# SUPPORTING INFORMATION
## Carbon-Deuterium Bonds as Probes of Electrostatics in Dihydrofolate Reductase

Megan C. Thielges, David A. Case, and Floyd E. Romesberg
*Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California, 92037.*

### *Sample preparation for solvent studies*

*(Methyl-$d_3$)* methionine (Met-$d_3$) (Cambridge Isotopes) was tert-butoxycarbonyl (BOC) protected to increase solubility in nonpolar solvents (1). 50 mM solutions of BOC Met-$d_3$ were prepared in water, isopropanol, mixtures of water and isopropanol, as well as in methanol, ethanol, *t*-butanol, benzyl alcohol, tetrahydrofuran (THF), *p*-dioxane, and toluene (Fisher or Sigma).

### *Protein Expression*

A pET22b-derived plasmid for the expression of dihydrofolate reductase (DHFR) was obtained from the Wright laboratory (The Scripps Research Institute). We constructed DHFR mutants that possessed only a single Met residue (Met1) or two Met residues (combinations of Met1 and Met16, Met20, or Met42), with the other Met residues mutated to leucine. Each protein retained the residue Met1, as it is required for translation. Mutagenesis was performed using the Quickchange protocol (Stratagene) and confirmed by DNA sequencing (TSRI nucleic acid core facility). DHFR was expressed in *E. coli* BL21(DE3) at 37 °C in the presence of 100 μg/mL ampicillin. Cells were grown in 2 L M9 minimal media containing 1 mM $MgSO_4$, 0.2% dextrose, 400 mg/L amino acids, and 1× Gibco MEM vitamin solution (Invitrogen). Isotopic labeling was accomplished by substitution with 50 mg/L Met-$d_3$. Protein expression was induced at $OD_{600}$ of 0.6-0.8 with 1 mM IPTG, and cells were grown 4-6 hr before harvesting.

The purification of DHFR was based on published protocol (2), and was performed at 4 °C. The cell pellet was resuspended in Ar-purged 50 mM Tris-Cl, pH 8.0, 1 mM EDTA, 1 mM DTT and treated with 3 mg/L lysozyme. The resuspension was brought to 0.4% Brij 58 (Anapoe) by the addition of a 10% solution in water and to 3 mM PMSF by the addition of a 100 mM stock in isopropanol. Cells were lysed by sonication, followed by clarification of the lysate by centrifugation at 27,000 x g for 60 min. DNA was removed through precipitation with 10 mg/mL streptomycin sulfate, followed by centrifugation at 27,000 x g for 40 min. The final lysate was dialyzed against 25 mM sodium phosphate, pH 7.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT and run in a loop over a MTX-agarose affinity column (Pharmacia) overnight. After washing the column with the above buffer, DHFR was eluted in 100 mM boric acid, 500 mM NaCl, 1 mM EDTA, 3 mM folate, 1 mM DTT. Pooled fractions were extensively dialyzed against 40 mM Tris-Cl, pH 7.5, 1 mM EDTA, 1 mM DTT to remove both salt and folate. The protein was further purified on a DEAE sepharose column (Bio-Rad), eluting with a 200-500 mM NaCl gradient. Fractions containing DHFR, as determined by SDS-PAGE, and were free of folate, as determined by no 363 nm absorbance, were pooled and dialyzed against 50 mM potassium phosphate, pH 7.0, 100 mM KCl, 1 mM DTT. The DHFR solutions were filtered and concentrated to 3.5-4 mM prior to use.

Several routes for removing the N-terminal methionine were explored in order to produce DHFR containing only a singly labeled methionine. A Factor X sequence was inserted at the N-terminus, but no cleavage occurred upon incubation with the enzyme. The DHFR sequence was placed after the intein cleavage site in the pTWIN1 construct of the IMPACT system (New

England Biolabs), but no cleavage was found. The penultimate residue, isoleucine, was mutated to glycine in attempt to utilize native methionine aminopeptidase, which preferentially cleaves N-terminal methionine residues preceding small residues (3). Unfortunately, this resulted in protein precipitation. We also attempted the semisynthesis of DHFR, which would not only have allowed the removal N-terminal methionine but would have enabled easy deuteration of further DHFR residues. Although we were able to produce both N-terminal and C-terminal pieces of DHFR that are amenable to native chemical ligation, as yet no ligation product has been obtained.

