Supporting Information for

Decomposition of Vibrational Shifts of Nitriles into Electrostatic and Hydrogen Bonding Effects

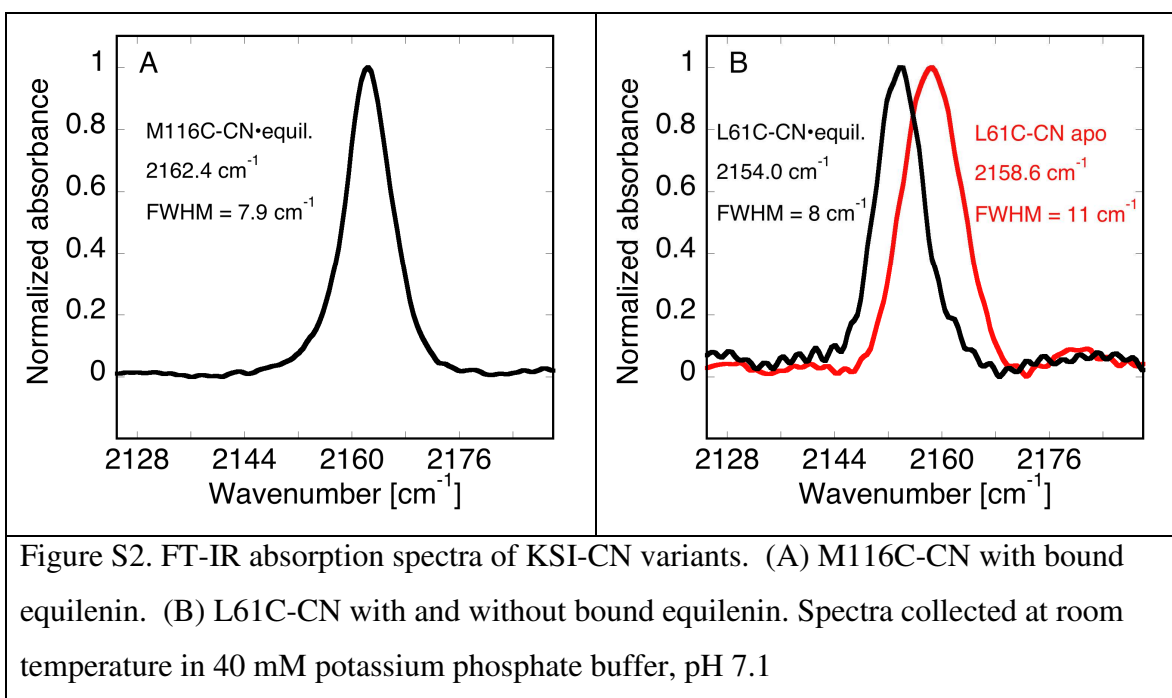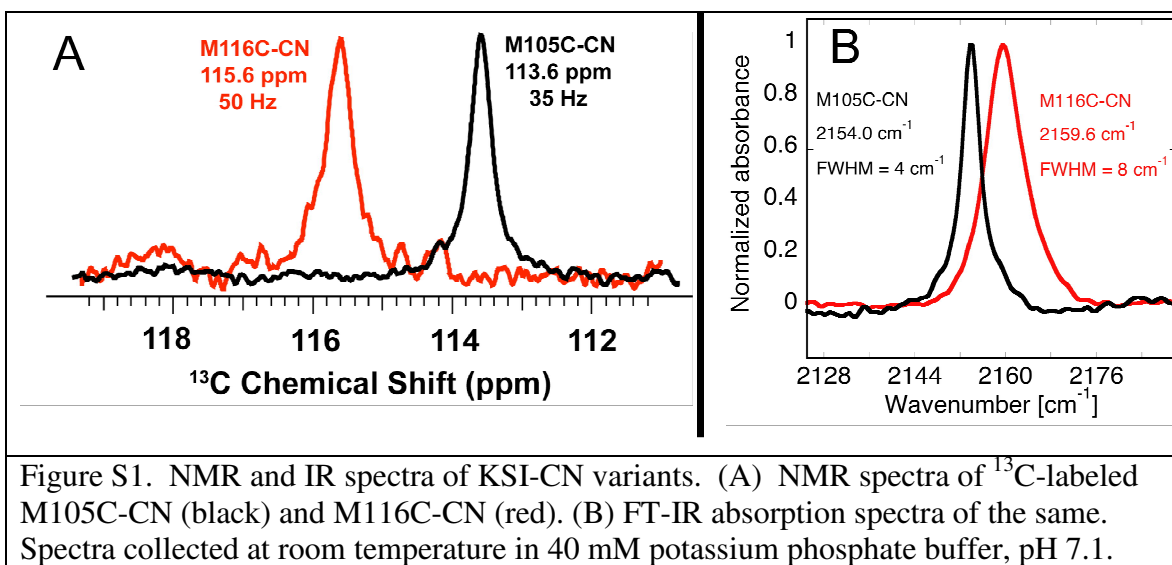Aaron T. Fafarman, Paul A. Sigala, Daniel Herschlag and Steven G. Boxer

