Supporting Information

# Substrate distortion contributes to the catalysis of orotidine 5'-monophosphate decarboxylase

Masahiro Fujihashi[1], Toyokazu Ishida[2], Shingo Kuroda[1], Lakshmi P. Kotra[3,4], Emil F. Pai[3,5*] and Kunio Miki[1*]

[1]Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto, 606-8502, Japan

[2]Nanosystem Research Institute (NRI), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Tsukuba 305-8568, Japan

[3]Center for Molecular Design and Preformulations and Division of Cell & Molecular Biology, Toronto General Research Institute/University Health Network, Toronto, Ontario, Canada M5G 1L7

[4]Departments of Pharmaceutical Sciences and Chemistry, McLaughlin Center for Molecular Medicine, University of Toronto, Toronto, Ontario, Canada M5S 3M2

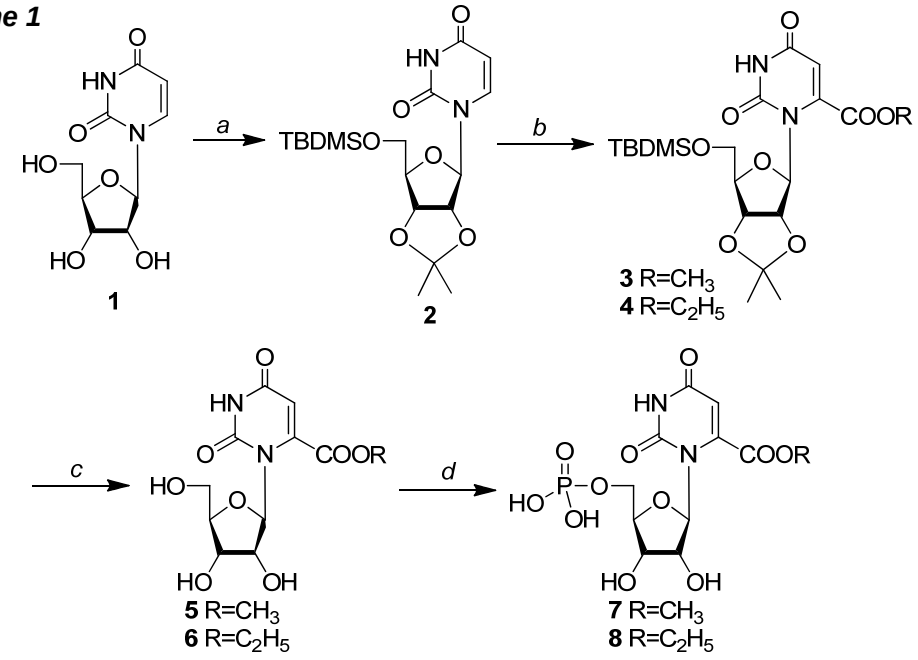
[5]The Campbell Family Cancer Research Institute, Ontario Cancer Institute/University Health Network, and Departments of Biochemistry, Medical Biophysics, and Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5G 1L7

## Supporting Material and Methods

**Synthesis of methyl and ethyl esters of OMP**

Methyl and ethyl esters of OMP were synthesized from uridine. First, the introduction of methoxy or ethoxycarbonyl moieties at the C-6 position was achieved *via* a lithium diisopropylamide (LDA)

***Scheme 1***




Reagents: (a) i. acetone/$H^+$; ii. TBDMSCl, imidazole/$CH_2Cl_2$, 0-25 °C; (b) LDA, $ClCO_2Me$ for **3** or $ClCO_2Et$ for **4**, THF, -78 °C; (d) 50% TFA, r.t.; (e) $POCl_3$, Py/$H_2O$/$CH_3CN$, 0 °C.

mediated reaction with the appropriate alkyl chloroformate. Deprotection of the protecting groups with trifluoroacetic acid (TFA)[1] followed by the monophosphorylation with phosphorus oxychloride afforded the mono-phosporylated nucleoside **5** (Scheme 1).[2,3] Finally, monophosphate compound **5** was transformed into the ammonium salt by neutralization with 0.5 M $NH_4OH$ solution at 0 °C and lyophilized to obtain the ammonium salts of the corresponding nucleotides.

**5'-*O*-*t*-Butyldimethylsilyl-2',3'-*O*-isopropylidene uridine (2)**. A stirred suspension of uridine (1g, 4.1 mmol) in dry acetone (50 mL) was treated with $H_2SO_4$ (0.5 mL) drop-wise at room temperature and the resulting mixture was stirred further for 1 h and neutralized with $Et_3N$. Evaporation of the solvent and purification of the crude by column chromatography (5-8% MeOH in $CHCl_3$) gave 2,3-*O*-isopropylidenuridine (1.15 g) in quantitative yield as a white solid. $^1$H NMR ($CDCl_3$) δ: 1.36 (s, 3H, -$CH_3$), 1.57 (s, 3H, -$CH_3$), 3.80 (dd, 1H, *J* = 3.3, 12.0 Hz, H-5'), 3.91 (dd, 1H, *J* = 2.7, 12.0 Hz, H-5''), 4.26-4.30 (m, 1H, H-4'), 4.95 (dd, 1H, *J* = 3.3, 6.3 Hz, H-3'), 5.02 (dd, 1H, *J* = 2.7, 6.3 Hz, H-2') 5.56 (d, 1H, *J* = 2.7 Hz, H-1'), 5.72 (d, 1H, *J* = 8.1 Hz, H-5), 7.36 (d, 1H, *J* = 8.1 Hz, H-6). A stirred solution of 2,3-*O*-isopropylidenuridine (0.2 g, 0.7 mmol) in dry $CH_2Cl_2$ (3 mL) was treated with imidazole (0.1 g, 1.4 mmol) and TBDMSCl (*tert*-butyldimethylsilyl chloride, 105 mg, 0.7 mmol) at 0 °C. The reaction mixture was brought to room temperature and stirred for 1 h. The solvent was evaporated under vacuum and the solid taken into ethyl acetate (30 mL), washed with water (15 mL), brine (15 mL) and dried ($Na_2SO_4$). Evaporation of the solvent and purification of crude by column chromatography (5% MeOH in $CHCl_3$) gave **2** (0.268 mg) in 96% yield as a foamy solid. $^1$H NMR ($CDCl_3$): δ 0.10 (s, 6H, -$CH_3$), 0.90 (s, 9H, -$CH_3$), 1.36 (s, 3H, -$CH_3$) 1.59 (s, 3H, -$CH_3$), 3.79 (dd, 1H, *J* = 2.7, 11.2 Hz, H5'), 3.92 (dd, 1H, *J* = 2.4, 11.2 Hz, H-5''), 4.30-4.33 (m, 1H, H-4'), 4.67 (dd, 1H, *J* = 2.7, 6.0 Hz, H-3'), 4.75 (dd, 1H, *J* = 3.0, 6.0 Hz, H-2'), 5.66 (d, 1H, *J* = 8.1 Hz, H-5), 5.96 (dd, 1H, *J* = 3.0 Hz, H-1'), 7.68 (d, 1H, *J* = 8.1 Hz, H-6), 8.47 (brs, 1H, -NH).