***Preparation of DHFR samples***

DHFR was prepared in complexes that mimic the intermediate states of the catalytic cycle (4). The holoenzyme is the NADPH binary complex, the Michaelis complex was modeled by the folate/$NADP^+$ complex, the transition state was modeled by the MTX/NADPH tertiary complex, and the THF binary product complex was modeled with the folate binary complex. DHFR complexes were formed by the addition of 200-300 mM stock solutions of $NADP^+$ and NADPH in $H_2O$ and approximately 150 mM stock solutions of folate and MTX in 400 mM NaOH. The $NADP^+$, folate, and MTX solutions were adjusted to pH 7 prior to use. Concentrations of DHFR and all compounds were determined spectrophotometrically using the following molar extinction coefficients: DFHR, $\varepsilon_{280}$ 31,100 $M^{-1}cm^{-1}$, MTX $\varepsilon_{302}$ 22,100 $M^{-1}cm^{-1}$, NADPH $\varepsilon_{339}$ 6,200 $M^{-1}cm^{-1}$, folate $\varepsilon_{282}$ 27,600 $M^{-1}cm^{-1}$, $NADP^+$ $\varepsilon_{259}$ 18,000 $M^{-1}cm^{-1}$. The final DHFR concentrations were 3 or 3.5 mM. The NADPH binary complex was formed by the addition of 2-fold excess NADPH. The MTX/NADPH complex sample contained 2-fold and 1.3-fold excess NADPH and MTX, respectively. $NADP^+$ and folate were added to 3.3-fold excess in the $NADP^+$/folate complex sample, while 6.7-fold excess folate was used to form the folate binary complex. At these concentrations less that 0.5% of DHFR should remain unbound according to the $K_d$ values for the various compounds (5,6). The introduced mutations are not expected to have catastrophic effects on binding, especially at the high concentrations used, as replacement of methionine residues by leucine in DHFR has been found to decrease the $k_{cat}$ and $K_M$'s for DHF and NADPH by at most four-fold (7).

***Data acquisition and analysis***

The spectra of DHFR containing only $d_3$Met1 were well fit by a Gaussian function (see Figure S1 below). These spectra were used to deconvolute the absorption of $d_3$Met1 from those of $d_3$Met16, $d_3$Met20, $d_3$Met42. Specifically, in the apoenzyme and each of the complexes, three spectra of DHFR containing only $d_3$Met1 were acquired, baseline-corrected, and fit to a single Gaussian function. Averages and standard deviations of the Gaussian parameters were then calculated from the individual fits. To account for $d_3$Met1 in the spectra of residues $d_3$Met1/$d_3$Met16, $d_3$Met1/$d_3$Met20 and $d_3$Met1/$d_3$Met42 labeled proteins, a Gaussian component was included in the fit with parameters restricted to values within a standard deviation of the averages found for $d_3$Met1 DHFR in the corresponding complex.

***BOC-Met-$d_3$ in various solvents***

**Table S2**: The dielectric constants (D), square of the refractive indices ($n^2$), acceptor numbers (AN), and frequency of the Met-$d_3$ symmetric stretch ($\nu_{obs}$, in $cm^{-1}$) for the different solvents (9,10).

| Solvent | D | $n^2$ | AN | $\nu_{obs}$ |
|---|---|---|---|---|
| H2O | 78.4 | 1.78 | 54.8 | 2135.4 |
| Methanol | 32.7 | 1.77 | 41.3 | 2130.6 |
| Ethanol | 24.6 | 1.85 | 37.9 | 2129.4 |
| Isopropanol | 18.3 | 1.89 | 33.8 | 2129.1 |
| benzyl alcohol | 13.1 | 2.37 | 34.5 | 2127.4 |
| t-butanol | 10.8 | 1.91 | 29.1 | 2129.4 |
| THF | 8.0 | 1.97 | 8.0 | 2128.1 |
| tolulene | 2.4 | 2.24 | 10 | 2127.3 |
| p-dioxane | 2.2 | 2.02 | 10.8 | 2127.9 |

The following equations describing dependence of vibrational frequencies on dielectric constant or refractive index were used in attempt to understand our BOC Met-$d_3$ spectra in various solvents (9,11). $\nu_0$ is the frequency in the vapor phase, $\nu_s$ is the frequency in solvent with dielectric constant, D, or refractive index, n; C is a constant.

Kirkwood-Bauer-Magat (KBM) equation:

$$\frac{(\nu_0 - \nu_s)}{\nu_s} = \frac{\Delta\nu}{\nu} = C\frac{D-1}{2D+1}$$

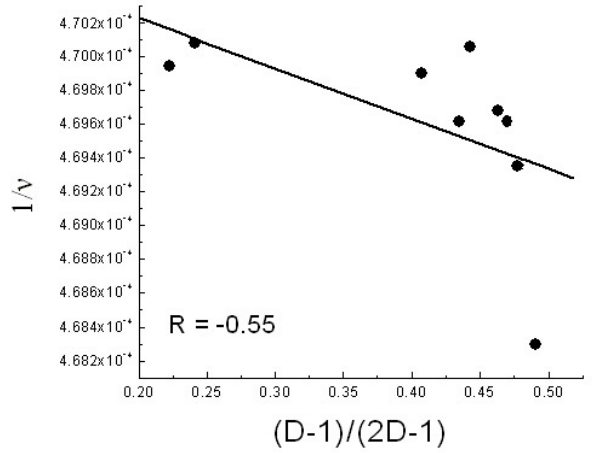
Modified KBM equation:

