sink, KCX201 might transfer the H3 proton to O2 and then abstract the O3 proton and remain H bonded to O3 (represented as template **B** in Figure 4). In a variation of this mechanism, King et al.⁷ have suggested that KCX201 only tautomerizes RuBP, while HIS294 might generate the enediolate anion. In this mechanism, H3 remains with the carbamate group (being more basic than O2) and is H bonded to the, formally, negatively charged O2, while the O3 proton is H bonded to HIS294 (template **C**). *A priori* discrimination of these three acid–base mechanisms is not possible as they involve subtle differences in effective pK_a 's and positioning of the relevant groups within the active-site environment. Also, positions of protons within H bonds is largely a moot point, depending greatly on their strength. Hence, we have investigated all three models of proton composition and H bond configuration in the present work.

2.3.3. Test of Enediolate-Template Models: Dependence on Carboxylation Energetics. The energetics of the carboxylation-reaction step were calculated using 10 different FMs, 3 belonging to each of the templates A and B and 4 FMs belonging to template C. Activation and reaction energies for all 10 FMs are given in Table 2. **FM1**, **FM2**, and **FM4** belong to template A, with improvement in representation of the coordination sphere from **FM1** to **FM4**. Similarly, **FM1a**, **FM10**, and **FM15** belong to template B. Within and between these two sets of FMs, the energetics are very variable, with significant differences in the magnitude of the activation barrier and degree of stabilization of product. This highlights their sensitivity to the mimics used to represent the Mg-coordinated residues and demonstrates that a realistic representation of the coordination sphere and key residues is essential to model the chemistry with fragment calculations and produce meaningful energetics. This problem has been well appreciated previously but could not be addressed because of computational limitations for an active site of Rubisco's complexity.

We can draw some useful conclusions from the results. **FM4**, based on template A, and **FM10**, based on template B, have structurally equivalent Mg coordination but differ in the protonation state and H bond configuration of KCX201. Both activation and reaction energies for **FM10** are much higher than those for **FM4**, and, in fact, the reaction energy is infeasible suggesting effectively

(59) Pierce, J.; Andrews, T. J.; Lorimer, G. H. *J. Biol. Chem.* **1986**, *261*, 10248–56.

Table 3. Relevant Geometrical Parameters for the Reaction Species Computed with the **FM20** Model at the B3LYP/6-31G(d,p) Level^a

	I	II [†]	III	IV [†]	V	VI [†]	VII	VIII [†]	IX
R_{C2-C3} (Å)	1.36	1.40	1.54	1.58	1.59	2.40	4.76	4.32	4.38
R_{C3-O3} (Å)	1.38	1.32	1.24	1.25	1.43	1.27	1.22	1.22	1.22
R_{C3-Ow} (Å) ^b	3.55	3.72	3.77	2.06	1.41	1.36	1.33	1.34	1.34
$R_{O3...H}$ (Å)	1.06	1.52	1.83	1.59	1.01	1.67	1.94	3.07	3.08
$R_{H...N\epsilon}$ (Å) ^c	1.52	1.09	1.03	1.06	1.05	1.05	1.05	1.10	1.09
$R_{OS2...H-LYS334}$ (Å)	2.18	1.86	1.50	1.63	1.70	1.50	1.70	1.55	1.86
D_{C2} (deg) ^d	174.9	160.6	123.7	116.5	121.5	147.7	178.2	128.9	122.7
D_{C3} (deg) ^d	179.1	177.1	178.2	153.6	117.6	151.2	179.6	180.0	180.0

^a TS structures are marked by †. Structures for all species are shown in Figure 5. ^b Ow is the oxygen atom derived from the Mg-coordinated water molecule. ^c N ϵ is the ϵ nitrogen atom of HIS294, and the reported distance is between N ϵ and the nearest H. ^d For species **I** to **V**, D_{C2} is the angle between planes C1–C2–C3 and O2–C2–C3, and D_{C3} is the angle between planes C4–C3–C2 and O3–C3–C2. For species **VI** to **IX**, D_{C2} is the angle between planes Cs–C2–C1 and O2–C2–C1 and D_{C3} is the angle between planes Ow–C3–C4 and O3–C3–C4.

no stabilization of product compared with TS (Table 2). This suggests that the template B model where the proton from O3 is abstracted by KCX201 while O2 is protonated (with the former H3 proton) is unlikely. Furthermore, as will be discussed in the Results for the **FM20** model, there is another reason why the template B model is unlikely if other residues in the active site are taken into consideration: transfer of the H3 proton from KCX201 to O2 is energetically unfavorable as it would hamper H bond donation to O2 from LYS175. This leaves the possibilities of templates A and C. The protonation state in template A relies on an uncharacterized means for transferring H3 to a proton sink, suggested to be near the P1 phosphate.⁴³ On the other hand, the protonation state for template C requires no such assumption and fits well with the crystallographically determined arrangement of amino acid residues at the active site, particularly in providing an active role for HIS294 which has been largely neglected in the experimental and computational literature. Hence, in the following we focus on the template C models, **FM16**, **FM17**, **FM18**, and **FM20**, in which HIS294 is explicitly modeled by an imidazole ring.

3. Results

After an initial assessment of the quality of the FMs, the complete reaction sequence starting from an enediolate intermediate has been studied using models **FM16** and **FM20**. **FM20** is a larger model generated to redress some artifacts in the energetics computed with **FM16** as well as difficulties in locating some transition states. Structures and energetics of postcarboxylation reaction intermediates are provided only for **FM20**. The relative energies and activation barriers are reported without and with zero-point energy (ZPE) correction for results from **FM20**. Cartesian coordinates of all the reported structures are provided in the Supporting Information. Table 3 summarizes key geometrical parameters for reaction species computed with **FM20**, while their structures are shown in Figure 5. Several alternate stationary points were located for some of the reaction species; these arise primarily from differences in the H bonding network or configuration of Mg-coordination. As we performed a final unconstrained minimization of all the reported structures, artificial interactions not present in the enzyme were noticed in some of the optimized structures. Such invalid structures were screened out based on the extent of structural deviation from the crystal geometry 8RUC¹⁶ (see Figure S2 of the Supporting Information). Validated lowest energy structures for each reaction species are denoted in roman numbers from **I** to **IX**, while alphabetical suffixes are used to denote alternate high energy structures (**IVa** and **Va**).

3.1. Addition of CO₂. The starting point of the reaction studied here is the vdW complex of CO₂ with the metal-bound enediolate intermediate. As in our previous work,³⁶ we obtained a loosely bound vdW complex of CO₂ with the coordination sphere using small FMs. However, as shown for structure **I** in

Figure 5, for **FM20** (and for the structure with **FM16**) the CO₂ is firmly anchored to LYS334 and LYS175 through strong H bonds and oriented appropriately for the addition reaction. In **I**, the Cs...C2 distance is 4.3 Å. Inclusion of HIS294 in the model also reveals its possible role in the mechanism. In structure **I**, HIS294 is poised to act as a base; it forms a short H bond (1.5 Å) with H–O3 through its N ϵ atom. It is also interesting to note that O2 is formally unprotonated in this reactive-complex structure, its proton being closer to KCX201. The protonation state of O2 has been a key point of debate in the literature on the protonation state of the enediol(ate) intermediate.^{36,43} It has been assumed that O2 must be protonated, as the expectation on chemical grounds is that an unprotonated O2 would direct the electrophilic addition of CO₂ to the C3 carbon, which is unwanted and is not observed experimentally. Conversely, deprotonation of O3 should assist carboxylation at C2. However, a closer look at the H bond network in **I** shows that O2 also has two short H bonds with LYS175 (1.7 Å) and KCX201 (1.6 Å), which we argue substantially nullifies its formal negative charge. This argument is supported by analysis of the atomic charges. The electrostatic potential (ESP) fitted atomic charges computed using the ChelpG⁶⁰ option within the program G03⁵⁰ are –0.33 and –0.46 for atoms O2 and O3, respectively. This clearly establishes that despite O2 appearing “unprotonated”, more negative charge resides on O3 and the proton configuration in geometry **I** could effectively lead to C2 carboxylation. We note, for completeness, that results from simplistic Mulliken partitioning of atomic charges give charges of –0.77 for O2 and –0.67 for O3, i.e., both quite negative, indicating that arguments based solely on formal charges deduced from molecular structures are misleading.

