Supporting Information

Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for *Escherichia coli* Dihydrofolate Reductase

C. Tony Liu,[†,‡] Joshua P. Layfield,[‖,‡] Robert J. Stewart III,[†] Jarrod B. French,[±] Philip Hanoian,[†] John B. Asbury,[†] Sharon Hammes-Schiffer,*[,‖] Stephen J. Benkovic*[,†]

[†] Department of Chemistry, Pennsylvania State University, University Park, PA 16802, United States

[‖] Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3364, United States

± Department of Chemistry, Stony Brook University, Stony Brook, NY 11794-3400, United States

Table of Contents

Enzyme activity. The enzyme activity of the *ec*DHFR variants was evaluated by their ability to catalyze the NADPH-dependent reduction of DHF at pH 7.0 and 25 ^{o}C. Table S1 summarizes both the pre-steady state hydride transfer reaction rates and the steady state enzyme turnover rates. The enzyme activity of the unlabeled mutants was also compared. Furthermore, the kinetic data showed no discrepancy in the enzyme activity between using MTEM buffer or phosphate buffer, both at pH 7.0.

Table S1. Summary of the Pre-Steady State Hydride Transfer Rate Constant (k_{hyd}) and the Enzyme Turnover Rate Constant (k_{cat}) for the Various *ec*DHFR Constructs at pH 7.0 and 25 ^{o}C.

	WT[1]	ΔCys* T46C	ΔCys* T46C-CN	ΔCys* L54C	ΔCys* L54C-CN
k_{hyd}, s^{-1} (MTEM buffer)	220	165 ± 5	120 ± 10	215 ± 7	250 ± 20
k_{cat}, s^{-1} (MTEM buffer)	12	8.2 ± 0.6	6.5 ± 0.5	17.4 ± 0.5	28 ± 3
k_{hyd}, s^{-1} (phosphate buffer)	ND**	158 ± 2	131 ± 8	202 ± 9	200 ± 10
k_{cat}, s^{-1} (phosphate buffer)	ND**	7.5 ± 0.2	6.0 ± 0.4	16.6 ± 0.1	23 ± 1

* ΔCys *ec*DHFR corresponds to the C85A/C152S *ec*DHFR mutant

** ND = not determined

Protein crystallization, data collection and structure determination. Crystallization was performed by the hanging-drop vapor diffusion method at 20 °C. Drops were set up using approximately 25 mg/mL protein solution containing 10 mM Tris, pH 7.5, 1 mM methotrexate and 1 mM NADPH. Crystals formed after 3 – 4 days using equal volumes of protein solution and well solution containing 50 – 100 mM calcium acetate, 30 – 36% Peg 400 and 100 mM Hepes, pH 6.5. Crystals were harvested, briefly soaked in a solution of 100 mM calcium acetate, 40% Peg 400 and 100 mM Hepes, pH 6.5 containing 1 mM methotrexate and 1 mM NADPH (freshly

made), and flash frozen in liquid nitrogen. Data were collected at 100 K at the A1 beamline of the Cornell High Energy Synchrotron Source (CHESS). Data collection statistics are provided in Table S2.

Table S2. Crystallization Data Collection Statistics.[a]

	T46C-CN *ecDHFR*	L54C-CN *ecDHFR*
resolution (Å)	1.84	2.26
wavelength (Å)	0.987	0.987
beam line	CHESS A1	CHESS A1
space group	$P2_12_12_1$	$P2_12_12_1$
a (Å)	34.41	34.13
b (Å)	44.86	42.65
c (Å)	94.43	98.59
no. of reflections	52399	35131
unique reflections	12684	6501
average I/σ	14.0 (3.32)	15.4 (3.1)
redundancy	3.9 (3.5)	5.1(3.9)
completeness (%)	97.3 (93.8)	95.9 (89.3)
R_{sym} [b] (%)	10.7 (31.1)	15.1 (42.0)

[a]Numbers in parentheses correspond to the highest resolution shell

[b] $R_{sym} = \sum\sum_i |I_i - \langle I \rangle| / \sum \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the *N* reflections with intensities I_i and common indices *h, k, l*

Data processing, structure determination and refinement. The data were indexed, integrated and scaled using HKL2000.[2] The crystals contained two molecules per asymmetric unit and had an approximate solvent content of 54%. Molecular replacement was employed for phasing, using MOLREP[3] with the structure of *E. coli* DHFR (PDB code 1RH3) as the search model. The resulting structure was refined using alternating cycles of refinement using REFMAC5[4] and manual model building with Coot.[5] The addition of water molecules took place only after the refinement converged and was followed by an additional round of refinement. The ligands and modified cysteine residue were placed into difference density using the models available from the PDB (MTX, NADPH, and SCN) and were included in the model for a final round of refinement. Data refinement statistics are provided in Table S3. The structures have been deposited to the Protein Data Bank with accession codes 4P66 (T46C-CN) and 4P68 (L54C-CN).

Table S3. Data Refinement Statistics.