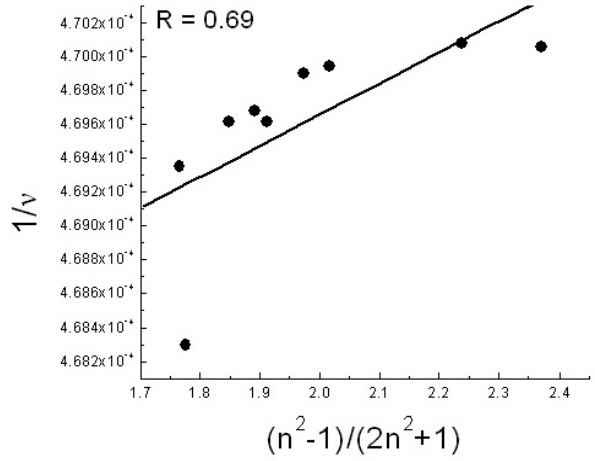
$$\frac{\Delta\nu}{\nu} = C\frac{n^2-1}{2n^2+1}$$

These equations were rearranged for a linear plot of $1/\nu_s$ against (D-1)/(2D+1) or $(n^2-1)/(2n^2+1)$:

$$\frac{1}{\nu_s} = \frac{C}{\nu_0}\frac{D-1}{2D+1} + \frac{1}{\nu_0}$$

$$\frac{1}{\nu_s} = \frac{C}{\nu_0}\frac{n^2-1}{2n^2+1} + \frac{1}{\nu_0}$$



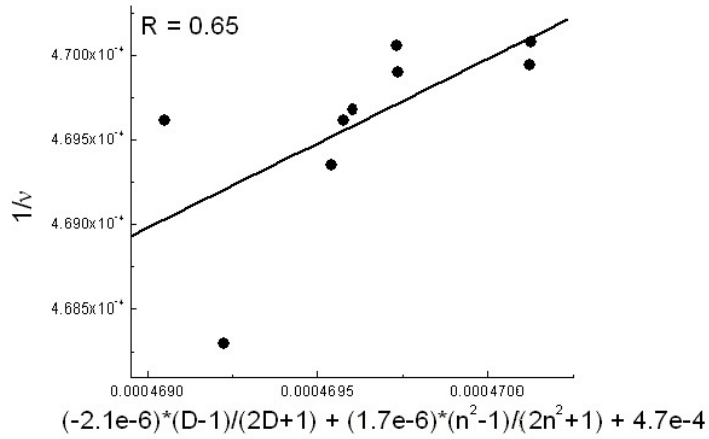

As the KMB relationships did not adequately describe our data, we also tried the following equation, presented by Buckingham to account for more polar solvents (11):

$$\frac{\Delta\nu}{\nu} = C + C_D\frac{D-1}{2D+1} + C_n\frac{n^2-1}{2n^2+1}$$

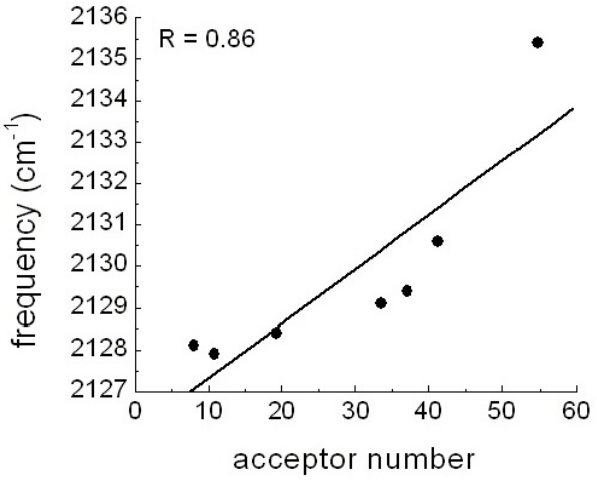
This can only be rearranged to

$$\frac{1}{\nu_s} = \frac{C_D}{\nu_o}\frac{D-1}{2D+1} + \frac{C_n}{\nu_o}\frac{n^2-1}{2n^2+1} + \frac{1}{\nu_o}$$

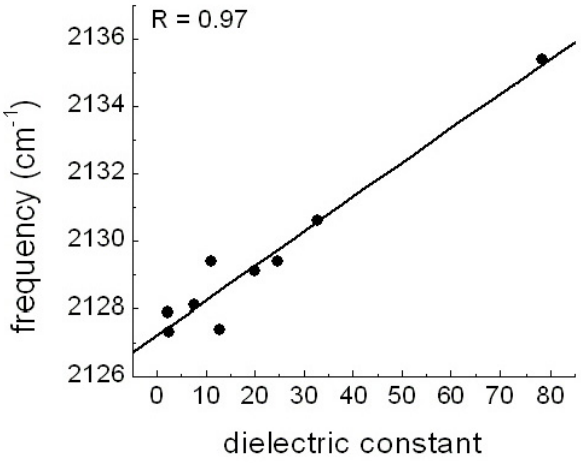
Minimization of the difference in measured frequencies and those calculated from the known dielectric constants and refractive indices with respect to unknown parameters $C_D/\nu_o$, $C_n/\nu_o$, and $1/\nu_o$ yielded parameter values of $-2.1\times10^{-6}$, $1.7\times10^{-6}$, and $4.7\times10^{4}$, respectively. Using these values, the following plot was obtained:



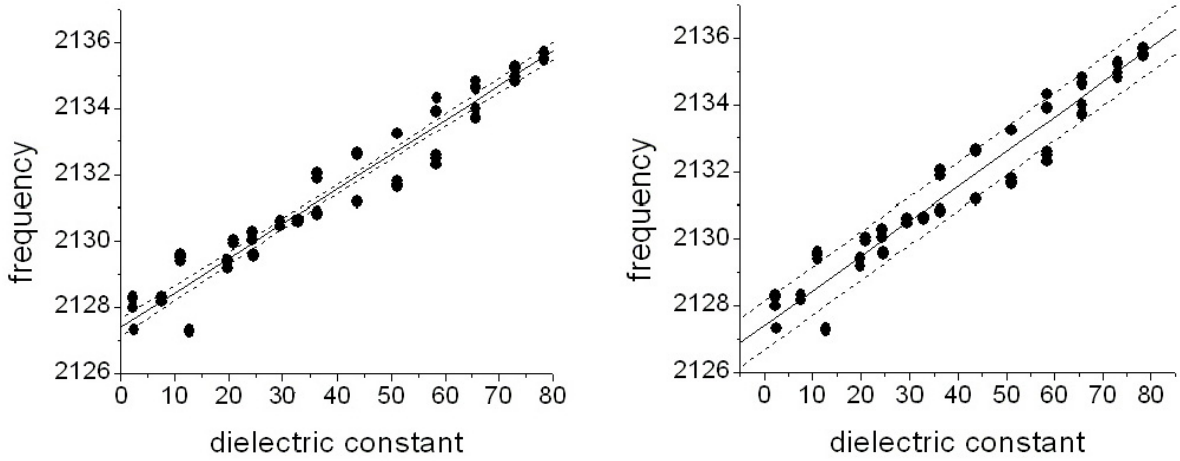

Previous studies have found solvent-induced frequency shifts to linearly correlate with the solvent acceptor numbers (10,12); however, a linear fit to our data was poor.




Although none of the above relationships adequately described the solvent-induced frequency shifts we measured for BOC-Met-$d_3$, a plot of the measured frequencies against solvent dielectric constant is remarkably linear.




Furthermore, we also measured the Met-$d_3$ spectrum of both free and BOC-protected amino acid in varying solutions of water and isopropanol, whose dielectric constants are also available (13). This data also followed a linear trend (as shown in Figure 1B of the main text). Shown below is the same data including the calculated confidence limits (95% confidence level, left) and prediction limits (75% confidence level, right).




If the combined free and Boc-protected amino acid data is used as an empirical reference, the Met16 and Met20 side chains in the apoenzyme are predicted to experience an environment approximately equivalent to a solvent dielectric of 24 and 13, respectively. Using the standard deviations in frequency and the calculated 95% confidence intervals we can estimate the error of the predicted dielectric constants to be ±3.5 and ±6.0, respectively, however, the use of the more rigorously correct prediction intervals at the 75% confidence level yields values of ±8.0 and ±11.5, respectively. The difference discussed in the text however are much larger. For example,

for Met16 in the apoenzyme and in the NADPH complex the predicted apparent dielectric constants are 24±3.5 (±8.0) and 50±4.2 (±9.7), where the confidence interval error is indicated and the prediction interval error is include in parenthesis. Thus, while greater accuracy may be obtainable with a better understanding of the forces contributing to the observed shifts (as discussed in the main text), the clear correlation supports the empirical interpretation of the observed frequencies in terms of environmental polarity.