The guess structure for the carboxylation TS was obtained by scanning the potential energy surface (PES) for decreasing Cs...C2 distance. Subsequent optimization led to an appropriate saddle point of order one (**II**). As with our previous work,³⁶ repeated attempts with several different starting geometries and PES scans failed to yield a valid TS for concerted C2-carboxylation and C3-hydroxylation. The O3...H distance in structure **II** shows that deprotonation of O3 is not in concert with CO₂ addition; rather O3 is effectively deprotonated ($R_{O3...H}$ = 1.52 Å) prior to CO₂ addition. In fact, this hydrogen now bonded to N ϵ of HIS294 has a negligible displacement vector for the imaginary frequency characterizing the carboxylation TS. The negative charge buildup in the newly forming C2-carboxylate group is stabilized by an H bond with LYS334 ($R_{OS2...H-LYS334}$ = 1.86 Å), which is also evidenced by a longer Cs–O_{S2} bond length relative to that for Cs–O_{S1} in both TS

(60) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.

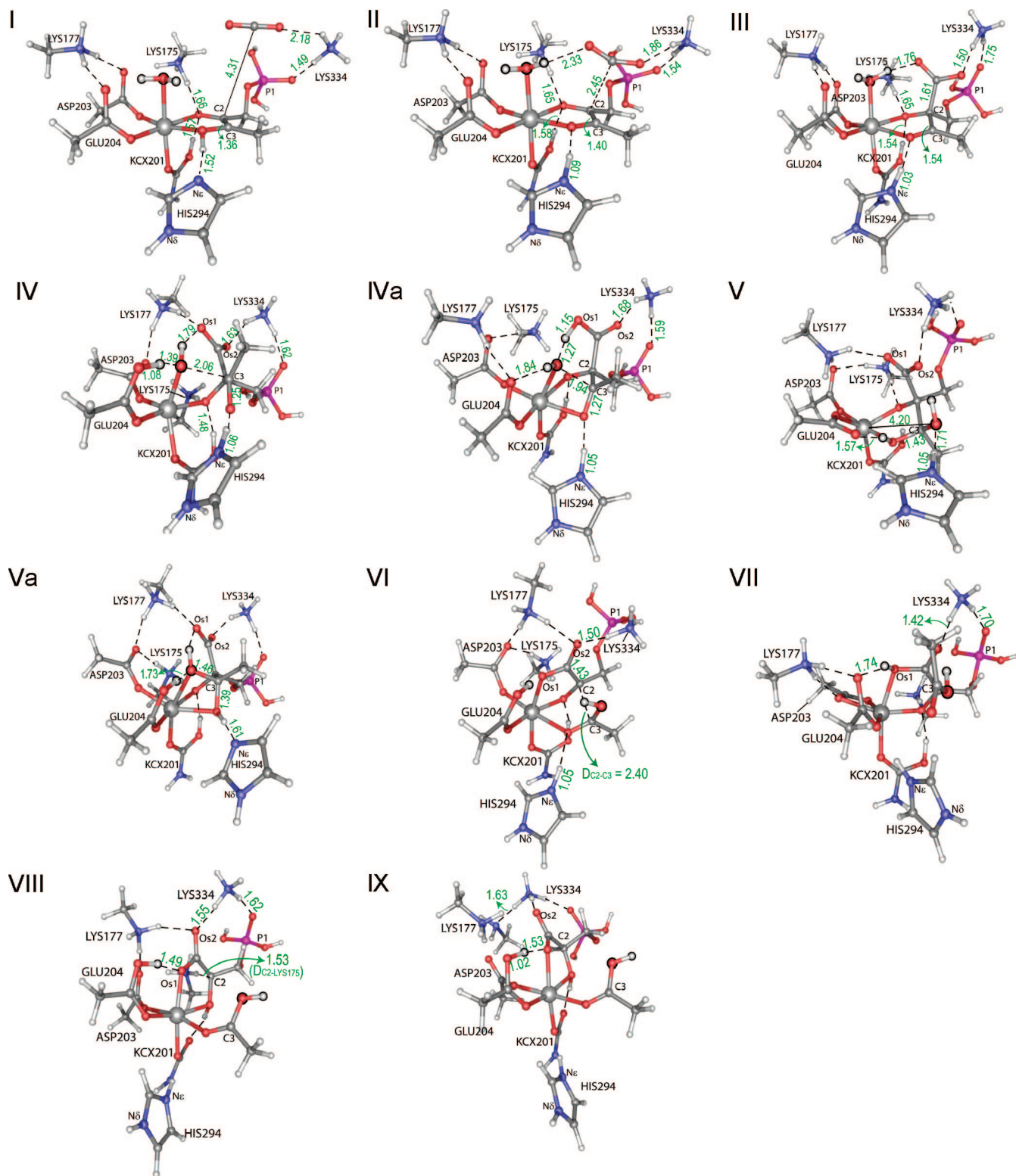


Figure 5. Geometries of local minima along the reaction pathway from enediolate to the end products in the Rubisco carboxylation reaction, obtained using **FM20** and optimization at the B3LYP/6-31G(d,p) level. The structures were drawn using Cartesian coordinates of the stationary points. Atoms of $\text{H}_2\text{O}[\text{Mg}]$ are highlighted by black circles. H bonds and partial bonds are shown as dashed lines. Relevant distances for each structure are shown to demonstrate changing interactions between component species as the reaction proceeds.

(II) and product (III). This is in agreement with the fact that LYS334 is absolutely essential for C2-carboxylation to take place. The extent of stabilization of reactive species II and III by LYS334, however, depends on the magnitude of the partial positive charge on the terminal amino group of LYS334, which in the crystal structure is seen to also form a salt bridge with

GLU60 of the neighboring LSU. In **FM20** a strong H bond between LYS334 and the P1 phosphate group, and the absence of phosphate-binding functional features, partially compensates for the lack of this salt bridge interaction with GLU60. This fact is brought out by comparing the reaction energies for the CO_2 addition step in Table 2. The relative stabilization of the

β -keto intermediate, 2C3KABP (**III**), is -31.6 and 7.5 kcal/mol for the **FM16** and **FM17** models, respectively. LYS334 is represented by an ammonium ion with a unit positive charge in **FM16** and by a neutral ammonia molecule in **FM17**. Comparing these numbers with the reaction energy for **FM20** (-7.7 kcal/mol) puts in perspective the effect of the partial charge on LYS334 in **FM20**. When the ZPE correction is included the reaction energy is increased to -4.8 kcal/mol for **FM20**. In the carboxylated product (**III**), the O3 proton is completely abstracted by N ϵ of HIS294, evidenced by elongation of the H...O3 distance (to 1.83 Å). The O $_{S1}$ atom of the newly formed carboxylate group forms strong H bonds with H $_2$ O[Mg] and H(LYS175), while the O $_{S2}$ atom H bonds with LYS334. The electron flow in the reaction can be tracked by following the C $_5$ –C2, C2–C3, C3–O3, H...O3, and H...N ϵ bond distances for structures **I**, **II**, and **III** (see Table 1 and Figure 5). Structures **I**, **II**, and **III** clearly bring out the role played by HIS294 as a base and LYS334 as an acid in the carboxylation step.