	T46C-CN *ec*DHFR	L54C-CN *ec*DHFR
resolution (Å)	50 – 1.84	50 – 2.26
no. of protein atoms	1239	1234
no. of ligand atoms	83	91
no. of water atoms	88	39
reflections in working set	12684	6501
reflections in test set	668	320
R factor[a] (%)	19.28	19.11
R_{free}[b] (%)	22.93	25.06
rmsd bonds (Å)	0.014	0.016
rmsd angles (^{o})	1.531	1.593
mean *B* factor (Å^2)	19.5	25.5
Ramachandran plot		
most favored (%)	99.4	98.0
additionally	0.6	1.4
allowed (%)		
disallowed (%)	0	0.6[c]

[a] $R = \sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl} \left| F_{obs} \right|$, where F_{obs} and F_{calc} are observed and calculated structure factors respectively.

[b] For R_{free}, the sum is extended over a subset of reflections (5.1 %) excluded from all stages of refinement.

[c] Pro21

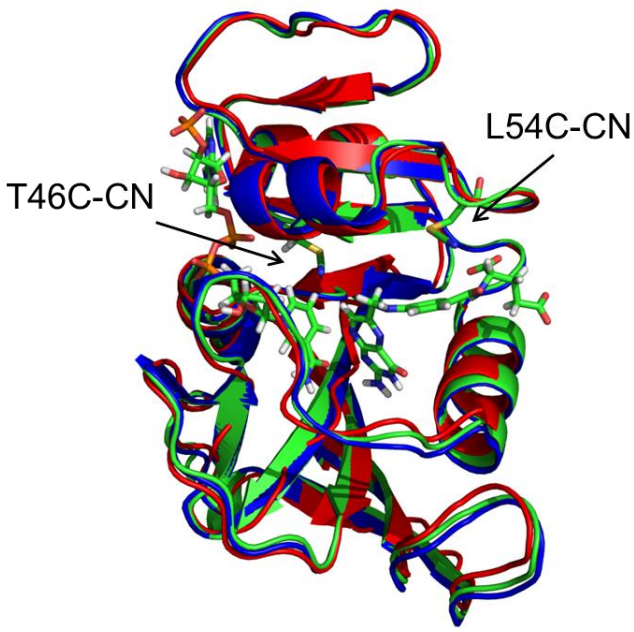

Figure S1. Superposition of the *ec*DHFR WT crystal structure (PDB Code:1RH3) with the NADPH and methotrexate ligands bound (red),[6] the T46C-CN mutant (PDB Code:4P66) with NADPH and methotrexate bound and the thiocyanate probe (green), and the L54C-CN mutant (PDB Code:4P68) with NADPH and methotrexate bound and the thiocyanate probe (blue). Only the ligands for the T46C-CN mutant are shown for the sake of clarity. The RMSDs for the backbone heavy atoms in the T46C-CN and L54C-CN mutants relative to the WT structure are 0.50 Å and 0.53 Å, respectively.

Table S4. Summary of ^{13}C NMR Shifts, Vibrational Frequency at Maximum Peak, and Full Width at Half Maximum (FWHM) Values for the IR spectra in two *ec*DHFR mutants.

L54C-CN Complexes			
	^{13}C Chemical Shift (ppm)	IR Frequency (cm^{-1})	FWHM (cm^{-1})
E	113.9 ± 0.1	2159.7	12.6 ± 0.1
E:NADPH	113.7 ± 0.0	2159.7	11.9 ± 0.1
E:NADPH:FOL	110.1 ± 0.1	2162.1	10.95 ± 0.08
E:$NADP^+$:FOL	110.1 ± 0.0	2162.1	6.43 ± 0.07
E:$NADP^+$:THF	110.1 ± 0.2	2161.1	6.2 ± 0.1
E:THF	110.2 ± 0.0	2160.9	8.08 ±0.08
E:NADPH:THF	110.2 ± 0.1	2160.6	6.61 ±0.08
T46C-CN Complexes			
	^{13}C Chemical Shift (ppm)	IR Frequency (cm^{-1})	FWHM (cm^{-1})
E	112.3 ± 0.2	2164.2	11.56 ± 0.09
E:NADPH	111.1 ± 0.0	2168.1	8.2 ± 0.1
E:NADPH:FOL	112.5 ± 0.2	2164.2	12.07 ± 0.03
E:$NADP^+$:FOL	113.1 ± 0.1	2164.0	14.1 ± 0.2
E:$NADP^+$:THF	113.6 ± 0.0	2159.9	11.8 ± 0.1
E:THF	112.9 ± 0.1	2161.1	12.2 ± 0.1
E:NADPH:THF	113.3 ± 0.0	2162.6	13.49 ± 0.06