**5'-*O*-*t*-Butyldimethylsilyl-2',3'-*O*-isopropylidene -6-methoxycarbonyl uridine (3)**. A stirred solution of compound **2** (0.25 g, 0.6 mmol) in dry THF (tetrahydrofuran, 2 mL) was treated with LDA (0.62 mL, 1.26 mmol, 2.0 M solution in THF) at -78ºC. After stirring for 1 h, methylchloroformate (0.048 g, 0.6 mmol) in dry THF (2 mL) was added and the mixture was stirred for another 5 h at the same temperature. The reaction was quenched with AcOH (0.3 mL), then brought to room temperature and dissolved in ethyl acetate (25 mL). The organic layer was washed with saturated $NaHCO_3$ solution (10 mL), 5% $Na_2S_2O_3$ solution (10 mL), brine (10 mL) and dried ($Na_2SO_4$). Evaporation of the solvent and purification of crude by column chromatography (hexane:ethyl acetate, 70:30) gave **3** (180 mg, 64% yield) as a syrup. $^1$H NMR ($CDCl_3$): δ 0.056 (s, 6H, -$CH_3$), 0.88 (s, 9H, -$CH_3$), 1.34 (s, 3H, -$CH_3$) 1.54 (s, 3H, -$CH_3$), 3.75 (dd, 1H, *J* = 7.2, 10.9 Hz, H5'), 3.81 (dd, 1H, *J* = 5.1, 10.9 Hz, H5'), 3.93 (s, 3H –$CH_3$), 4.06-4.12 (m, 1H, H-4'), 4.71 (dd, 1H, *J* = 4.8, 6.4 Hz, H-3'), 5.15 (dd, 1H, *J* = 2.0,1 6.4 Hz, H-2'), 5.89 (d, 1H, *J* = 2.1 Hz, H-1'), 6.07 (s, 1H, H-5), 9.32 (brs, 1H, -NH).

**6-Methoxycarbonyl uridine (5).** A stirred solution of compound **3** (0.23 g, 0.5 mmol) was treated with 50% aqueous TFA (3 mL) at 0ºC and then brought to room temperature and stirred for 2 h. Evaporation of solvent and purification of crude by column chromatography (10-15% EtOH in $CHCl_3$) gave **5** (135m g, 89% yield) as a solid. $^1$H NMR (DMSO (dimethyl sulfoxide) -$D_2O$): δ 3.37 (dd, 1H, *J* = 6.6, 12.0 Hz, H-5'), 3.54 (dd, 1H, *J* = 3.6, 12.0 Hz, H-5''), 3.62-3.67 (m, 1H, H-4'), 3.80 (s, 3H, -$CO_2CH_3$), 388-3.97 (m, 1H, H-3'), 4.41 (dd, 1H, *J* = 4.2, 6.3 Hz, H-2'), 5.34 (d, 1H, *J* = 4.2 Hz, H-1'), 5.95 (s, 1H, H-5).

**6-Methoxycarbonyluridine-5'-*O*-monophosphate (7)**. A stirred solution of $H_2O$ (0.02 g, 1.1 mmol) and $POCl_3$ (0.16 mL, 1.7 mmol) in dry acetonitrile (3 mL) was treated with pyridine (0.154 mL, 1.91 mmol) at 0°C and stirred for 10 min. Compound **5** was added (0.12 g, 0.4 mmol) and the mixture was stirred for another 5 h at the same temperature. The reaction mixture was quenched with 25 mL of cold water and stirring was continued for 1 h. Evaporation of solvent and purification of crude by column chromatography (Dowex ion-exchange basic resin, 0.1 M formic acid) gave **7** as a syrup. UV ($H_2O$): λmax = 274 nm; ; $^1$H NMR ($D_2O$): δ 3.99 (s, 3H –$CO_2CH_3$), 4.02-4.08 (m, 2H, H-5',5''), 4.16-4.23 (m, 1H, H-4'), 4.37 (t, *J* = 6.6 Hz 1H, H-3'), 4.75 (dd, 1H, *J* = 3.3, 6.6 Hz, H-2'), 5.70 (d, 1H, *J* = 3.6 Hz, H-1'), 6.26 (s, 1H, H-5).