At this point it is useful to review the protonation states of key active-site residues. Both KCX201, the Lewis base in enolization, and HIS294 (N ϵ), the Lewis base in carboxylation, are protonated. KCX201 and LYS175 H bond with O2, and HIS294 H bonds to O3. Both O2 and O3 have formal negative charges, but H bonds neutralize the charges to some extent as analyzed above. Thus, protonated states of both KCX201 and HIS294 are critical to the stability of 2C3KABP. As detailed in the Discussion, we hypothesize that unsuitable protonation states of these (and possibly other) active-site residues would result in slower binding of 2C3KABP if externally fed to Rubisco, and this would account for the slow catalytic turnover reported for such Rubisco experiments.⁵⁸

3.2. Hydration at C3 Center. By performing a PES scan on structure **III** by decreasing the C3–O $_w$ distance, we obtained reasonable guess geometries for the hydration TS from which to start geometry optimization. It appears that a base is required to activate the Mg-coordinated water molecule for nucleophilic attack on C3. We obtained two starting geometries for the hydration TS: one with GLU204 acting as a base (starting geometry for **IV**) and the other with the C2-carboxylate as a base (starting geometry for **IVa**). The guess geometries led to two first-order saddle points. **IV** is the lowest energy structure with the displacement vector of imaginary frequency corresponding to C3–O $_w$ bond formation and simultaneous abstraction of a proton (H $_{w1}$) by GLU204 from H $_2$ O[Mg]. In this geometry, O $_w$ remains coordinated to Mg (2.02 Å) while forming a partial bond with C3 (2.06 Å). The TS is stabilized by several strong interactions. The H $_{w1}$ atom of the water molecule is positioned 1.39 Å from O $_w$ and stabilized by effective bonding to the free oxygen of GLU204 (distance 1.08 Å). Part of the partial negative charge developed on O $_w$ is further stabilized by stronger coordination to Mg, evidenced by a decrease in the coordinate bond distance by 0.07 Å relative to the distance in **III**. In the alternate structure **IVa**, which is higher in energy by 1.7 (1.0) kcal/mol, the displacement vector for the imaginary frequency corresponds to C3–O $_w$ bond formation and simultaneous abstraction of H $_{w1}$ from H $_2$ O[Mg] by the C2-carboxylate group. In both structures **IV** and **IVa**, it is evident that HIS294 contributes to the reaction by polarizing the C3–O3 double bond as manifested by $R_{N\epsilon-H...O3}$ distances of 1.59 and 1.64 Å, respectively. In addition to **IV** being the lowest energy structure, optimized geometries of subsequent reaction species also indicate that GLU204 is the preferred base in the PES.

Thus, the geometrical parameters of **IV** reveal new roles for GLU204 (as a base) and HIS294 (as an acid) in coordinating the hydration step. There are only minor changes in the H bond lengths involving LYS175, KCX201, and O2.

The lowest energy structure for the gem-diol intermediate (**V**) has a slightly different protonation state from **IV**. H $_{w1}$, abstracted by GLU204 in **IV**, is donated to O3, but H $_{w1}$ remains H bonded to GLU204. Protonated N ϵ of HIS294 now H bonds with O $_w$. The geometry for the gem-diol intermediate that connects to **IV** has also been located in the PES (**Va** in Figure 5) but is higher in energy than **V** by more than 9 kcal/mol. Similarly, the intermediate resulting from structure **IVa** is ~ 15 kcal/mol higher in energy than **V**. Alternate structures (12) have been located for the gem-diol intermediate; their coordinates are provided in the Supporting Information. It should be noted that structure **V** can be achieved in the reaction through isomerization/reorganization of **Va** which succeeds **IV** in the reaction sequence. The C3–O $_w$ distance for bond formation decreases from 3.77 Å in **III**, through 2.06 Å in **IV**, to 1.43 Å in **V**. The gradual pyramidalization of C3 in moving from structure **III** through **IV** to **V** is also apparent from the torsion angle D_{C3} (angle between planes C4–C3–C2 and O3–C3–C2), as shown in Table 3. HIS294 is seen to actively assist the completion of the hydration by back-donation of the proton to O3: the H...O3 distance changes from 1.83 Å (**III**) through 1.59 Å (**IV**) to 1.01 Å (**V**).

A particularly interesting feature of the hydration reaction revealed by our new mechanism is the protonation of GLU204. To date, no specific *chemical* role for this residue has been suggested in the literature on Rubisco catalysis. This protonation produces other structural changes in the H bonded network of the active site by inducing significant reorientations of ASP203 and LYS177. In many of the alternate structures for the gemdiol intermediate such as **Va**, LYS177 does not form bridging H bonds to ASP203 and GLU204 as in structures **I**, **III**, and **IV** (Figure 5). Instead, it interacts only with ASP203 and moves farther from the coordination shell. However, this movement of LYS177 would be restricted to some extent in the actual enzyme environment. Although for this fragment model the activation energy for the hydration step appears somewhat high at 21.8 (20.6) kcal/mol, it is less than that for the (next) bond-scission step, which is known to be rate-limiting in Rubisco catalysis (see Figure 6). Although the activation energy for the reverse reaction is slightly lower (21.2 (17.9) kcal/mol), it should be noted that O $_w$ has now moved farther from Mg (4.2 Å) and is inappropriately oriented for recoordination with Mg, effectively preventing the reverse reaction. This supports the experimental finding that the carboxylated intermediate exists in the enzyme active site predominantly as the C3-gemdiol.⁶¹ Our mechanism thus suggests that the reverse reaction is hindered due to structural repositioning and that the forward reaction has a relatively high activation barrier, leading to a significant half-life for the C3-gemdiol form.

Structures **IV** and **V** demonstrate the feasibility of use of H $_2$ O[Mg] as the source of the water molecule for hydration, whereas both the intramolecular mechanism of Oliva et al.³⁸ and other enzyme-assisted mechanisms discussed in the literature have ignored this possibility and assume a different water molecule approaching C3 from beneath (in the orientation shown in Figure 2) as the source.

(61) Lorimer, G.; Andrews, T. J.; Pierce, J.; Schloss, J. V. *Philos. Trans. R. Soc. London, Ser. B* **1986**, *313*, 397–407.