FTIR spectroscopy. FTIR Spectra were obtained at room temperature using a Mattson Research Series FTIR Spectrometer (Madison Instruments, Inc.) equipped with a liquid-nitrogen cooled HgCdTe (MCT) detector (Infrared Associates, Inc.). Cyanylated DHFR samples were mixed with appropriate ligands (1:10 mole equivalents) in aqueous buffer and placed in a liquid cell with two CaF_2 windows separated by a 56μm Teflon spacer. All nitrile (CN) stretch absorbance spectra for cyanylated DHFR samples consisted of 2000 scans and were collected with 1.0 cm^{-1} resolution. The CN stretch absorbance spectra for cyanylated DHFR samples were obtained by subtracting a spectrum of the aqueous buffer solution from the protein spectrum. The baseline for each spectrum was corrected by fitting a 5^{th} order polynomial function to the data, as done previously by Fafarman[7] and McMahon[8]. The 5^{th} order polynomial local fit functions were obtained by defining roots at least15 cm^{-1} from the peak maximum, meaning the experimental data and fits were forced to be equal at points more than 15 cm^{-1} from the central frequency. All published spectra are the averages of two independently prepared protein samples after fitting and subtracting the baseline with the polynomial and normalizing the peak height to unit amplitude. The peak positions obtained from independently prepared protein samples typically differed by less than 0.5 cm^{-1}, substantially less than the changes among peak positions due to variation by protein binding and release of the substrate and cofactor.

The IR vs. ^{13}C NMR correlation data are shown in Figure S4. The species that appear to exhibit significant hydrogen bonding can be described by a linear expression (slope = -2.6 ± 0.5 cm^{-1}/ppm, $R^2 = 0.80$), which is within one standard deviation from that of EtSCN in non-hydrogen bonding solvents (slope = -1.7 ± 0.5 cm^{-1}/ppm, $R^2 = 0.68$;[9] dashed line). In water, hydrogen bonding between a water molecule and acetonitrile induces a 7 cm^{-1} blue shift in the nitrile stretching frequency.[10] This is consistent with the degree of shift we found in the nitrile

stretching frequency for all T46C-CN complexes, the binary E:NADPH complex of L54C-CN, and the L54C-CN apoenzyme.