***Fits of $d_3$ Met1 labeled DHFR***

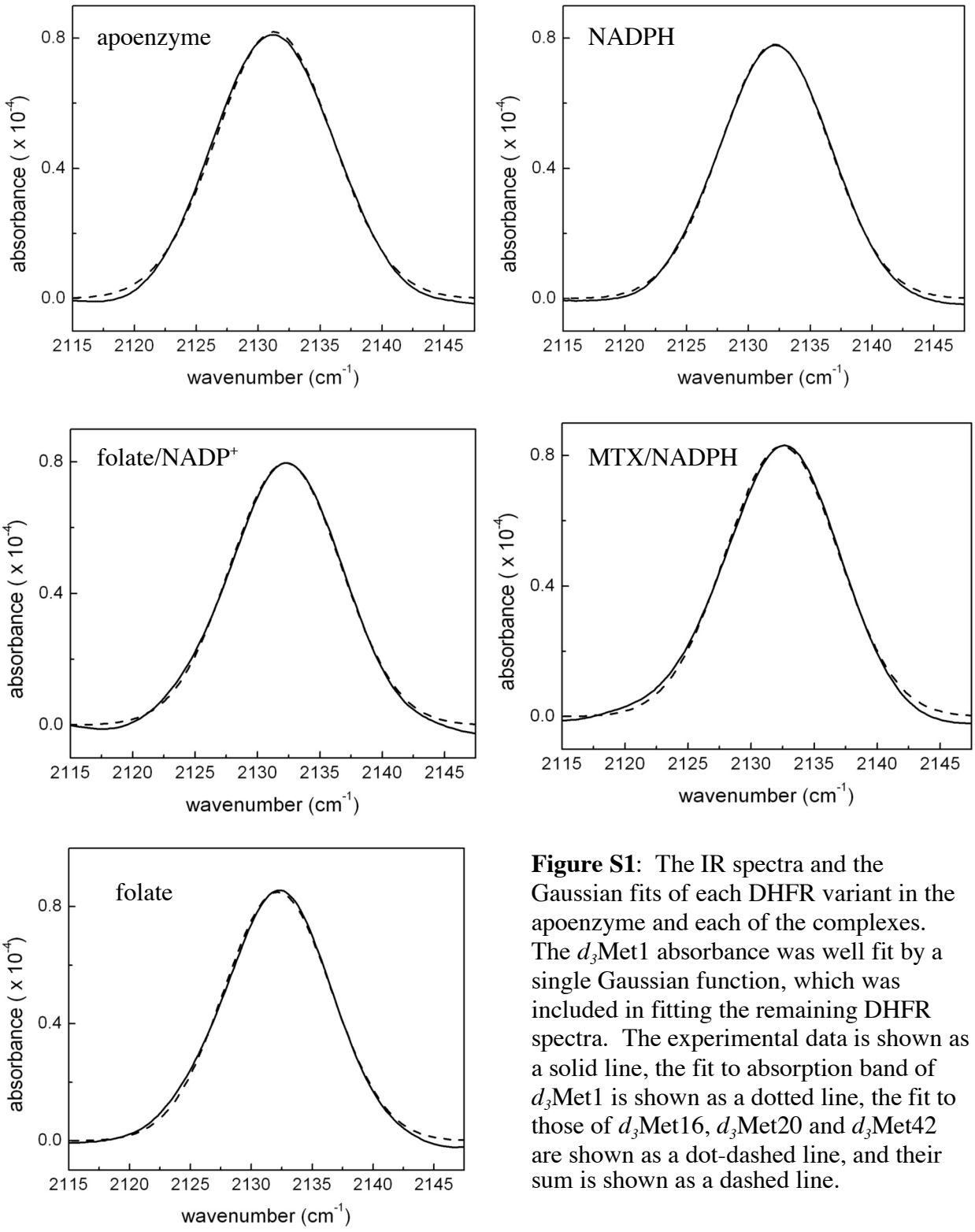
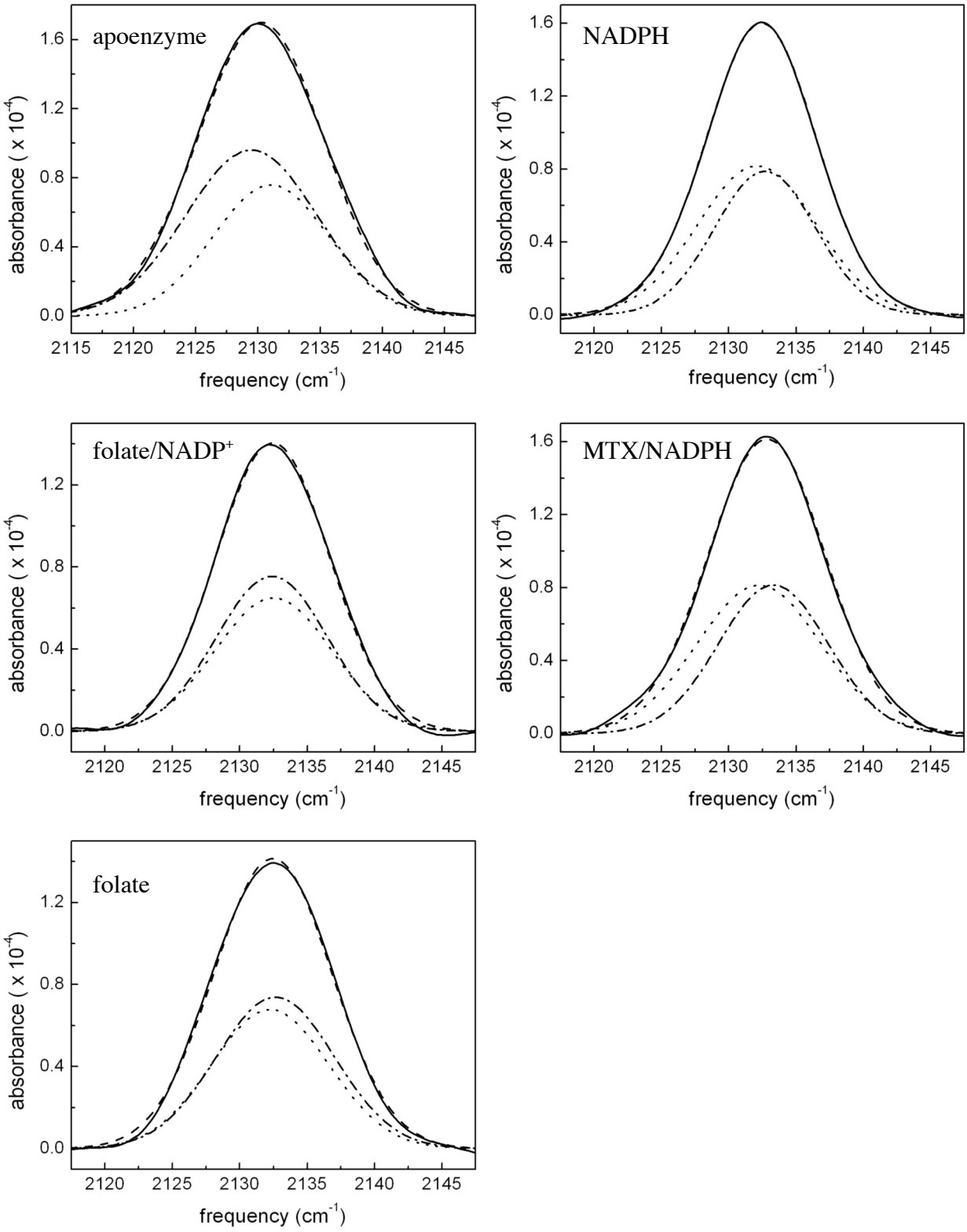
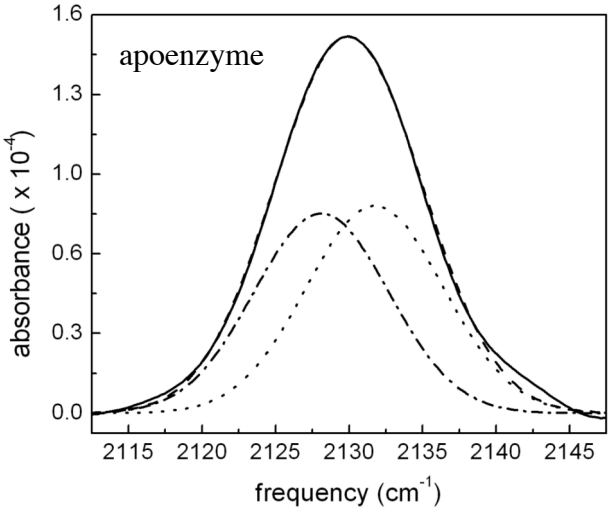