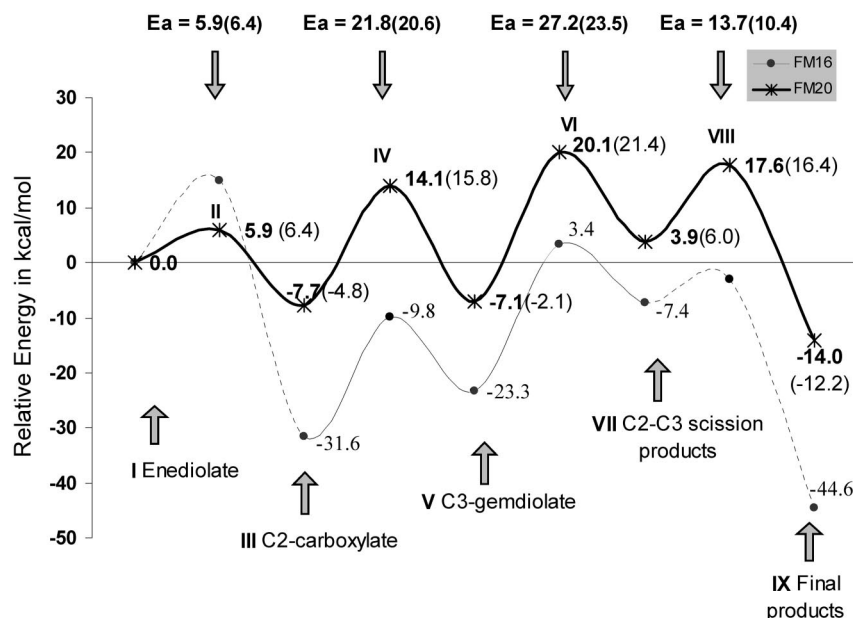


Figure 6. Potential energy surface for the carboxylation and subsequent reactions in the carboxylase pathway for **FM20** (thick line) and **FM16** (thin line), plotted using the local minima points optimized at the B3LYP/6-31G(d,p) level (shown in Figure 5 for **FM20**). The numbers in brackets are the energies with zero-point energy correction. For **FM16**, structures **II** and **VII** could not be located, and hence the relevant sections of the energy curve are shown by dashed lines.

3.3. C2–C3 Bond Scission. We obtained a starting geometry for the transition state for C2–C3 bond scission by a constrained optimization of a guess with an elongated C2–C3 bond. Full geometry optimization produced a saddle point of order one (**VI**), with the eigenvector of the imaginary frequency corresponding to breaking of the bond between C2 and C3. The TS occurs at a C2–C3 distance of 2.4 Å. Examination of interatomic distances shows convincingly that this C2–C3 bond scission is, again, facilitated by HIS294. In the TS (**VI**), the H–O3 proton is within a strong H bond but mostly abstracted by HIS294 ($N_{\epsilon} \cdots HO3$ and $H \cdots O3$ distances of 1.05 and 1.67 Å; Table 3); this configuration imparts partial double bond character to the C3–O3 bond (1.27 Å). Interestingly, compared with the proton network in **V**, GLU204 regains H_{w1} . The partial negative charge developing on C2 is distributed to O_{S1} and O_{S2} as evidenced by a partial double bond between C2 and C_S (1.43 Å). This electron flow is facilitated by H bonds $O_{S1} \cdots H_{w1}$ (bonded to GLU204) and of $O_{S2} \cdots H(LYS334)$. At this stage, the proton of LYS175 is 2.44 Å from C2. Due to the increasing C2–C3 bond distance and the loss of a proton from O3 that imparts double bond character to the C3–O3 bond, the coordination of the C3 fragment with Mg weakens ($R_{Mg \cdots O3} = 2.39$ Å). This allows the free oxygen atom of the C2-carboxylate (O_{S1}) to coordinate with Mg.

Moving forward along the reaction coordinate from the TS involves complete deprotonation of O3 by HIS294 facilitating the C2–C3 bond break and eventually leading to the *aci*-acid intermediate and 3PGA, **VII**. During this stage, the deprotonation is specifically at O3, leading to the formation of the complete C3 fragment, which moves farther from Mg coordination. In structure **VII**, the C3 fragment bears no significant net charge (sum of ESP-derived atomic charges of atoms in C3 fragment is -0.04), whereas the C2 fragment has a net partial negative charge of -0.57 . As in the TS **VI**, the charge on C2 is delocalized by further shortening of the C2– C_S bond (1.36 Å), while O_{S1} and O_{S2} are both effectively protonated leading to an enetriol-like entity. This C2-fragment entity is similar in

structure to the transition state structures 8a, 9a, and 8b reported by Mauser et al.³⁶ and structure B reported by Oliva et al.⁶² Structure **VII** is 11.0 (8.1) kcal/mol higher in energy than the gem-diol intermediate (**V**), which can be attributed to the charge separation in **VII**. However, the corresponding energy difference reported by Mauser et al.³⁶ is more than 28 kcal/mol, emphasizing the extent of stabilization of intermediate **VII** in the model used here (**FM20**) by residues LYS175, LYS334, and GLU204 that are absent in FMs used by Mauser et al.³⁶ During this reaction step, LYS177 moves back to its original position forming an H bonded bridge between GLU204 and ASP203. The bond-scission step has an activation barrier of 27.2 (23.5) kcal/mol (Figure 6), the highest in the reaction sequence; this is consistent with our earlier computational work and conclusions from experiment. In **VII**, the C2 and C3 atoms are 4.76 Å apart, reducing the likelihood of the reverse reaction, while the proton suggested in the literature⁴³ to complete the forward reaction (H-LYS175) is only 2.4 Å from C2.

3.4. C2 Protonation. As in the work reported by Mauser et al.,³⁶ we were initially unable to find a transition state structure that would connect structures **VII** and **IX** using **FM16**. However, with the inclusion of P₁1 in **FM20** it was possible to locate an appropriate TS (**VIII**) for the transfer of a proton from LYS175 to C2. The guess geometry was generated by scanning the PES for decreasing the $C2 \cdots H-LYS175$ distance starting from structure **VII**. In structure **VIII**, the proton is 1.53 Å from C2 and 1.23 Å from LYS175. The displacement vector for the imaginary frequency of the TS corresponds to the transfer of a proton from LYS175 to C2 and recoordination of the C3 fragment with Mg. C2 undergoes substantial pyramidalization ($D_{C2} = 128.9^\circ$) accompanied by an increase in C2– C_S bond length to 1.46 Å. This bond elongation enables release of protons from both O_{S1} and O_{S2} to GLU204 and LYS334, respectively. The decrease in negative charge on the C2 fragment and its

(62) Oliva, M.; Safont, V. S.; Andres, J.; Tapia, O. *Chem. Phys. Lett.* **2001**, *340*, 391–399.

effect on coordinate bonds to Mg may drive the recoordination of the C3 fragment with Mg. In the case of **FM16**, when the proton of LYS175 was moved slightly closer to C2 and the geometry reoptimized, it spontaneously transferred to C2 without any energy barrier despite repeated attempts from several starting geometries (results not shown). This result, in conjunction with the fact that **FM16** and **FM20** differ only in the presence of P₁l, implies a possible role for P₁l in C2-protonation. In structures **VII** and **VIII** LYS175 is 2.25 Å from one of the oxygen atoms of the P₁l group. Interestingly, in the crystal structure of activated spinach Rubisco in complex with 3-phosphoglycerate (PDB code 1rus), LYS175 is 3.8 Å from an oxygen atom of the P₁l group with appropriate orientation for H bond interaction. This interaction is absent in crystal structures of complexes of Rubisco with 2CABP (PDB code 8ruc) and RuBP (1rxo). This indicates possible catalytic involvement of P₁l in the last reaction step. The energy barrier for C2-protonation in the current scheme is a modest 13.7 (10.4) kcal/mol.

The geometry for final products is obtained by transferring the proton from LYS175 completely to C2 followed by optimization (**IX**). This results in some reorganization of the coordination sphere. Having lost a proton to C2, LYS175 moves away and forms a strong H bond with one of the H atoms of LYS334, which is still H bonded to O_{S2} and P₁l. LYS334 occupies a spatial position in between LYS175 and P₁l and, hence, LYS175 is positioned farther from P₁l than in the reaction species **VII** and **VIII**. In the enzyme this configuration could be affected by the presence of the GLU60...LYS334 salt bridge. The energy of the final products relative to the enediolate intermediate **I** is −14.0 (−12.2) kcal/mol.