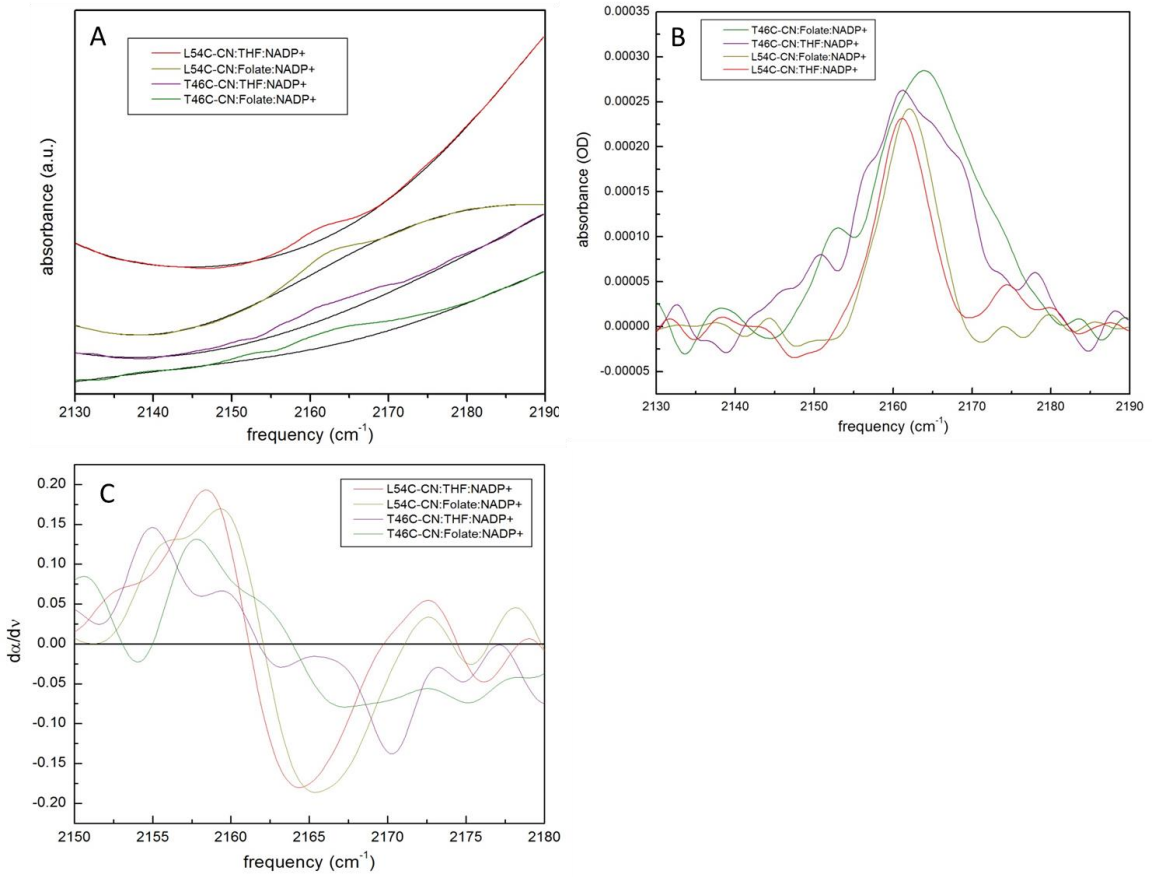

Figure S2. Details of FTIR signal processing and analysis. The CN stretch absorbance spectra for the protein samples on top of large baselines arising from the bend + libration combination band of water (A). The black curves correspond to local best fit 5th order polynomials with roots defined more than 15 cm^{-1} from peak maximum of each spectrum. (B) represents the CN stretch band following subtraction of the best fit functions from raw absorbance spectra. The center frequency has little dependence on the points at which polynomial and data are forced to be equal in our analysis scheme (see Figure S3). The first derivative spectra (C) were calculated from the baseline-corrected CN absorption spectra appearing in Panel B. The nodes in the first derivative spectra correspond to the peaks of the absorption spectra (where the derivate is zero). This analysis scheme was used to identify the center vibrational frequencies of the *ec*DHFR-CN stretch modes with negligible influence from the 5th order polynomial function or background subtraction procedure (see Figure S3).

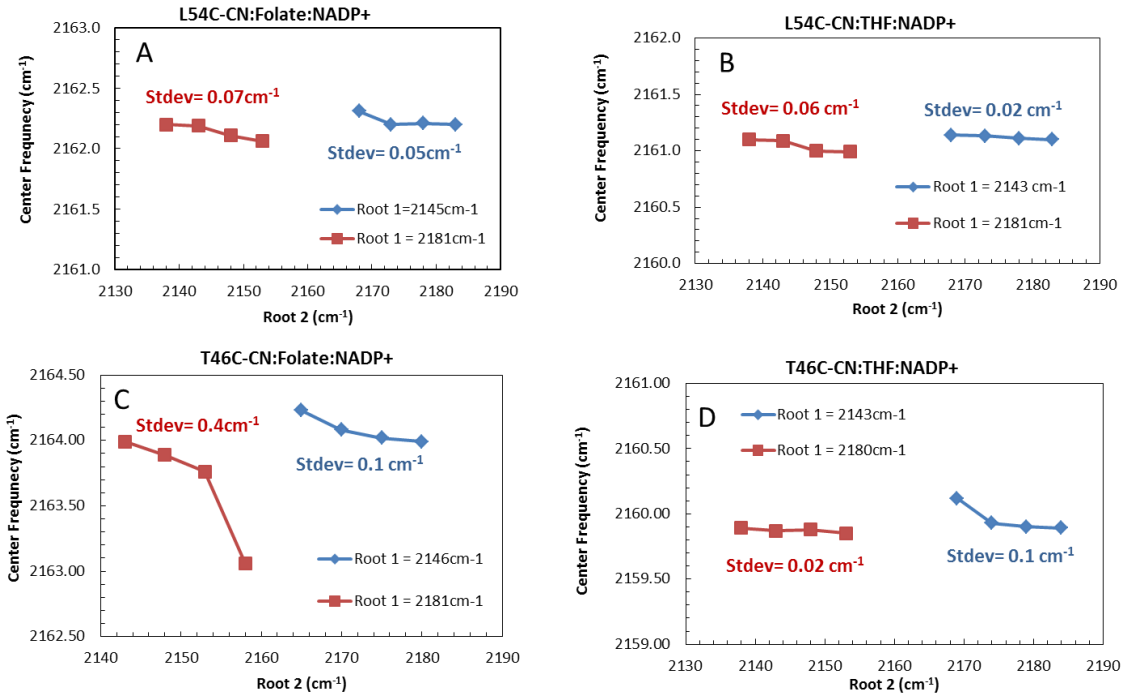

Figure S3. Analysis of the sensitivity of the center frequency with definition of 5th order polynomial roots. To quantitatively assess the variation of the center frequencies obtained from the CN stretch peaks versus locations of where the 5th order polynomial is forced to equal the baseline, we systematically varied the position of the roots on either side of the CN stretch peak and plotted the corresponding center frequencies for each protein, substrate and cofactor complex. Panels A – D represent the results of calculating the center frequencies of the CN stretch peaks while holding one root position constant and systematically varying the other root. The root that is held constant is labeling Root 1 in the panels, and the root that is varied is labeled Root 2. Symmetric variation of both roots simultaneously had no effect on the center frequencies obtained from our analysis scheme. We performed this analysis for both roots: we fixed the high frequency root and varied the low frequency one (data appear as squares); we also varied the high frequency root and fixed the low frequency one (data appear as diamonds). Panels A, B, C, and D show the variation in the center frequency for a variety of root positions for L54C-CN:FOL:NADP$^+$, L54C-CN:THF:NADP$^+$, T46C-CN:FOL:NADP$^+$, and T46C-CN:THF:NADP$^+$ samples, respectively. The standard deviation in the center frequency for each root position is displayed in the figure, with the largest variation in center frequency of 0.4 cm^{-1} and most variations less than 0.05 cm^{-1}, showing that the center frequency is insensitive to the position of the polynomial roots in comparison to the variation among peak positions due to variation by protein binding and release of the substrate and cofactor. It is noteworthy that the largest variation of the center frequency with the location of roots occurs in the T46C-CN:FOL:NADP$^+$ complex, which has the greatest line width. The lowest extracted frequency in Panel C occurs where the low frequency root is actually within the line width of the CN stretch. Therefore, the experimental uncertainty in frequency is actually substantially less than the 0.4 cm^{-1} standard deviation indicated in Panel C.