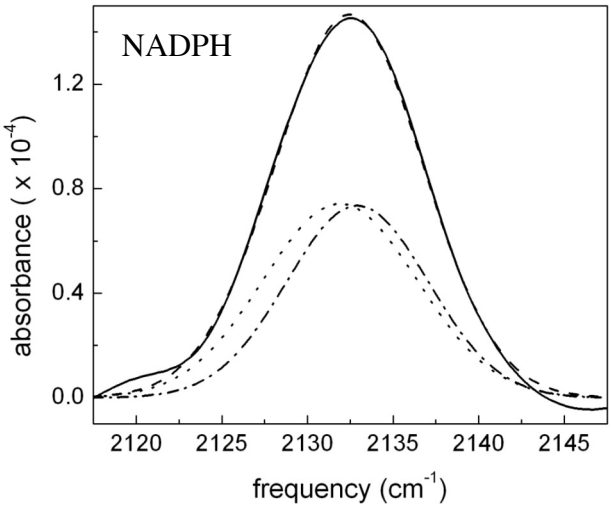

**Figure S1**: The IR spectra and the Gaussian fits of each DHFR variant in the apoenzyme and each of the complexes. The $d_3$Met1 absorbance was well fit by a single Gaussian function, which was included in fitting the remaining DHFR spectra. The experimental data is shown as a solid line, the fit to absorption band of $d_3$Met1 is shown as a dotted line, the fit to those of $d_3$Met16, $d_3$Met20 and $d_3$Met42 are shown as a dot-dashed line, and their sum is shown as a dashed line.

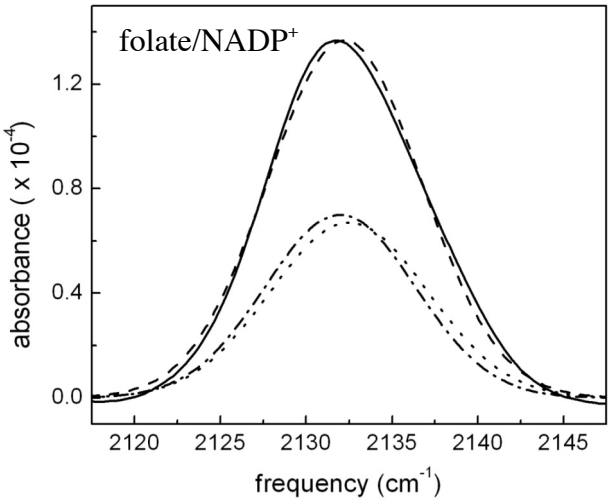
***Fits of $d_3$ Met1/$d_3$ Met16 labeled DHFR***