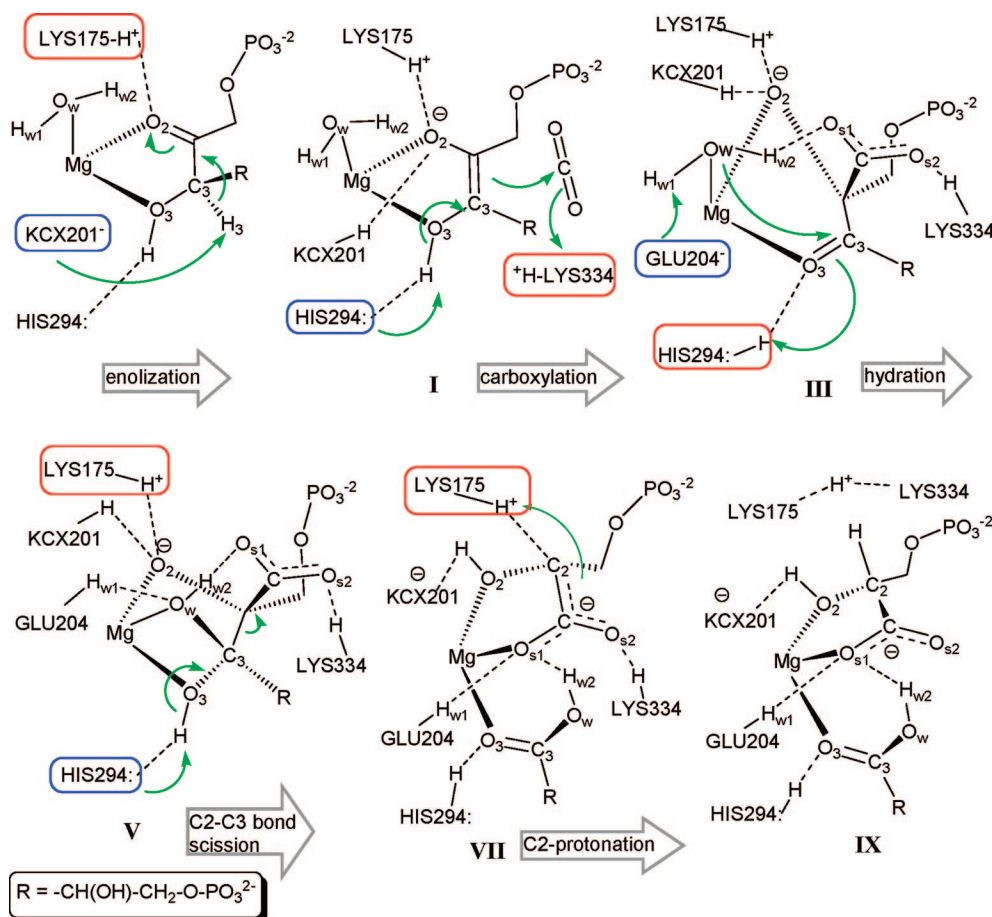
3.4. Energies with IEFPCM. Dielectric continuum calculations were carried out for all the structures on the reaction pathway for **FM20** and for the reaction species involved in the CO₂ addition step for **FM10**. For **FM10**, the activation energy for the carboxylation reaction is reduced by 6.3 kcal/mol relative to the gas-phase computed barrier, while the reaction energy is reduced by 7.8 kcal/mol. This large difference between the energetics computed in the gas phase and in solution suggests that the charge centers in the **FM10** model are unshielded. On the other hand, for **FM20** all the relative energies, except the final products, were shifted by only 0.2 to 1.6 kcal/mol. Comparison of the IEFPCM and gas-phase energies are provided in Table S1 of the Supporting Information. The IEFPCM-computed (and gas-phase) activation barriers are 7.1 (5.9), 20.0 (21.8), 25.2 (27.2), and 14.3 (13.8) kcal/mol, respectively, for the carboxylation, hydration, C2–C3 bond-scission, and C2-protonation steps. Thus, the barriers are elevated slightly for the carboxylation and C2-protonation steps and slightly reduced for the hydration and C2–C3 bond-scission steps. There is no significant difference in the trend in the energetics from inclusion of the enzyme environment as a dielectric continuum. The C2–C3 bond-scission step is still the slowest step with a barrier of 25.2 kcal/mol.

4. Discussion

4.1. Assessment of Reliability of Computational Model FM20. Investigation of the CO₂ addition step with several different FMs enabled us to evaluate the effect of the protonation state and H bond configuration, and the ligands used to fill the coordination sphere, on the energetics of the reaction, and to determine a minimum model composition within computational constraints. We found both factors have a major effect on the

energetics (see Table 2). For example, **FM1**, **FM2**, and **FM4** represent the same protonation state but differ by improving the representation of the ligands that mimic ASP203 and GLU204 in going from **FM1** to **FM4**. The carboxylation energetics (reaction energy and activation barrier) for these FMs does not show any definitive trend. Similarly, the pairs **FM1**/**FM1a** and **FM4**/**FM10** differ only in the protonation states of the enediolate and KCX201, but the change in reaction energy between the two pairs displays opposite trends. However, for larger FMs the changes in energetics between FMs accompanying structural change (see **FM16**, **FM17**, **FM18**, and **FM20** in Table 2) are more readily interpretable and show how a decision as to whether the model is reliable can be reached. Both the reaction energy and the activation barrier decrease when a proton donor is added to stabilize the C2-carboxylate group in the product. This demonstrates the stability of the FM to further extension; i.e., that minor change in the composition of the fragment will produce changes in the energetics in predictable directions. However, **FM16** has some shortcomings, namely its charge (+2) and the absence of residues from the second shell that interact strongly with LYS334. This latter deficiency was most apparent for calculation of the carboxylation TS and 2C3KABP intermediate. It was not possible to locate the carboxylation TS with **FM16**, while 2C3KABP is overstabilized relative to the preceding intermediate. This shortcoming is rectified in **FM20** by the inclusion of the P₁l group, which provides a satisfactory representation of the charge on LYS334. The neutralization of the P₁l group in **FM20** to avoid further artificial interactions is justified by consideration of its environment in the enzyme, where interactions with several phosphate-binding groups reduce the effective charge on P₁l. By building fragment models of incremental size, we have, step-by-step, incorporated all active-site features that have a major effect on the energy and mechanism of the carboxylase reaction sequence. Inclusion of a second shell of residues around the coordination sphere and the outer enzyme environment can be expected to cause second-order effects only on the energetics of the reaction and not change the mechanism we have proposed. Hence, although the magnitudes of the energies from the large **FM20** still need to be considered as estimates only, the trend in relative energetics described by this model can be considered reliable. Comparison of energies for **FM16** and **FM20** (see Figure 6) support this contention. Although the energetics of the carboxylation reaction for **FM16** are flawed due to the artificial charge localization in LYS334 (rectified in **FM20**), the trends in energetics for the postcarboxylation reactions are similar for **FM16** and **FM20**. Thus, considering the computational impediments to employing an even larger fragment model, we used **FM20** with some confidence to study the complete reaction pathway. ZPE corrections were applied to the energies of reaction species optimized with **FM20**, but this does not alter the trend in energetics of the reaction sequence.