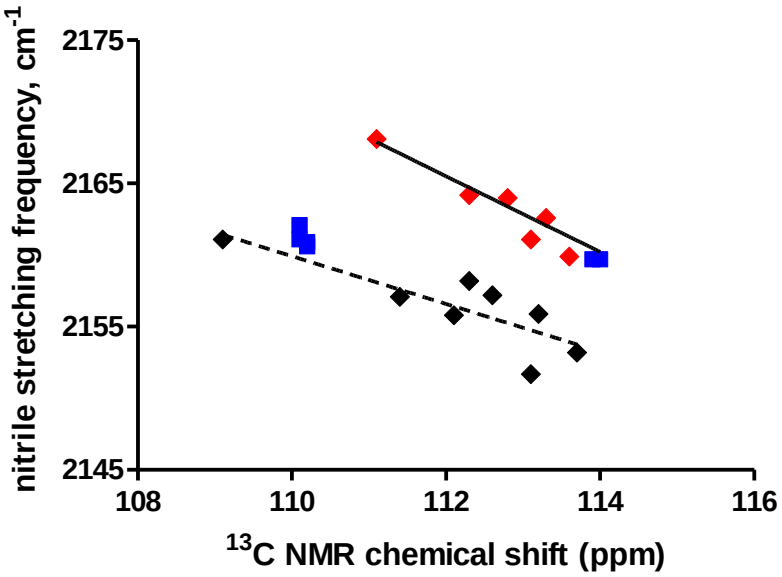

Figure S4. Plot of nitrile vibrational stretching frequencies versus the ^{13}C NMR chemical shifts for the labeled *ec*DHFR variants (T46C-CN: ♦ and L54C-CN: ■). The dashed line and the black solid data points are from a previous publication illustrating the expected trend for EtSCN in non-hydrogen bonding solvents.[9] The solid line (slope = –2.6 ± 0.5, R^2 = 0.8008) is fitted to all data points that exhibit significant deviations (~ 7 cm^{-1}), which can be attributed to specific hydrogen-bonding interactions between the CN probe and a hydrogen bond donor.

QM/MM parameterization. We reparameterized the PM3 method using a model system that contains a methyl thiocyanate (MeSCN) molecule solvated in 2900 TIP3P water molecules. We propagated a 5 ns molecular dynamics (MD) trajectory and calculated the vibrational frequency of the nitrile stretching mode for configurations sampled at intervals of 10 ps (500 total) using the methodology described in the main paper. The QM region in these calculations included the 12 water molecules closest to the CN probe, determined by the distance between the nitrile nitrogen and the water oxygen, for each configuration sampled in the MD trajectory. The remaining water molecules were treated as atomic point charges using the partial charges from the TIP3P water model. The QM region was treated at the B3LYP/6-311++G(d,p) level of theory as implemented in Gaussian09,[11] and the MM force field for the MeSCN utilized the parameters given in our previous work.[12] After generating the one-dimensional potential energy curve along the CN axis, the nitrile vibrational frequencies were calculated using the Fourier Grid Hamiltonian method.[13,14]

We used a non-linear-least-squares method to minimize the root-mean-square deviation (RMSD) between a target function calculated with DFT and with the reparameterized PM3 method by optimizing the 18 free nitrogen parameters in the PM3 method and two parameters that describe the interaction between the QM and MM regions.[15] The target function that was minimized was a weighted sum of the calculated potential energies relative to the minimum energy for the CN potential energy curve and the associated vibrational frequency. The potential energy curve was represented by nine points evenly spaced along a grid,[12] and the target function for each configuration weighted each of these points equally and multiplied the associated vibrational frequency by nine, which was the number of points along the potential energy curve. The full target function that was minimized was averaged over the 500 configurations sampled.

The advantage of simultaneously fitting both the calculated frequency and the QM/MM relative potential energies, rather than only the frequency, is that the equilibrium CN bond length as well as the frequency is fit during the reparameterization.

The optimized PM3 and QM/MM parameter set is given in Table S5. While we did not constrain the parameters during the optimization, all of the PM3 parameters are within 12% of the original values, and the average deviation is less than 7%. The α parameter that governs the QM/MM interactions varies by ~50% from the original value. This value is similar to that obtained in previous work of Skinner and co-workers[16] as well as in our previous reparameterization.[12] The other parameter that controls QM/MM interactions, ρ, varies only slightly from default value of zero.