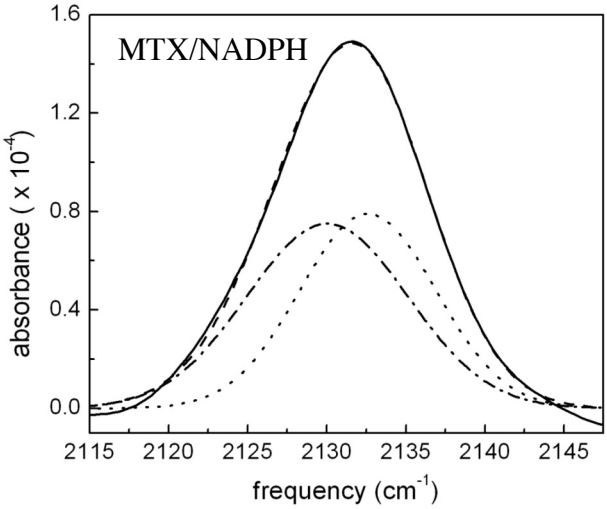


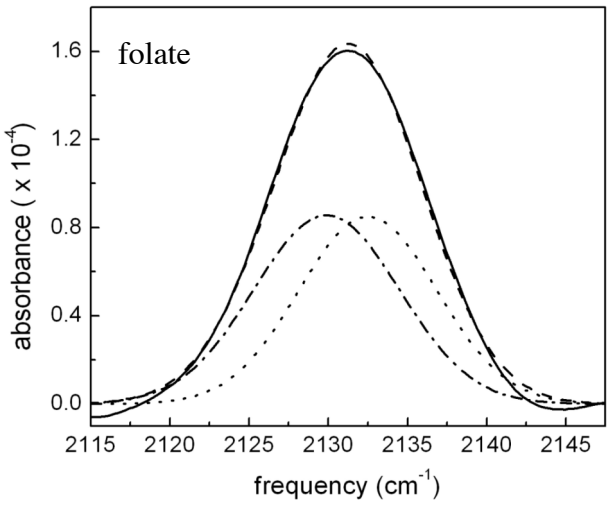
***Fits of $d_3$ Met1/$d_3$ Met20 labeled DHFR***