Departments of Chemistry and Biochemistry, Stanford University, Stanford, California 94305-5080

Preparation and FT-IR spectroscopy of nitrile-modified proteins

Mutation, cyanylation, purification and spectral acquisition were performed as previously described[5]. Peak maxima were determined using a second-derivative-based method built into the OPUS FT-IR software (Bruker Photonics, Billerica, MA).

NMR Spectra of KSI-[13]CN

KSI-[13]CN samples were prepared as previously described[5], except that ^{13}C-labeled K[13]CN (Cambridge Isotope Labs) was used in place of KCN. One-dimensional ^{13}C NMR spectra were acquired on a 500 MHz Varian [UNITY]INOVA NMR spectrometer (125 MHz ^{13}C frequency) running VNMR v6.1C and equipped with a 5 mm PFG switchable probe operating at ambient temperature (20 ˚C). Samples contained 1 mM KSI, 2 mM ligand (bound samples only), 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% D_2O as the lock solvent. Spectra were acquired in a 5 mm Shigemi symmetrical microtube for ~5000 scans and processed using a 10 Hz line broadening. Peak positions are reported as the maximum value. Chemical shifts were referenced to an external standard of sodium 3-trimethylsilylpropionate-22,33-d_4 (0 ppm).

Figure S1. NMR and IR spectra of KSI-CN variants. (A) NMR spectra of ^{13}C-labeled M105C-CN (black) and M116C-CN (red). (B) FT-IR absorption spectra of the same. Spectra collected at room temperature in 40 mM potassium phosphate buffer, pH 7.1.

Figure S2. FT-IR absorption spectra of KSI-CN variants. (A) M116C-CN with bound equilenin. (B) L61C-CN with and without bound equilenin. Spectra collected at room temperature in 40 mM potassium phosphate buffer, pH 7.1

NMR and IR Spectra of Ethyl Thiocyanate in Solvents

Solutions of ethyl thiocyanate (EtSCN, Acros) were prepared in different deuterated solvents (Cambridge Isotopes); for dimethylformamide and trifluoroethanol

the non-deuterated solvent was used (Acros). Infrared absorption spectra were obtained on a Bruker Vertex 70 FT-IR spectrometer (Billerica, MA) equipped with an InSb detector. A band-pass (2000-2500 cm^{-1}) interference filter from Spectrogon (Parsippany, NJ) was used. A gas-tight demountable liquid cell (Bruker Photonics, Billerica, MA) with sapphire windows and offset spacers (one 75 μm, and one 100 μm spacer on either side) was used. Absorbance spectra were measured relative to a background taken with pure solvent. Baselines were calculated using a polynomial fit (0th to 4th order, depending on curvature) with roots defined at least 15 cm^{-1} distant from the peak maximum. Peak positions were determined using a second-derivative-based method built into the OPUS FT-IR software (Bruker Photonics, Billerica, MA).