The optimized parameter set reproduced the vibrational frequencies calculated with DFT with a mean unsigned error of 0.07 cm^{-1} and an RMSD of 5.6 cm^{-1}. This level of agreement between the DFT frequencies and those calculated with the reparametrized PM3 method is similar to that in our previous parameterization.[12] In our new parameterization scheme, we also simultaneously fit the relative energies along the one-dimensional potential energy curve for each configuration. The mean-unsigned error and RMSD for these energies are 0.28 and 1.85 kcal/mol, respectively. Furthermore, the average CN bond length from the reparametrized PM3 calculations is 1.17 Å, which is in excellent agreement with the average CN bond length of 1.16 Å obtained with DFT. This agreement contrasts with the average CN bond length of 1.21 Å from our previous work,[12] which did not directly include any information about the underlying potential energy curve in the fitting procedure. In addition, the potential energy curve was fit to a Morse potential, rather than utilizing the Fourier Grid Hamiltonian method, to calculate the vibrational frequencies in our previous work.[12] We confirmed that the new parameter set, in

conjunction with the Fourier Grid Hamiltonian method, reproduces the vibrational frequency shifts calculated in our previous work on the enzyme ketosteroid isomerase (Table S6).

Table S5. Original and Optimized PM3 Nitrogen Parameters and QM/MM Interaction Parameters.

Parameter	Original	Optimized
U_{ss} (eV)	-49.3357	-52.400
U_{pp} (eV)	-47.5097	-46.906
β_s (eV)	-14.0625	-14.563
β_p (eV)	-20.0438	-18.045
ζ_s (eV)	2.0281	2.0520
ζ_p (eV)	2.3137	2.6104
α ($Å^{-1}$)	2.8305	2.9084
G_{ss} (eV)	11.905	10.4908
G_{sp} (eV)	7.3486	7.1824
G_{pp} (eV)	11.7557	12.3479
G_{p2} (eV)	10.8072	10.7442
H_{sp} (eV)	1.1367	1.0723
a_1 (Å)	1.5017	1.4093
b_1 (Å)	5.9011	6.6059
c_1 (Å)	1.7107	1.7561
a_2 (Å)	-1.5058	-1.3919
b_2 (Å)	6.0047	6.5365
c_2 (Å)	1.7161	1.7644
α ($Å^{-1}$)	5.0000	2.3246
ρ (Å)	0.0000	-0.003

Table S6. Nitrile Stretching Frequency of the Maximum Peak for Simulated IR Spectra of Two Mutants of Ketosteroid Isomerase using Two Parameter Sets[a]

M116C-CN			
	Deprotonated Site[b]	New Parameters[c]	Old Parameters[d]
No Ligand	Tyr57	2159.5	2162.1
EQU	EQU	2161.5	2164.1
EQU	Tyr16	2162.6	2165.1
F86C-CN			
	Deprotonated Site	New Parameters	Old Parameters
No Ligand	Tyr57	2177.4	2186.8
EQU	EQU	2178.7	2187.8
EQU	Tyr16	2179.3	2188.5

[a]Units are cm^{-1}. Absolute frequencies differ between the old and new parameter sets, but the shifts between no ligand and bound equilinen (EQU) for each mutant are nearly identical for the two parameter sets.

[b]The location of the negatively charged site in the KSI active site on either the indicated tyrosine residue or the EQU ligand.

[c]The parameter set obtained and used in this work. The frequencies are calculated using the Fourier Grid Hamiltonian method described in this work.

[d]The parameter set published in our previous work on KSI.[12] The frequencies are calculated by fitting the potential energy curve to a Morse potential as described in that work.

Table S7. CN Vibrational Frequencies and Electric Fields Calculated from Two Independent Trajectories for each *ec*DHFR mutant.[a]

Construct	T46C-CN		L54C-CN	
	Traj 1	Traj 2	Traj 1	Traj 2
ν_{max} ($NADP^+$/FOL)	2165.8	2165.9	2167.5	2167.1
ν_{max} ($NADP^+$/THF)	2162.1	2162.4	2165.0	2164.3
$\Delta\nu$	3.7	3.5	2.5	2.8
E_{CN} ($NADP^+$/FOL)	-15.9	-15.2	-8.4	-8.8
E_{CN} ($NADP^+$/THF)	-19.0	-19.3	-13.8	-14.6
ΔE_{CN}	3.1	4.1	5.4	5.8
Predicted $\Delta\nu$	2.2	2.9	3.8	4.1
ν_{max} ($NADP^+$/FOL) Probe only	2162.7	2163.0	2165.5	2165.1
ν_{max} ($NADP^+$/THF) Probe only	2160.1	2160.3	2162.0	2161.3
$\Delta\nu$	2.6	2.7	3.5	3.8

[a]Units of frequency are cm^{-1} and units of electric field are MV/cm.