**5'-*O*-*t*-Butyldimethylsilyl-6-ethoxycarbonyl-2',3'-*O*-isopropylidene uridine (4)**. A stirred solution of compound **2** (0.25 g, 0.63 mmol) in dry THF (2 mL) was treated with LDA (0.62 mL, 1.23 mmol, 2.0 M solution in THF) at -78°C. After stirring for 1 h, ethyl chloroformate (0.048 g, 0.6 mmol) in dry THF (2 mL) was added and the mixture was stirred for another 5 h at the same temperature. The reaction was quenched with AcOH (0.3 mL), then brought to room temperature and dissolved in ethyl acetate (25 mL). The organic layer was washed with saturated $NaHCO_3$ solution (10 mL), 5% $Na_2S_2O_3$ solution (10 mL), brine (10 mL) and dried ($Na_2SO_4$). Evaporation of the solvent and purification of crude by column chromatography (hexane:ethyl acetate, 70:30) gave **4** (0.18 g) in 64% yield as a syrup. $^1$H NMR ($CDCl_3$) δ (ppm) 9.55 (bs, 1H, NH), 6.09 (s, 1H, H-5), 5.91 (d, J = 1.5Hz, 1H, H-1' ), 5.19 (dd, J = 1.5, 6.6 Hz, 1H, H-2' ), 4.75 (dd, J = 4.8, 6.6 Hz, 1H, H-3' ), 4.40 (m, 2H, -O$\underline{CH_2}$$CH_3$), 4.11 (m, 1H, H-4' ), 3.80 (m, 2H, H-5', 5" ), 1.55 (s, 3H, -C-$\underline{CH_3}$), 1.39 (t, J = 7.2 Hz, 3H, -$OCH_2\underline{CH_3}$), 0.89 (s, 9H, -Si-$\underline{tBu}$), 0.05 (s, 6H –Si-$\underline{(Me)_2}$). 13C NMR ($CDCl_3$) δ (ppm) 162.48, 161.58, 150.16, 145.78, 114.42, 106.02, 93.70, 89.06, 84.88, 81.77, 64.13, 63.70, 27.43, 26.11, 25.56, 18.64, 14.03, -5.061, -5.09. MS EI (+) m/z = 455.27 $[M\text{-}CH_3]^+$, 413.23 $[M\text{-}t\text{Bu}]^+$.

**6-Ethoxycarbonyl uridine (6).** A stirred solution of compound **4** (0.23 g, 0.5 mmol) was treated with 50% aqueous TFA (3 mL) at 0°C and then brought to room temperature and stirred for 2 h. Evaporation of solvent and purification of crude by column chromatography (10-15% EtOH in $CHCl_3$) gave **6** (140 mg, 89% yield) as a solid. $^1$H NMR ($CDCl_3$) δ (ppm) 6.15 (s, 1H, H-5), 5.59 (d, J = 4.0 Hz, 1H, H-1' ), 4.68 (dd, J = 4.0, 6.9 Hz, 1H, H-2' ), 4.43 (m, 2H, -O$\underline{CH_2}$$CH_3$), 4.23 (t, J = 6.9 Hz, 1H, H-3' ), 4.38 (m, 1H, H-4' ), 3.81 (bd, J = 3.0, 1H, H-5'), 3.68 (dd, J = 6.6, 12.0 Hz, 1H, 5" ), 1.38 (t, J = 7.2 Hz, 3H, -$OCH_2\underline{CH_3}$). 13C NMR ($CDCl_3$) δ (ppm) 165.52, 163.41, 151.97, 147.75, 106.17, 95.16, 85.46, 73.31, 70.55, 65.42, 62.89, 14.19. EI (+) m/z = 339.1 $[M+Na]^+$. $\lambda_{max}$ ($H_2O$) 272nm.

**6-Ethoxycarbonyluridine-5'-*O*-monophosphate (8)**. A stirred solution of $H_2O$ (0.02 g, 1.1 mmol) and $POCl_3$ (0.16 mL, 1.7 mmol) in dry acetonitrile (3 mL) was treated with pyridine (0.15 mL, 1.9 mmol) at 0°C and stirred for 10 min. Compound **6** was added (0.12 g, 0.4 mmol) and the mixture was stirred for another 5 h at the same temperature. The reaction mixture was quenched with 25 mL of cold water and stirring was continued for 1 h. Evaporation of solvent and purification of crude by column chromatography (Dowex ion-exchange basic resin, 0.1 M formic acid) gave **8** as a syrup. 1H NMR ($H_2O$) δ (ppm) 6.00 (s, 1H, H-5), 5.53 (bs, 1H, H-1' ), 4.72 (m, 1H, H-2' ), 4.32 (m, 1H, H-3' ), 4.12-3.81 (m, 3H, H-4', H-5', 5" ), 4.61 (q, J = 6.9 Hz, 2H, -O$\underline{CH_2}$$CH_3$), 1.14 (t, J = 6.9, 3H, -$OCH_2\underline{CH_3}$). EI (-ve) m/z = 366.0 $[M\text{-}Et]^-$. $\lambda_{max}$ ($H_2O$) 272nm.

**Computational Details**

Computational details based on hybrid QM/MM modeling were reported in previous

publications.[4-6] Throughout this article, we employed the electrostatic embedding scheme in all *ab initio* QM/MM calculations; QM/MM and MM (modeling and simulation) calculations were based on our original codes; the *ab initio* QM/MM program has been developed based on the HONDO package,[7] and MM modeling/simulation routines were added to the MO calculation part. The AMBER parameter set (parm.96) was used for the force field calculations.[8,9]

**1. Preparation of initial enzyme structure**

The initial coordinates of proteins were adopted from the X-ray geometry of wild-type ODCase complexed with barbituric acid ribosyl 5'-monophosphate (BMP) determined at 1.45 Å resolution (PDB code 1X1Z). The natural substrate (OMP) structure was modeled and placed at the two active sites to retain maximum overlap with the original BMP positions. Hydrogen atoms were added to the ES complex in the standard modeling procedure by assuming a standard protonation state for all of the polar residues under physiological pH conditions. This assumption is consistent with the p$K_a$ of the individual residues estimated by PROPKA[10,11] (supporting table S2). With this definition, the total charge of the ES complex was -18. We added 18 sodium ions to neutralize the total charge of this initial model. These counterions were placed at the positions of largest negative electrostatic potentials. We also considered several crystallographic water molecules clearly observed in the X-ray coordinates. Unfavorable steric contacts were removed by initial MM energy minimizations.