4.2. Key Features of the New Rubisco Mechanism. Based on the computational results, we propose a novel mechanism for Rubisco carboxylase activity. This is summarized in Scheme 2. A pair of amino acid residues, a Lewis base and a Lewis acid, has been identified for each reaction step: enolization, KCX201 (base) and LYS175 (acid); carboxylation, HIS294 (base) and LYS334 (acid); hydration, GLU204 (base) and HIS294 (acid); and C2–C3 bond scission, HIS294 (base) and LYS175 (acid). Table 4 provides a comparison of our new mechanism with the intramolecular mechanism of Oliva et al.³⁹ and the previous consensus enzyme-assisted mechanism in the

Scheme 2. Proposed Reaction Mechanism for the Conversion of RuBP to Two Molecules of 3PGA at the Rubisco Active Site^a

^a The red and blue boxes highlight residues acting as a Lewis acid and a Lewis base, respectively, in the indicated reaction step. In the optimized structure of **VII**, the negative charge on the C2 fragment is substantially delocalized from C2 to O2, O_{S1}, and O_{S2} as shown by ESP-derived charges.

literature.⁴³ Key features and points of difference of the present mechanism are as follows:

1. The H₂O[Mg] molecule is not displaced from Mg-coordination by CO₂ during carboxylation; the water molecule, in fact, assists in binding CO₂ to the active site and contributes to the stability of the carboxylated product and the corresponding TS.
2. Contrary to the common belief that O3, and not O2, of substrate must be deprotonated in the enediol intermediate to direct CO₂ addition to C2 exclusively, in our mechanism O2 remains deprotonated. Our ESP-derived atomic-charge analysis shows that contrary to arguments based on simplistic deductions from the apparent structure, O3 is more negative than O2 in the enediolate **I**. Also, as LYS334 is also H bonded to the P1-phosphate group of the substrate in the enzyme, its interaction with the incoming CO₂ molecule and stabilization of the TS would not be possible if C3 carboxylation was to take place.
3. KCX201 has a direct role only in the initial enolization reaction, and it remains in the protonated state (with abstracted H3 atom) after enolization.
4. KCX201 and LYS175 play roles in inhibiting C3-carboxylation by partially neutralizing the negative charge on O2.
5. HIS294 plays several roles in the multistep catalysis of Rubisco. It shuttles the proton between its Nε atom and O3 of substrate, modulating the C3–O3 bond length appropriately during the different steps of the reaction.

6. GLU204 activates the H₂O[Mg] molecule for hydration by abstracting its proton. We note that both carboxylation and hydration take place on the same face of the enediolate intermediate.
7. The charge on the *aci*-acid intermediate is stabilized by LYS175 and LYS334.
8. LYS175 ensures stereospecific protonation of the C2 carbon to yield the final products. LYS334 facilitates this proton transfer by sharing a proton with LYS175.

The current mechanism predicts C2–C3 bond scission to be the rate-limiting step, consistent with the accepted view in Rubisco catalysis. Hydration and the corresponding reverse reaction have similar energy barriers; with ZPE correction the barrier for the reverse reaction is lower than the hydration step by 2.7 kcal/mol. However, as discussed in the Results, spatial separation of Mg and O_w in the C3-gemdiol would effectively inhibit the reverse reaction. For similar structural reasons, the current mechanism predicts C2–C3 bond scission to be essentially irreversible even though the reverse reaction has a lower activation barrier.

4.3. Specific Roles for Residues. In addition to defining a new mechanism, the present study has allowed us to assign specific roles to key amino acid residues in the enzyme catalysis. These roles are consistent with known experimental findings based on mutagenesis studies, as summarized below.

4.3.1. KCX201. As noted, reports in the literature assume a wide catalytic role for this residue in several steps. However, there is no experimental evidence to suggest a specific role for

Table 4. Comparison of Features of Our New Catalytic Mechanism with Those in the Literature

	intramolecular mechanism ^a	enzyme-assisted mechanism ^b	our proposed mechanism
enolization	H3 is directly transferred to O2 through a twisted TS. KCX201 assists only by H bonding with O3, and thereby creating a twisted enediol	KCX201 transfers H3 to O2. LYS175 possibly assists by polarizing C2 carbonyl group	LYS175 polarizes C2–O2 bond. HIS294 binds to H–O3 keeping it away from KCX201. KCX201 abstracts H3
CO ₂ addition	Occurs by pericyclic homolytic mechanism, involving one bond of CO ₂ and C2–C3–O3–H moiety in homolytic electron movement. During reaction, H–O3 is transferred to O atom of CO ₂ . KCX201 and HIS294 stabilize twisted form of enediol	O3 is deprotonated by either KCX201 or HIS294 prior to CO ₂ addition. Carboxylated intermediate stabilized by LYS334	HIS294 deprotonates O3. LYS334 stabilizes C2-carboxylate. Charge on O2 is neutralized by two short H bonds with LYS175 and KCX201
hydration	Occurs by concerted addition of H ₂ O to carbonyl bond via a four-centered TS, resulting in gem-diol. No external groups involved in reaction	H ₂ O attacks C3 from a direction <i>anti</i> to C2-carboxylate. KCX201 activates H ₂ O by abstracting a proton, and transferring it to O3	H ₂ O[Mg] attacks C3, assisted by C2-carboxylate and abstraction of proton by GLU204. HIS294 reprotonates O3
C2–C3 bond cleavage	H atom from C3-gem-diol group is transferred to either O2 or C2-carboxyl group, resulting in either hypervalent O2 or both O atoms of C2-carboxylate protonated. The transferred proton is then passed on to C2, with an inversion of stereochemistry	C2–C3 bond scission is coupled with a proton transfer from O _w to O2. LYS175 donates a proton to C2	HIS294 triggers C2–C3 bond scission by abstracting O3 proton. KCX201 donates proton to O2. LYS175 and LYS334 stabilize partial negative charge on C2 fragment

^a Reference 39. ^b Reference 43.

this residue in any of the reaction steps, although it is known that Mg²⁺ binds to the active site only after LYS201 has been activated by carbamylation. It was deduced to be the primary base in enolization based on structural analysis, which systematically ruled out all other possibilities, and on assessment that its pK_a is suitable for it to abstract H3.⁷ The current study proposes a catalytic role for this residue only in the initial enolization of RuBP. In the subsequent reaction steps, it acts to reduce the charge on O2 through H bonding. The abstracted H3 proton is held by the carbamate group right up to the C2–C3 bond-breaking reaction step.