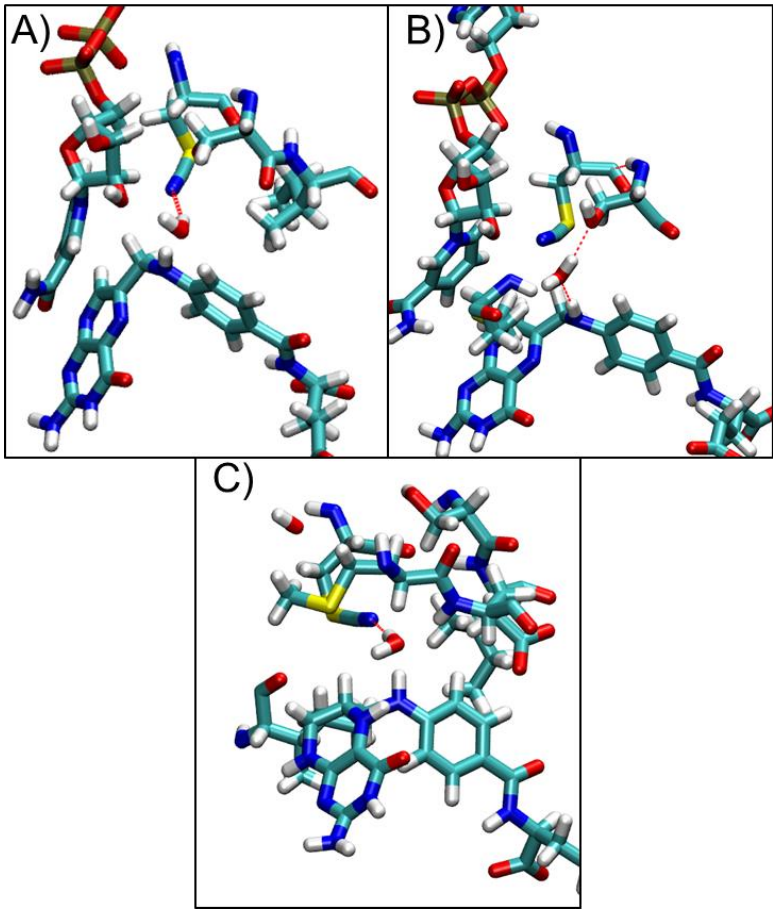

Figure S5. Representative snapshots from an MD trajectory showing three different water orientations relative to the thiocyanate probe in the T46C-CN mutant of *ec*DHFR. (A) A water molecule hydrogen bonding with the thiocyanate probe in the closed conformation. (B) A water molecule hydrogen bonding with the hydroxyl group of Ser49, with the oxygen atom oriented toward the nitrile probe in the closed conformation. (C) A water molecule hydrogen bonding with the thiocyanate probe and positioned in the pocket vacated by the cofactor in the occluded conformation.

QM/MM partitioning. As shown in our previous work,[12] the quantum mechanical effects of hydrogen bonding and polarization must be included in the QM/MM frequency calculations to properly describe the experimentally observed shifts. To determine which residues to include in the QM region, we employed a scheme in which the CN stretching frequencies were calculated for a subset of configurations with all of the proximal residues in the probe microenvironment treated quantum mechanically. These benchmark calculations were then used to determine which residues are important to include in the QM region.

For the T46C-CN mutant, we considered the entire probe microenvironment to consist of the residues Asn18, Ser49, Ile50, and the ligand bound in the substrate pocket (FOL/THF). Additionally, since our MD simulations showed that the T46C-CN probe is solvent accessible, we also included the solvent molecule closest to the nitrogen of the nitrile group. Investigation of treating additional solvent molecules quantum mechanically showed that the vibrational frequencies were converged with only a single water molecule in the QM region. Table S8 provides the root-mean-square deviations (RMSDs) of the nitrile stretching frequencies calculated with the indicated QM regions relative to those calculated with the entire probe microenvironment. Treatment of the nearest solvent molecule quantum mechanically reduces the RMSD dramatically for both the closed and occluded conformations. During the MD trajectory, a water molecule was hydrogen bonded to the nitrile probe in the closed and occluded conformations for 25% and 50% of the MD trajectories, respectively. Inclusion of the substrate or product (FOL/THF) in the QM region also reduces the RMSD significantly, most likely because the amino group acts as a hydrogen bond donor either directly to the nitrile probe or to the water molecule that is hydrogen bonded to the probe. The negligible impact of treating Asn18, Ser49, or Ile50 quantum mechanically indicates that these residues do not need to be

treated quantum mechanically and can be reasonably approximated as point charges. On the basis of these studies, we included the nearest water molecule and the substrate/product ligand, as well as the probe residue, in the QM region for our vibrational frequency calculations for the T46C-CN mutant. This QM region resulted in an RMSD of 0.92 cm^{-1} and 1.06 cm^{-1} in the closed and occluded conformations, respectively, relative to treating the entire probe microenvironment quantum mechanically.