**2. Molecular Dynamics simulation for the ES complex**

Next, we performed molecular dynamics (MD) simulation to reliably model the ES complex structure, which has a stable conformation at the free energy minimum regions. The initial MM-refined model of the ES complex was solvated in a sphere of TIP3P water molecules[12] with a 45 Å radius centered on the center of mass of the complex. Any water molecules that came within 3.0 Å of the ES complex were removed: the resultant system consisted of ~13,000 water molecules, 18 sodium ions and the ES complex. The solvated enzyme complex was fully relaxed by performing MD simulations for more than 2 ns periods. The Nose-Hoover-chain (NHC) method was employed to generate the NVT (Number of atoms, Volume, Temperature) ensemble, and the system temperature was maintained at 303 K by attaching five chains of thermostat.[13] In all simulations, spherical boundary conditions were employed to keep the solvation structure. A weak harmonic constraint potential was added on the surface boundary of the solvation sphere. No cut-off for the non-bonded interaction was introduced in all simulations. After collecting stable trajectories for more than 2 ns periods, we sampled 10 representative structures from the MD trajectories for the following QM/MM structural optimizations.

**3. QM/MM structural modeling of the decarboxylation pathway**

Initial models of QM/MM calculations were selected from the 10 sampling geometries extracted from the stable region estimated from the MD simulations. After annealing and quenching each solvated protein complex by MM calculations, we performed *ab initio* QM/MM structural optimizations for the whole solvated enzyme complexes. In all QM/MM calculations, we only considered a single catalytic site of the dimer when following the reaction path. The QM region in each sampling structure contains the side chains of Lys42, Aps70, Lys72, Asp75'[†] and the substrate OMP molecule. Boundaries between QM and MM regions ($C_\alpha$-$C_\beta$) were saturated using dummy hydrogen atoms, which were allowed to move freely during the QM/MM geometry

[†] ' indicates that the residue belongs to the second subunit.

optimizations. Considering the system size of an ODCase complex, we employed in all QM/MM reaction path optimizations the restricted Hartree-Fock (RHF) method with the 6-31(+)G** basis set and diffuse functions added to the carboxylic groups of two Asp residues and the six-membered ring of OMP. For following the direct decarboxylation path, we optimized the reaction path using constrained QM/MM optimizations, in which the reaction coordinate was defined as a bond distance between the C6 and C7 atoms. After confirming geometries along the reaction coordinate, we finally performed the second-order Moller-Plesset perturbation (MP2) energy correction with the same basis sets.

For a more detailed discussion of molecular electronic properties, we employed a partial charge model derived from the molecular electrostatic potential (ESP). To compare the surrounding environment in the aqueous phase and the enzyme on an equal basis, the whole OMP structure was redefined as a QM region in QM/MM computations. By adding a constraint to conserve the total electron density of molecules in the QM fragment, the electron distributions of molecules are appropriately reduced into a simple charge model. The basic procedure to calculate and fit ESP is similar to the standard methods within the framework of QM/MM computation. In both cases, derived ESP charges well reproduce the total dipole moments of the QM fragment.

**4. All-Electron QM calculations by the Fragment Molecular Orbital (FMO) Method**

The molecular interaction energies between OMP and surrounding amino acid residues were estimated by all-electron QM computations for the entire protein complex using the Fragment Molecular Orbital (FMO) method.[14] In all FMO computations, we employed the two-body expansion (FMO2) due to the complex size of the enzyme system, and used the GAMESS implemented version.[15] Technical details of FMO2 computations are as follows. Both the atomic and molecular orbital accuracies were increased to $10^{-12}$ using ICUT=12, ITOL=24, and CUTOFF=$10^{-12}$, and the self-consistent field (SCF) convergence was tightened to $10^{-7}$. The same values were used during the monomer SCF cycle where monomer densities converge. Since using diffuse functions in the fragment-based methods often leads to problems, we used the 6-31G* basis set in all FMO2 calculations. The option to remove *s* contaminants from *d* functions was used. For simplicity, we followed a one amino acid residue per fragment partition scheme for fragmentation. The substrate was treated as a single fragment in all FMO2 calculations. The protein backbone was divided into fragments at the $C_\alpha$ positions, keeping peptide bonds intact. The hybrid $sp^3$ orbitals of the carbon atom were used to appropriately divide the molecular orbital space at bond fraction points.

**5. Calculation of the intrinsic electronic energy (side chain rotation / out of plane distortion) for model OMP analogs.**

To complement energy component analyses in the ligand distortion, we also evaluated the intrinsic electronic (QM) energy of ligand distortion for two analog molecules.

The rotational energy barrier of the C6 substituent group was estimated by calculating the potential energy profiles of COO rotation at the MP2/aug-cc-pVDZ level both for 1-methyl-orotate methyl ester and 1-methyl-orotate. All the internal degrees of freedom with fixed C5-C6-COO dihedral angles were optimized by appropriate selection of the internal coordinates of these molecules. The rotational potential energy profiles were scanned and optimized by rotating in 30º steps for this dihedral angle. Convergence thresholds of optimizations are default parameters of GAMESS.[15]

The energy cost to deform the planar form of the orotate structure (Fig. 6E) was also estimated by computational experiments. The intrinsic energy cost (= additional work) necessary

to distort the OMP ligand was evaluated in two steps; (1) extracting the optimized OMP geometry, which was determined by QM/MM structural refinement, from the enzyme active site, and (2) gradually releasing the external steric and electrostatic force created by the protein environment. All the degrees of freedom are relaxed in the aqueous environment with the fixed reaction coordinate of C6-COO distance, and the intrinsic QM energies of orotate inside the enzyme and in the aqueous phase were analyzed. The intrinsic energy was calculated using the PCM-MP2/aug-cc-pVDZ method implemented in the GAMESS package. Convergence threshold and PCM parameters are default values of GAMESS.