4.3.2. LYS175. We found that LYS175 stabilizes the initial enediolate intermediate, but by forming a short H bond with O2 and not by donating a proton as reported by King et al.⁷ Retention of the proton by LYS175 even after the enolization step in the current mechanism avoids the need to reprotonate LYS175, by unknown means,⁴² for it to participate in the stereospecific protonation of the C2 carbon in the last reaction step. As shown by the geometries in Figure 5, LYS175, together with KCX201, forms short H bonds with O2 in all structures from **III** to **VI**. The H bond of LYS175 with O2 breaks only with the formation of the *aci*-acid intermediate **VII**, in which LYS175 uses its proton to stabilize the negative charge on C2. Only after this stage does O2 abstract the proton (initial H3) from KCX201. The mechanistic roles we ascribe for LYS175 agree well with experimental findings; the K175G mutant does not support enolization but processes 2C3KABP to 3PGA and pyruvate.⁴⁵ In our mechanism, LYS175's action as a proton donor to O2 plays a significant role in enolization by polarizing the C2 carbonyl group; in its absence, the chain of electron flow shown in Scheme 2 is broken and enolization is inhibited. On the other hand, we now predict an active role for LYS175 (not as part of the electron-flow chain) in the subsequent reaction steps, and thus, its absence would be expected to affect only the

stabilization of intermediates. This is consistent with the K175G mutant findings; i.e., the mutant does catalyze forward processing of the β -keto intermediate, *albeit* at a rate 1 order less than that for the wild-type enzyme.⁴⁵

4.3.3. HIS294. Our mechanism envisages a much greater role for HIS294 in the overall reaction sequence than previously suggested. Our results support the suggestion of King et al.⁷ that HIS294 is the base that initiates carboxylation, by abstracting H–O3. Subsequently, HIS294 shuttles this proton between its N ϵ atom and the O3 atom, thus modulating the C3–O3 bond distance. This is evident from the R_{C3-O3} and $R_{O3...H}$ distances in Table 3 for structures **III** to **VI** (see Figure 5). Thus, we propose that HIS294 has active roles in carboxylation, hydration, and C2–C3 bond scission. To date, there has been no acceptable explanation in the literature for the observed inactivity of the ASN and GLN mutants of HIS294 for both enolization and forward processing of 2C3KABP.¹⁹ Our mechanism provides a convincing explanation. The roles of HIS294 (participation in the chain of electron flow; Scheme 2) in carboxylation, and again in subsequent reaction steps, clearly explain the lack of activity for forward processing of 2C3KABP by the mutants. Although a role for HIS294 in enolization is not so obvious, further analysis reveals one. In the current scheme, the N ϵ atom of HIS294 H bonds with H–O3 in the Rubisco•RuBP complex, and we propose that this strong H bond is the key to understanding the indirect role played by HIS294 in enolization. For KCX201 to abstract H3, H–O3 needs to be held away from KCX201. If rotated appropriately, the O3 proton could, in fact, adopt a position much closer to the free oxygen atom of the carbamate group than H3. Thus, H–O3 being more acidic would bind (or be transferred) to the free oxygen atom of the carbamate in the absence of an alternative base that binds H–O3. In the H294N and H294Q mutants, there is no alternate base to bind H–O3, and hence, it is free to bind tightly with KCX201. This

strong H bond will prevent KCX201 from abstracting H3 and, thereby, inhibit enolization altogether, as found for these HIS294 mutants.¹⁹

4.3.4. LYS334. Our mechanism supports the consensus view in the literature that LYS334 is crucial for stabilization of the carboxylated intermediate and the corresponding TS. The short $RO_{S2} \cdots H_{LYS334}$ distance for structure **III** (1.50 Å) demonstrates the extent of stabilization of the carboxylated intermediate by LYS334. Note that with the **FM16** model we observe proton donation from LYS334 to the C2-carboxylate group until the stereoinversion at C2, when it regains its proton from the C2-carboxylate and shares one of its protons with LYS175 (see structure **IX**) to compensate for the loss of a proton from LYS175 to C2. The shortening of the $H_{LYS334} \cdots O_{S2}$ length from structure **V** to **VII** also suggests a minor role for LYS334 in stabilizing the *aci*-acid intermediate (**VII**).

4.3.5. GLU204. No role for GLU204 in the carboxylase mechanism has been proposed in the literature. However, the geometries of reaction species defined by our mechanistic pathway suggest an active role for GLU204 in the hydration reaction. We see that GLU204 is H bonded to H_{w1} of $H_2O[Mg]$ in the β -keto intermediate **III** and activates the water for hydration by abstracting this proton. This mechanism is supported by the TS for the hydration step in which the transition vector corresponds to simultaneous formation of a bond between O_w and C3 and transfer of H_{w1} from O_w to the free oxygen atom of GLU204. Even though another TS (**IVa**) was located with the C2-carboxylate acting as a base, its energy is 1.8 (1.1) kcal/mol higher than **IV**, and the lowest energy structure for the subsequent intermediate (**V**) indicates that the energetically preferred scheme involves GLU204 as a base. GLU204 then remains protonated, even in the geometry for the final products, **IX**. Note that this proton is well positioned for relay to LYS334/LYS175 to return the enzyme back to its original state, ready to bind a new RuBP molecule.

LYS177 and ASP203 are not assigned any active role in the present mechanism. However, their positioning with respect to the active groups suggests they are likely essential for maintaining the delicate balance in the acid–base chemistry proposed in the mechanism, through a network of H bonds.

4.4. Residues Involved in the Gas-Addition Step. The mechanism suggests roles in the gas-addition step for residues in three spatially distinct regions of the active site. As shown in Table 4, proposed roles for HIS294 and LYS334 in the C-terminal domain of the large subunit (LSU) are similar to those of the published consensus enzyme-assisted mechanism.⁴³ HIS294 is the essential base that initiates the CO_2 or O_2 additions, whereas LYS334 is essential for stabilization of the reaction product from the gas addition. However, the mechanism also reveals hitherto unsuspected roles for a cluster of three residues, ASN123, GLU60, and TYR20, in the N-terminal domain of the partner LSU of the dimer containing the active site. These three residues are positioned at the “top” of the active site (shown in Figure 2). Our modeling shows that GLU60 directly affects the positioning of the gas molecule for its addition to the enediol intermediate and that ASN123 and TYR20 participate in this positioning through their interactions with GLU60. These roles are a novel feature of the current mechanism and arise from the active-site structure with the water molecule still coordinated to Mg during and after the gas-addition step. The salt bridge between GLU60 and LYS334 also contributes to the energetic effects of LYS334's interactions with the C2-carboxylate group during the reaction as mapped in Figure 5. Implication of

ASN123, GLU60, and TYR20 in the mechanism points to a greater influence of the N-terminal domain than has been suggested previously.

4.5. Reinterpretation of Other Experimental Data. In addition to being consistent with a wide range of mutational and other data, as already discussed, our mechanism provides the basis for reinterpreting some other experimental results. One new point of intersection is the displacement of the water molecule from Mg coordination. As indicated by our preliminary analysis in Materials and Methods of the feasibility of our new hydration mechanism, displacement of the water molecule, as observed in the crystal structure of the Rubisco·2CABP inhibitor complex, would require substantial local reorganization in the structure, including the breaking and formation of Mg-coordinate bonds. The results of our test calculations suggest this rearrangement is energetically less likely for the β -keto intermediate of the reaction (2C3KABP) compared with the analogue 2CABP, but in any case thermodynamically it would be expected to occur over the long time period for crystallization of the Rubisco·2CABP inhibitor complex. Our proposition is also supported by other experimental evidence. Pierce et al.⁵⁹ showed that the binding of 2CABP takes place in two steps, a fast first step and a slower second step (0.04 s^{-1}) that accounts for the tighter binding of 2CABP to Rubisco. While Pierce et al.⁵⁹ interpreted the slow step as an “isomerization” of the enzyme complex, our mechanism suggests this slow process is the displacement of the third Mg-coordinated water molecule from the coordination sphere by the C2-carboxylate group of 2CABP.