For the L54C-CN mutant, the entire probe microenvironment consisted of Leu28, Phe31, Val40, Ile94, the substrate/product ligand, and Arg 57, a positively charge residue that forms a salt bridge with one of the two carboxylate groups of the bound ligand. Table S8 provides the RMSDs of the vibrational frequencies calculated with the indicated QM regions relative to those calculated with the entire probe microenvironment. Including the aliphatic residues Leu28, Val40, and Ile94 in the QM region does not significantly reduce the RMSD. The aromatic Phe31 residue reduces the RMSD when it is included in the QM region, indicating that the close proximity of the phenyl ring to the nitrile probe exerts polarization effects. Inclusion of the positively charged Arg57 in the QM region significantly reduces the RMSD of the calculated frequencies. Surprisingly, including the substrate/product ligand in the QM region does not significantly impact the calculated frequencies, despite the negative formal charge on its carboxylate groups and the presence of a salt bridge with Arg57. Including both the Phe31 and Arg57 residues in the QM region appears to have a cooperative effect by strongly reducing the RMSD of the calculated frequencies, most likely due to the interaction of the polarizable phenyl ring in Phe31 with the positively charged Arg57 side chain. On the basis of these studies, we included Phe31 and Arg57, as well as the probe residue, in the QM region for our vibrational frequency calculations for the L54C-CN mutant. This QM region resulted in an RMSD of 1.13

cm^{-1} and 1.01 cm^{-1} in the closed and occluded conformations, respectively, relative to treating the entire probe microenvironment quantum mechanically.

Table S8. RMSDs of the Calculated CN Frequencies for Multiple QM Regions Relative to the Entire Probe Microenvironment.

	RMSD (cm^{-1})[a]	
QM Residue	Closed	Occluded
	T46C-CN[b]	
probe only	4.04	2.72
H_2O	1.81	1.61
Asn18	2.92	2.71
Ser49	2.29	1.85
Ile50	2.97	2.14
Ligand	1.57	2.71
Ligand / H_2O	0.92	1.06
	L54C-CN[c]	
probe only	2.75	3.51
Phe31	2.41	3.10
Val40	2.86	3.87
Arg57	2.21	1.9
Ile94	2.66	3.55
Ligand	2.75	3.51
Phe31 / Arg57	1.13	1.01

[a]RMSD among a set of 500 configurations. All calculations include the full probe residue in the QM region.

[b]The entire probe microenvironment for the T46C-CN mutant includes the thiocyanate probe, Asn18, Ser49, Ile50, the nearest water molecule, and the ligand bound in the substrate binding pocket (FOL/THF).

[c]The entire probe microenvironment for the L54C-CN mutant includes the thiocyanate probe, Leu28, Phe31, Val40, Ile50, Arg57 and the ligand bound in the substrate pocket (FOL/THF).

Table S9. Calculated Contributions from Specific Residues to the Electric Field along the Hydride Transfer Donor-Acceptor Axis for WT *ec*DHFR in the Closed Conformation ($NADP^+$/FOL).

Residue	E_{CN}, MV/cm[a]
Ile5	–1.4
Ala6	1.5
Ala7	–1.2
Asp11	–1.9
Arg12	1.7
Ile14	–1.2
Glu17	–1.4
Asp27	5.4
Lys32	–3.2
Arg33	–2.1
Asp37	1.1
Lys38	–1.4
Arg44	1.6
Arg52	–3.9

Residue	E_{CN}, MV/cm
Arg57	–4.2
Arg71	–1.0
Ile94	–3.5
Gly95	2.3
Gy96	1.1
Arg98	3.4
Tyr100	–4.2
Glu101	–2.3
His114	1.2
Glu120	–1.1
Asp122	–2.4
His124	3.0
Asp127	–1.6
Ligands	–32.4

[a]Component of the electric field along the donor-acceptor axis. Only residues contributing to the field with a magnitude greater than 1.0 MV/cm are listed. The sum of the values given in this table is -48.0 MV/cm, in comparison to the total field along this axis of –48.7 MV/cm, which includes contributions from other residues and solvent molecules.

References

(1) Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085.

(2) Otwinowski, Z.; Minor, W. *Method Enzymol* **1997**, *276*, 307.

(3) Vagin, A.; Teplyakov, A. *J Appl Crystallogr* **1997**, *30*, 1022.

(4) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355.

(5) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126.

(6) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586.

(7) Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. *Proc Natl Acad Sci U S A* **2012**, *109*, E299.

(8) McMahon, H. A.; Alfieri, K. N.; Clark, K. A.; Londergan, C. H. *J Phys Chem Lett* **2010**, *1*, 850.

(9) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. *J. Am. Chem. Soc.* **2010**, *132*, 12811.

(10) Reimers, J. R.; Hall, L. E. *J Am Chem Soc* **1999**, *121*, 3730.

(11) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Wallingford CT, 2009.

(12) Layfield, J. P.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2013**, *135*, 717.

(13) Marston, C. C.; Balint-Kurti, G. G. *J. Chem. Phys.* **1989**, *91*, 3571.

(14) Balint-Kurti, G. G.; Ward, C. L.; Marston, C. C. *Comput. Phys. Commun.* **1991**, *67*, 285.

(15) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(16) Li, S. Z.; Schmidt, J. R.; Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. *J. Chem. Phys.* **2006**, *124*, 204110.