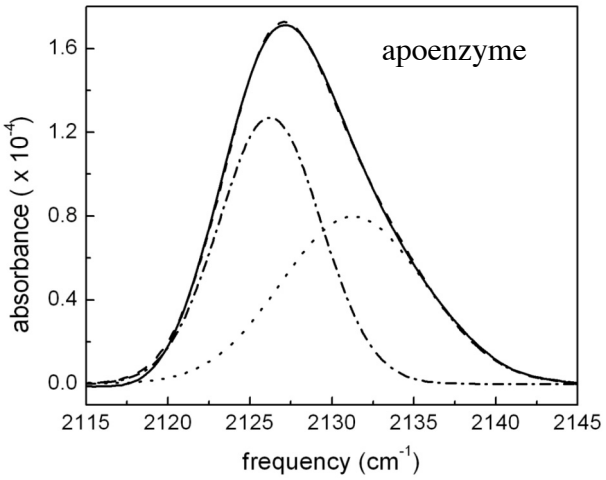



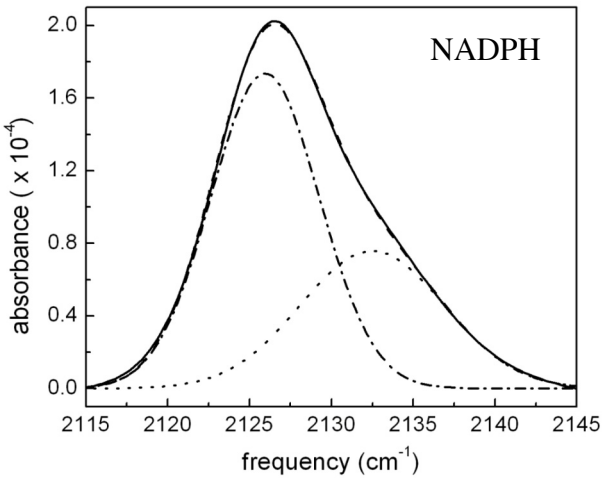



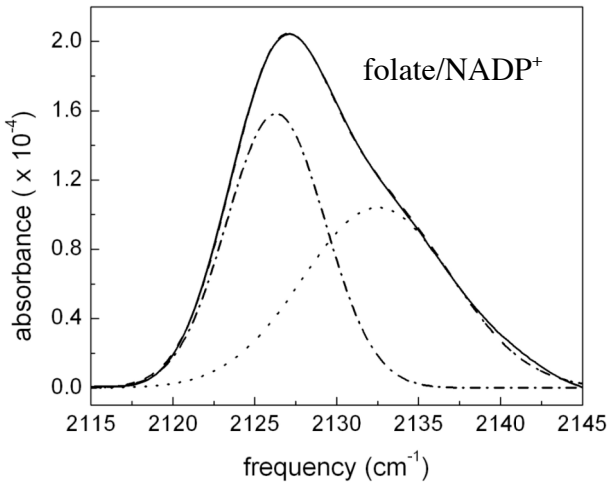



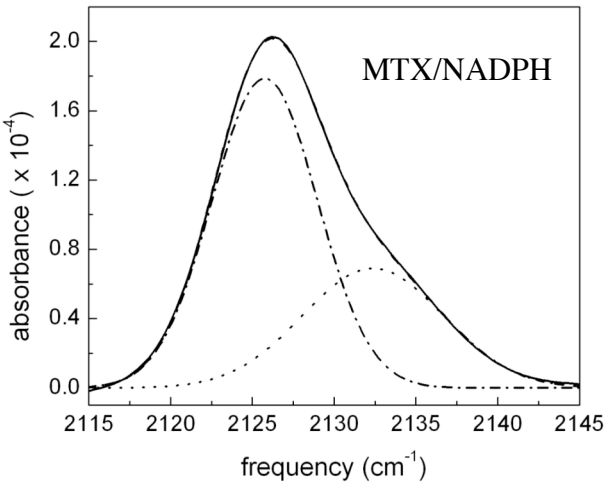



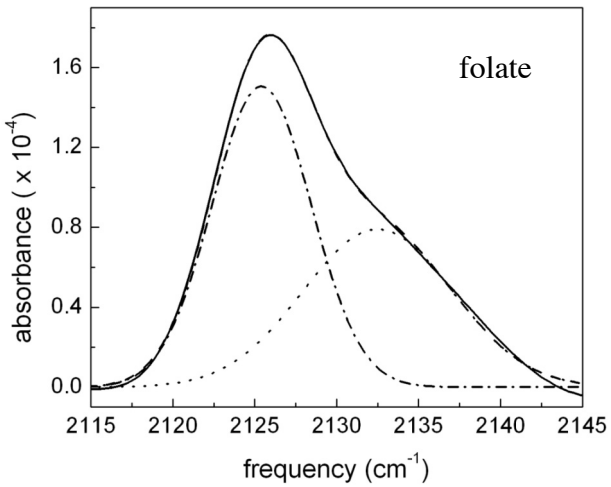
***Fits of $d_3$ Met1/$d_3$ Met42 labeled DHFR***



***Quantum chemistry calculations.***

Density functional (DFT) calculations on $CH_3SCD_3$ were carried out using the Gaussian 03 program, with various external electric fields. All calculations used the B3LYP functional, and either the cc-pVTZ or aug-cc-pVTZ basis sets. Structures were geometry optimized at each external field, and the frequencies for the $CD_3$ modes are given in the following Table.

**A. cc-pVTZ basis set**

| Field along the S-C direction (a.u.) | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| 0.010 | 2157.4 | 2282.3 | 2305.3 |
| 0.005 | 2164.2 | 2292.4 | 2310.7 |
| 0.002 | 2167.7 | 2297.8 | 2313.5 |
| 0.0 | 2169.6 | 2300.8 | 2315.5 |
| -0.002 | 2171.8 | 2304.2 | 2316.9 |
| -0.005 | 2174.4 | 2308.4 | 2319.0 |
| -0.010 | 2177.8 | 2314.6 | 2322.0 |
| Field perpendicular to the CSC plane | | | |
| 0.005 | 2169.3 | 2300.4 | 2315.1 |
| 0.0 | 2169.6 | 2300.8 | 2315.5 |
| -0.005 | 2169.3 | 2300.4 | 2315.1 |
| Field perpendicular to both S-C direction and CSC plane | | | |
| 0.005 | 2168.8 | 2300.2 | 2315.3 |
| 0.0 | 2169.6 | 2300.8 | 2315.5 |
| -0.005 | 2170.1 | 2301.7 | 2313.7 |

**B. aug-cc-pVTZ basis set**

| Field along the S-C direction (a.u.) | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|
| 0.010 | 2157.6 | 2282.3 | 2305.2 |
| 0.005 | 2164.7 | 2292.6 | 2310.7 |
| 0.002 | 2168.3 | 2298.0 | 2313.6 |
| 0.0 | 2170.4 | 2301.2 | 2315.3 |
| -0.002 | 2172.4 | 2304.3 | 2316.9 |
| -0.005 | 2174.9 | 2308.4 | 2319.0 |
| -0.010 | 2178.1 | 2314.2 | 2321.9 |

***References***


(1) Cremeens, et al. (2006) *J. Am. Chem. Soc.* 128, 6028-6029.
(2) Schnell, et al. (2004) *Biochem.* 43, 374-383.
(3) Lowther and Matthews (2000) *Biochem. Biphys. Acta* 1477, 157-167.
(4) Sawaya, M.R., Kraut, J. (1997) *Biochem.* 36(3), 586-603.
(5) Fierke, C.A., et al. (1987) *Biochem.* 26, 4085-4092.
(6) Taira and Benkovic (1988) *J. Med. Chem.* 31, 129-137.
(7) Shaw, et al. (1999) *Biochim. Biophys. Acta* 1429, 401-410.
(8) Gans, P. *Data Fitting in the Chemical Sciences.* John Wiley & Sons: New York, 1992; Chapter 6.
(9) Reimers, J.R. and Hall, L.E. (1999) *J. Am. Chem. Soc.* 121, 3730-3744.
(10) Gutmann, V. (1976) *Coord. Chem. Rev.* 18(2), 225-255.
(11) Rao, C.N.R., et al. (1976) *Chem. Soc. Rev.* 5(3), 297-316.
(12) Fawcett, W.R., et al. (1993) *J. Phys. Chem.* 97(37), 9293-9298.
(13) Akerlof, G. (1932*) J. Am. Chem. Soc.* 54(11) 4125-4139.