An obvious conclusion from the intricacy of specific patterns of protonation and H bonding of residues and intermediates at different stages of the reaction defined by our mechanism is that each intermediate would have a different preferred pattern for energetically favorable binding. In the normal reaction sequence starting with RuBP, this pattern will naturally occur during progression through the steps. Thus, the Rubisco active site has an inbuilt memory of where it is up to in the sequence of steps. However, an attempt to start the reaction from an intermediate point could be expected to result in less efficient or no catalysis. One interesting experimental result mentioned briefly in the Results is the finding that Rubisco catalyzes the forward processing of externally fed 2C3KABP at a rate only $\sim 3\%$ of the overall turnover rate with RuBP as substrate.⁵⁸ We surmise that when 2C3KABP is fed to Rubisco, residues KCX201 and HIS294, in particular, will be in protonation states suitable for binding RuBP but not 2C3KABP. Thus, the complex of exogenous 2C3KABP and Rubisco must undergo proton reorganization before the hydration can proceed. We suggest that this accounts for the reduced rate of forward processing of 2C3KABP by Rubisco.

4.6. Regenerating Active Rubisco. A coherent reaction mechanism for enzyme catalysis must incorporate means for replenishing the active form of enzyme after a reaction cycle has completed. With the newly proposed mechanism for Rubisco carboxylase activity reported here, the protonation states of residues at the completion of reaction are as follows: GLU204 and HIS294 are in the protonated state and LYS175 is short of a proton, while all other residues retain their initial chemical state. For the reaction products, the C3 fragment is a neutral carboxylic acid, while the C2 fragment is an anionic species tightly coordinated to Mg. When these two species dissociate from the Rubisco active site, the C2

fragment, being anionic, can abstract a proton from either HIS294 or GLU204, leaving only one of these residues in a protonated state. We have not performed calculations to determine the subsequent chemical states of the Rubisco active site after the completion of a catalytic cycle, but we can speculate on possible ways by which the enzyme can be returned to its native active state by proton shifts. However, it must be kept in mind that once the active site is again exposed to solvent after the opening of loop 6, restoration of the protonation states of the active-site residues could also be effected by proton shifts mediated by water. If the additional proton is located on GLU204 after the dissociation of products, it could be donated to LYS175 as they are spatially accessible to each other, directly in the absence of any oxygen atom occupying the top coordination position to Mg, or indirectly through the coordinated oxygen atom as a relay point. Alternatively, if HIS294 has the additional proton it could be relayed to LYS175 through GLU204. As loop 6 needs to open to release the final products from the active site, the LYS175–LYS334 interaction would be broken in this process. Hence, it is likely that the proton reorganization (involving GLU204/HIS294 and LYS175) and reorganization of the coordination sphere (removal of reaction products and coordination of water molecules to Mg) take place simultaneously.

5. Conclusions

Rubisco is a prime target for re-engineering, as better Rubiscos would have major impacts on global efforts to improve the photosynthetic, water, and nutrient efficiency of plants. Systematic mutagenesis and crystallographic studies to define the roles of active-site residues known to affect its activity have not led to the understanding necessary to underpin rational re-engineering. The chemical complexity, including acid–base catalysis, of the multistep mechanism has made attribution of the functional roles of the residues intractable by experiment. Consequently, no mutants with the desired improved efficiency have resulted from all this work. Similarly, although earlier computational studies provided some useful clues to complement experiment, the level of modeling which was computationally feasible was inadequate to represent the active-site chemistry realistically. However, as a result of major increases in computational capability we have now been able to define the complete mechanistic pathway for Rubisco carboxylase catalysis at the required level of detail, by performing comprehensive QM calculations using the B3LYP functional and a large 6-31G(d,p) basis set on large active-site fragment models (FMs) with up to 77 atoms. Use of such large FMs enabled us to include fragments from all the active-site residues in the immediate environment of substrate and intermediates undergoing reaction in all steps of the pathway, viz. LYS175, LYS177, KCX201, ASP203, GLU204, HIS294, and LYS334. These results, coupled with some original ideas for what might be possible mechanistically, have allowed us to propose a novel mechanistic pathway for the complete carboxylase activity that defines precise roles for all the key active-site residues.

Our new mechanism identifies two amino acid residues, one acting as a proton donor and the other as an acceptor, for each step in the sequence of reactions at the Rubisco active site. This dissection allows us to identify amino acid residues that could be targeted by re-engineering strategies aimed at influencing the kinetics of any of the partial

reactions. Solutions to several key questions, hitherto unanswered, are revealed by the new mechanism. First, the use of the third Mg-coordinated water molecule as the source of the water for hydration resolves the question of the fate of this water molecule which had not been defined in the literature. It had simply been assumed it was displaced by the newly formed C2-carboxylate/peroxide group formed after the gas-addition step. Our new mechanism both demonstrates the possibility of this water attacking the C3 atom of the β -keto intermediate (2C3KABP) after the gas addition step and defines the source of the water molecule for hydration. This possibility, which might now appear intuitive, had not been considered previously as in the crystal structure of the Rubisco complex with Mg^{2+} and the inhibitor 2CABP, a structural analogue of 2C3KABP, the C2-carboxylate group occupies the coordination position of the third water molecule. However, this static picture could result from either long-time equilibration in the crystal and/or differences in stability of the 2CABP and 2C3KABP complexes. In support of the latter proposition, our calculations show that displacement of the third water molecule by 2C3KABP is energetically less favorable compared with its displacement by 2CABP. A second question answered is the role of KCX201 subsequent to enolization. There have been conflicting suggestions in the literature of the identity of the active base in the reactions that follow enolization, with some proposals for a continued role for KCX201. However, our mechanism proposes no active role for KCX201 beyond enolization, although it does act as a H bond donor/acceptor and assists in modulating the electronic charge on the reactive center of the substrate and intermediate through its Mg-coordination. Third, the new mechanism identifies the bases involved in hydration and the C2–C3 bond-scission steps. We propose that GLU204, which had not been assigned any role in the literature, abstracts a proton from the Mg-bound water molecule and initiates hydration, while HIS294 triggers the breaking of the C2–C3 bond.

Our mechanism confirms the roles for LYS175 and LYS334 suggested in the literature, with LYS175 playing a direct role in enolization by polarizing the C2-carbonyl group and LYS334 acting to stabilize the carboxylated intermediate 2C3KABP and the corresponding transition state. But finally, a significant finding of our work is that HIS294 plays multiple active roles in the carboxylation, hydration, and C2–C3 bond-scission reactions by shuttling a proton between its N_ϵ atom and the O3 atom of the substrate and intermediates, thus modulating the C3–O3 bond distance, and also a passive structural role in the enolization reaction by binding H–O3. We have also explained possible ways of returning the enzyme back to its initial chemical state after the catalytic cycle. In summary, we expect that our new reaction mechanism with redefined roles for many of the active-site residues will provide the impetus for renewed attempts by researchers to probe Rubisco's chemistry and to engineer mutants to tweak its catalytic efficiency and specificity.

Acknowledgment. This work has been carried out with funding from the Australian Research Council Linkage Grant LP034746, with Rio Tinto as the industry partner, and support from the Australian National University Institute for Advanced Studies (IAS) block grant. Generous grants of computer time from the Australian Partnership for Advanced Computing National Facility (APAC NF) are gratefully acknowledged. We thank Dr. Vlad Vassiliev (APAC

NF) for access to his program JmolEditor which was used to prepare Figures 2 and 5.

Supporting Information Available: Full author list for ref 50; absolute energies and Cartesian coordinates for the geometries shown in Figure 5; MESP-derived atomic point charges for structures **I** and **VII**; Figure S1 providing structures m1, m1', m2, and m2'; Figure S2 providing

graphical representation of structural deviation of optimized structures compared with crystal geometry of 2CABP•Rubisco complex; Figure S3 comparing PES for **FM16** and **FM20**; and Table S1 with energetics for **FM20** computed using IEFPCM. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA803464A