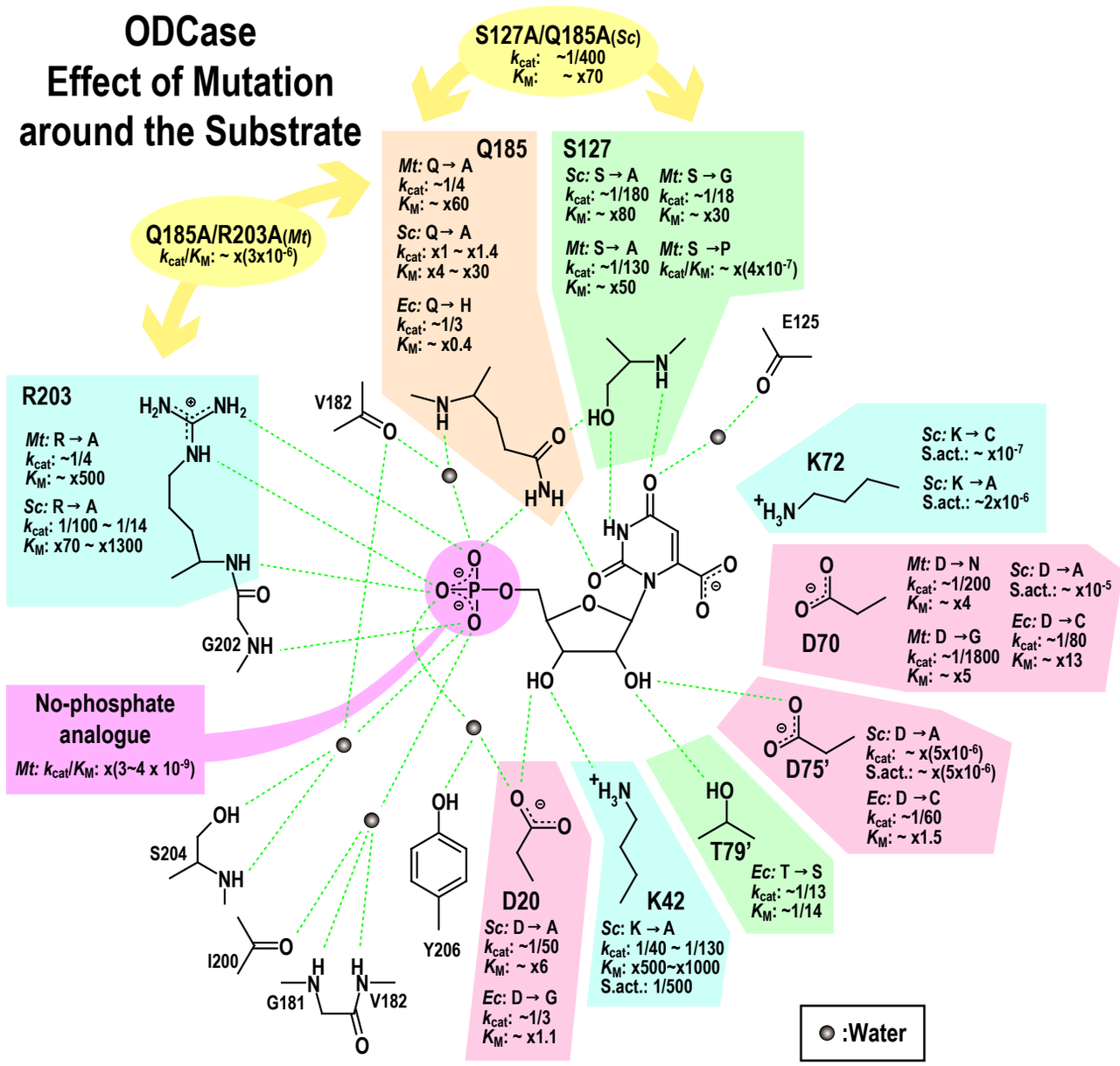



**Figure S1**

*Summary of results of mutagenesis experiments in the substrate binding site of ODCase.* Numbering of the residues is based on *Mt*ODCase. The hydrogen-bonding or charge networks are drawn by green dotted lines based on the BMP binding structure. ' indicates the residue belonging to the second subunit. Abbreviations; *Ec*:, ODCase from *Escherichia coli*, *Mt*: ODCase from *Methanothermobacter thermoautotrophicus*, *Sc*: ODCase from *Saccharomyces cerevisiae,* S. act, specific activity. References; D20A(*Sc*),[16,17] D20G(*Ec*),[18] K42A(*Sc*),[17,19-21] D70N(*Mt*),[22] D70G(*Mt*),[22] D70A(*Sc*),[19] D70C(*Ec*),[18] K72A (*Sc*),[19] K72C(*Sc*),[23] D75A(*Sc*),[19] D75C(*Ec*),[18] T79S(*Sc*),[18] S127A(*Sc*),[24,25] S127A(*Mt*),[26] S127G(*Mt*),[26] S127P(*Mt*),[26] Q185A (*Mt*),[27] Q185A(*Sc*),[19,24,25] Q185H(*Ec*),[18] R203A(*Mt*),[27] R203A(*Sc*),[17,24,28] S127A/Q185A double mutant (*Sc*),[24,25] Q185A/R203A double mutant (*Mt*).[27]

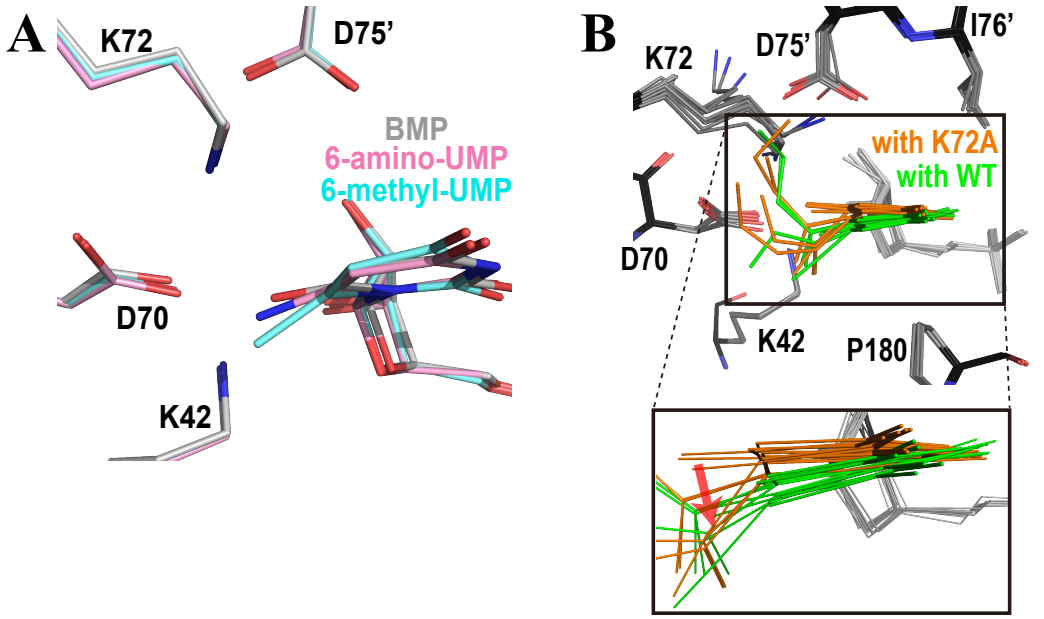



**Figure S2.**

*Superposition of selected structures.* (A) WT-MtODCase with 6-methyl-UMP (cyan) and conformer A of WT-*Mt*ODCase with 6-amino-UMP (pink) superimposed on WT-MtODCase with BMP (gray). The characteristic K42-D70-K72-D75' networks from the three structures superimpose very well. (B) Comparison of the various ligands bound to WT-*Mt*ODCase and K72A-*Mt*ODCase, respctively. The BMP, UMP, 6-cyano-UMP, 6-methyl-UMP, 6-amino-UMP, OMP-methyl-ester and OMP-ethyl-ester complexes bound to WT-*Mt*ODCase (models in green) and K72A-*Mt*ODCase (models in orange) are superimposed. In the lower panel, the red arrow indicates the effect on ligands caused by K72. Note that conformation B of 6-amino-UMP in complex with WT-*Mt*ODCase is included in orange, since in this complex K72(B) is flipped from the typical K72 position in other complexes and cannot influence the ligand positions.

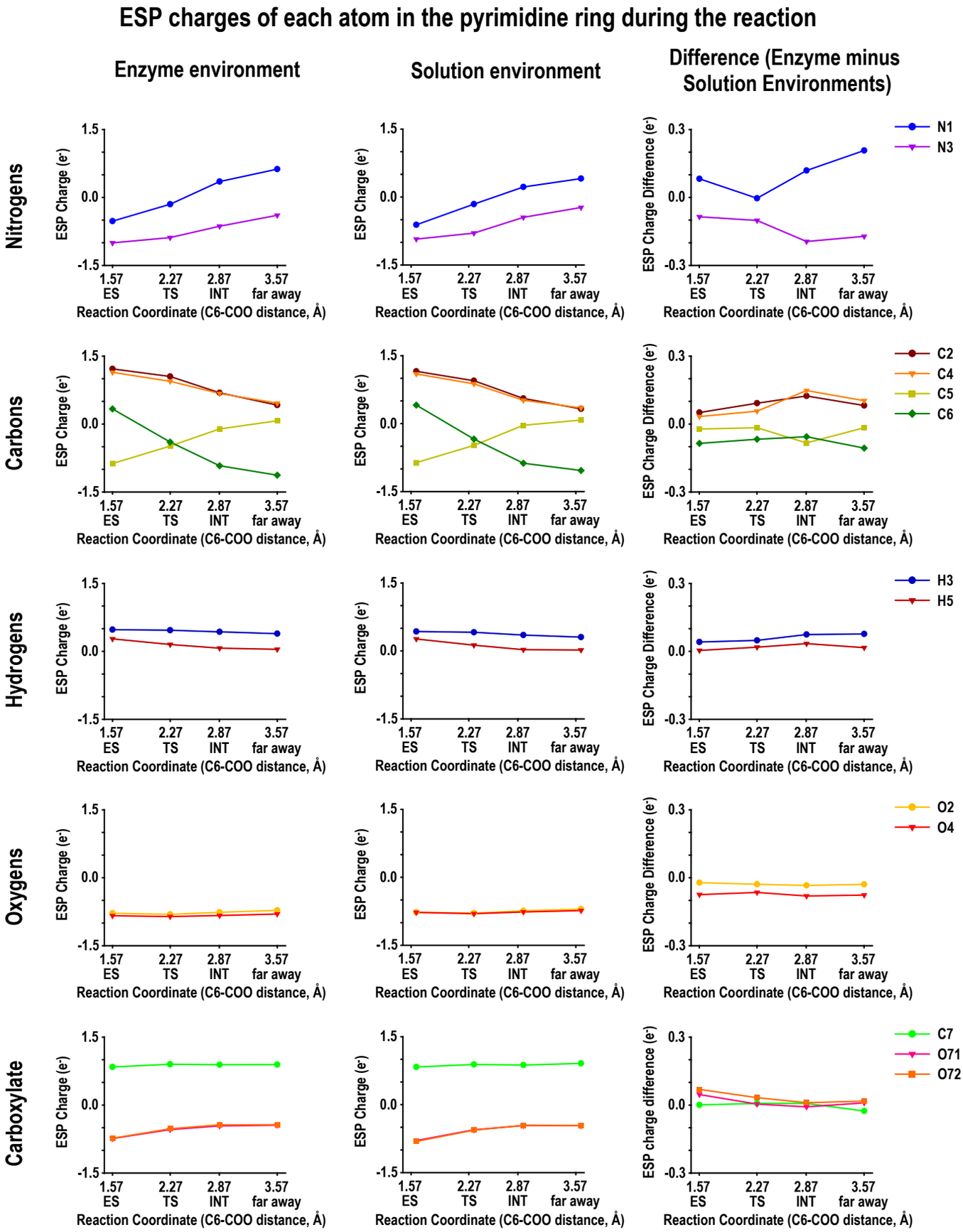
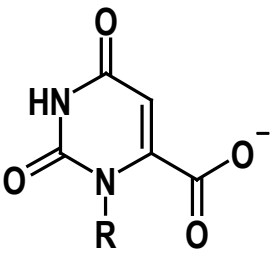
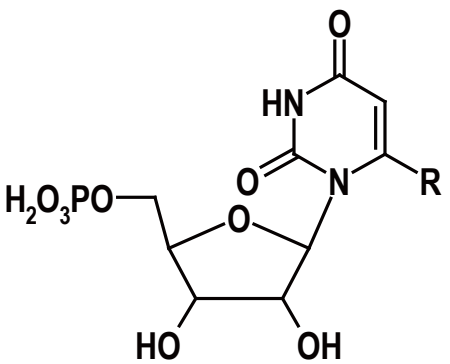



**Figure S3**

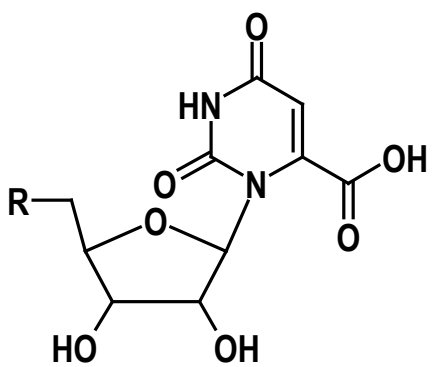
*ESP charges of each atom in the pyrimidine ring during the reaction.* Left and center panels represent the absolute ESP charges of each pyrimidine atom in the enzyme complex and in solution, respectively. Right panels represent the ESP charge difference, which was calculated by subtracting the values in solution from those in the enzyme complex. From the top, the panels show the ESP charges and their differences for the nitrogen, carbon, hydrogen, and oxygen atoms in the ring and for atoms in the carboxylate group, respectively.



| | |
|---|---|
| R = H | orotate |
| $CH_3$ | 1-methyl-orotate |



| | |
|---|---|
| R = $COO^-$ | OMP |
| H | UMP |
| $CH_3$ | 6-methyl-UMP |
| $NH_2$ | 6-amino-UMP |
| OH | BMP |
| $COOCH_3$ | OMP-methyl-ester |
| $COOC_2H_5$ | OMP-ethyl-ester |



| | |
|---|---|
| R = H | 1-(β-D-erythrofuranosyl)orotic acid (EO) |
| OH | orotidine |

**Figure S4**

Chemical structures of the compounds discussed in this paper.

Supporting Table 1. Data and refinement statistics

| Protein | | Wild-type | Wild-type | K72A | D75N | Wild-type | Wild-type | K72A | K72A |
|---|---|---|---|---|---|---|---|---|---|
| Ligand | | 6-methyl-UMP | 6-amino-UMP | 6-amino-UMP | 6-amino-UMP | OMP-methyl-ester | OMP-ethyl-ester | OMP-methyl-ester | OMP-ethyl-ester |
| **Data Statistics** | | | | | | | | | |
| Space group | | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ |
| Unit cell | *a* (Å) | 58.03 | 57.88 | 58.12 | 58.07 | 58.03 | 57.80 | 57.83 | 57.99 |
| | *b* (Å) | 103.57 | 103.35 | 103.53 | 103.60 | 103.56 | 103.49 | 103.88 | 103.47 |
| | *c* (Å) | 73.51 | 73.58 | 74.05 | 73.71 | 73.60 | 73.59 | 73.27 | 73.63 |
| Resolution† (Å) | | 50-1.59 (1.63-1.59) | 100-1.22 (1.24-1.22) | 100-1.72 (1.76-1.72) | 100-1.39 (1.41-1.39) | 100-1.41 (1.43-1.41) | 100-1.59 (1.63-1.59) | 100-1.69 (1.73-1.69) | 100-1.26 (1.28-1.26) |
| No. of Reflections observed / uniq | | 235,777 / 30,174 | 548,146 / 65,804 | 198,000 / 24,073 | 336,947 / 45,280 | 313,526 / 43,044 | 217,637 / 30,030 | 182,335 / 24,736 | 381,754 / 57,994 |
| Completeness† | | 100.0% (100.0%) | 99.9% (100.0%) | 99.9% (100.0%) | 100.0% (100.0%) | 99.6% (99.4%) | 99.9% (99.7%) | 99.0% (98.3%) | 96.6% (93.5%) |
| $I/\sigma(I)$ † | | 29.5 (6.4) | 33.9 (6.8) | 34.8 (6.5) | 39.8 (6.6) | 45.6 (6.5) | 38.2 (6.9) | 35.4 (6.1) | 49.6 (6.0) |
| $R_{merge}$†, ‡ | | 6.9% (30.2%) | 5.9% (31.0%) | 5.8% (31.5%) | 4.8% (30.7%) | 4.2% (30.9%) | 5.1% (29.8%) | 5.2% (30.6%) | 4.0% (21.8%) |
| **Refinement Statistics** | | | | | | | | | |
| $R$ / $R_{free}$§ | | 16.4% / 18.6% | 15.3% / 16.0% | 15.4% / 18.4% | 15.7% / 17.4% | 16.9% / 18.8% | 16.6% / 18.2% | 16.4% / 18.2% | 16.5% / 17.7% |
| No. of Atoms Protein / Ligand / Solvent | | 1,704 / 22 / 153 | 1,915 / 38 / 205 | 1,721 / 22 / 142 | 1,812 / 22+Cl /186 | 1,798 / 29 /156 | 1,885 / 26 / 180 | 1,895 / 29 /146 | 1,871 / 31 / 195 |
| RMSD | Bond (Å) | 0.016 | 0.010 | 0.013 | 0.011 | 0.018 | 0.016 | 0.013 | 0.016 |
| | Angle (º) | 1.593 | 1.437 | 1.423 | 1.442 | 1.809 | 1.630 | 1.413 | 1.813 |
| | Plane (Å) | 0.013 | 0.010 | 0.010 | 0.010 | 0.012 | 0.010 | 0.009 | 0.012 |
| Ramachandran Favored / Allowed | | 98.1% / 100.0% | 97.7% / 100.0% | 97.2% / 100.0% | 97.7% / 100.0% | 98.1% / 100.0% | 97.7% / 100.0% | 97.2% / 100.0% | 97.7% / 100.0% |
| RMSD (Å) to 2zz7 (K72A-BMP) (Superposed Cα atoms) | | 0.11 (215) | 0.09 (215) | 0.08 (215) | 0.09 (215) | 0.10 (215) | 0.09 (215) | 0.09 (215) | 0,08 (215) |
| PDB code | | 3WJW | 3WJX | 3WJY | 3WJZ | 3WK0 | 3WK1 | 3WK2 | 3WK3 |

†Values in parentheses are for the highest resolution shell.
‡$R_{merge} = \Sigma\Sigma_i|I(h) - I(h)_i| / \Sigma\Sigma_i I(h)$, where $I(h)$ is the mean intensity after rejections.
§$R_{free}$ is calculated with a randomly selected 5 % set of reflections.

Supporting Table 2 p*K*a[†] of individual residues estimated by PROPKA[10,11]

| | WT-apo | WT-UMP | WT-BMP |
|---|---|---|---|
| PDB used for estimation | 1dv7[29], 3g18[22] | 1loq[30], 3g1d[22] | 1lor[30], 1x1z[31], 3ltp[32] |
| K42 | 12.56 ± 1.03 | 11.88 ± 0.56 | 12.29 ± 0.56 |
| D70 | 2.37 ± 0.08 | 2.50 ± 1.05 | 1.84 ± 0.84 |
| K72 | 11.46 ± 0.90 | 10.10 ± 1.04 | 10.24 ± 0.74 |
| D75 | 3.45 ± 0.10 | 3.76 ± 0.49 | 3.37 ± 0.06 |

[†]Averaged p*K*a calculated for multiple structures are presented with standard deviations.

## References


(1) Tanaka, H.; Hayakawa, H.; Miyasaka, T. *Tetrahedron* **1982**, *38*, 2635.

(2) Sowa, T.; Ouchi, S. *Bull. Chem. Soc. Jpn.* **1975**, *48*, 2084.

(3) Ueda, T.; Yamamoto, M.; Yamane, A.; Imazawa, M.; Inoue, H. *J. Carb.-Nucleos.-Nucl.* **1978**, *5*, 261.

(4) Ishida, T. *J. Chem. Phys.* **2008**, *129*, 125105.

(5) Ishida, T. *J. Am. Chem. Soc.* **2010**, *132*, 7104.

(6) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035.

(7) Dupuis, M.; Marquez, A.; Davidson, E. R. In *Quantum Chemistry Program Exchange (QCPE)*; Indiana University: Bloomington, In 47405.

(8) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(9) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer Academic Publishers: Leiden, **1997**; Vol. 3.

(10) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525.

(11) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284.

(12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(13) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.

(14) Fedorov, D.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, **2009**.

(15) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(16) Miller, B. G.; Butterfoss, G. L.; Short, S. A.; Wolfenden, R. *Biochemistry* **2001**, *40*, 6227.

(17) Miller, B. G.; Snider, M. J.; Short, S. A.; Wolfenden, R. *Biochemistry* **2000**, *39*, 8113.

(18) Yuan, J.; Cardenas, A. M.; Gilbert, H. F.; Palzkill, T. *Protein Sci.* **2011**, *20*, 1891.

(19) Miller, B. G.; Snider, M. J.; Wolfenden, R.; Short, S. A. *J. Biol. Chem.* **2001**, *276*, 15174.

(20) Shostak, K.; Jones, M. E. *Biochemistry* **1992**, *31*, 12155.

(21) Van Vleet, J. L.; Reinhardt, L. A.; Miller, B. G.; Sievers, A.; Cleland, W. W. *Biochemistry* **2008**, *47*, 798.

(22) Chan, K. K.; Wood, B. M.; Fedorov, A. A.; Fedorov, E. V.; Imker, H. J.; Amyes, T. L.; Richard, J. P.; Almo, S. C.; Gerlt, J. A. *Biochemistry* **2009**, *48*, 5518.

(23) Smiley, J. A.; Jones, M. E. *Biochemistry* **1992**, *31*, 12162.

(24) Barnett, S. A.; Amyes, T. L.; Wood, B. M.; Gerlt, J. A.; Richard, J. P. *Biochemistry* **2008**, *47*, 7785.

(25) Porter, D. J.; Short, S. A. *Biochemistry* **2000**, *39*, 11788.

(26) Iiams, V.; Desai, B. J.; Fedorov, A. A.; Fedorov, E. V.; Almo, S. C.; Gerlt, J. A. *Biochemistry* **2011**, *50*, 8497.

(27) Desai, B. J.; Wood, B. M.; Fedorov, A. A.; Fedorov, E. V.; Goryanova, B.; Amyes, T. L.; Richard, J. P.; Almo, S. C.; Gerlt, J. A. *Biochemistry* **2012**, *51*, 8665.

(28) Toth, K.; Amyes, T. L.; Wood, B. M.; Chan, K. K.; Gerlt, J. A.; Richard, J. P. *Biochemistry* **2009**, *48*, 8006.

(29) Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2017.

(30) Wu, N.; Pai, E. F. *J. Biol. Chem.* **2002**, *277*, 28080.

(31) Fujihashi, M.; Bello, A. M.; Poduch, E.; Wei, L.; Annedi, S. C.; Pai, E. F.; Kotra, L. P. *J. Am. Chem. Soc.* **2005**, *127*, 15048.

(32) Wood, B. M.; Amyes, T. L.; Fedorov, A. A.; Fedorov, E. V.; Shabila, A.; Almo, S. C.; Richard, J. P.; Gerlt, J. A. *Biochemistry* **2010**, *49*, 3514.