In the apolar solvents cyclohexane-d_6 and toluene-d_8, we tested for nitrile self-association by monitoring the concentration dependence of the EtSCN nitrile stretch peak frequency. Our observation of a constant IR frequency at concentrations of ≤1% in toluene and ≤0.5% in cyclohexane suggest that EtSCN did not self-associate at these concentrations. Conditions for the NMR and IR samples were therefore: 0.5% EtSCN in cyclohexane, and 1% (vol/vol) in all other solvents. The IR absorption spectra exhibit roughly symmetric peaks within the experimental certainty, which is limited by the curvature of the baseline. An exception to this general behavior is seen for EtSCN in trifluoroethanol, where the spectral envelope seems to consist of a dominant peak with a shoulder at lower energy. No attempt was made to deconvolve the two contributions in this spectrum. The same samples were used for IR and NMR measurements, with IR detecting $^{12}C^{14}N$, and NMR detecting the ^{13}C at natural abundance. NMR measurements were obtained on a 300 MHz instrument otherwise identical to that described above. Peak positions are reported as the maximum value. The nitrile ^{13}C chemical shift was referenced to observed solvent peaks, whose 13C chemical shift values were referenced to 1% tetramethylsilane[41]; for D_2O, the nitrile chemical shift was referenced to an external standard of sodium 3-trimethylsilylpropionate-22,33-d_4 (0 ppm).

MD modeling of L61C-CN

A molecular dynamics-based structural model for L61C-CN was created by
placing the nitrile in the space occupied by the leucine side chain in the crystal structure
of KSI (PDB: 1OPY[36]). This structure was then relaxed by performing 4 ns of molecular
dynamics in a box of explicit SPC/E water (FFamber99 port[42,43], simulated in the
GROMACS environment[44,45], nitrile parameterized using Antechamber[46]; see Fafarman
and Boxer[26] for an example). The equilibrated structure resulted in a solvent-exposed
nitrile positioned at the mouth of the active site, presenting the possibility of a nitrile-
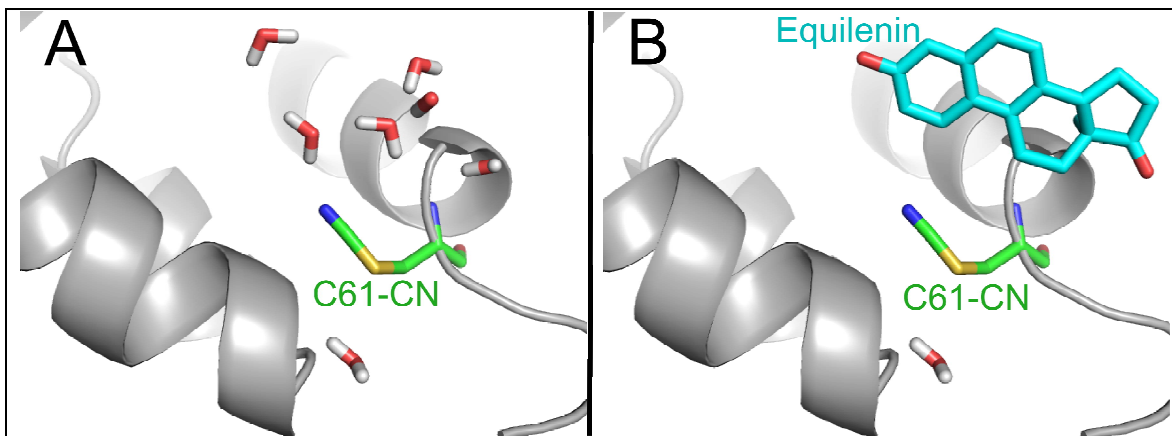water hydrogen bond as discussed in the text (Fig. S3).

Figure S3. Molecular dynamic model of L61C-CN with and without equilenin bound.
(A) A close-up view of a snapshot of a molecular dynamics model of KSI L61C-CN in a
box of SPC/E explicit water. Ribbon cartoon of the protein in grey and the
thiocyanatoalanine residue at position 61 and surrounding waters in stick representation
with carbon in green, sulfur in gold, nitrogen in blue and oxygen in red. (B) The same as
A, with equilenin modeled in to the active site based on alignment of the MD model with
the X-ray structure of KSI with equilenin bound (PDB: 1OH0)[36]; equilenin carbon atoms
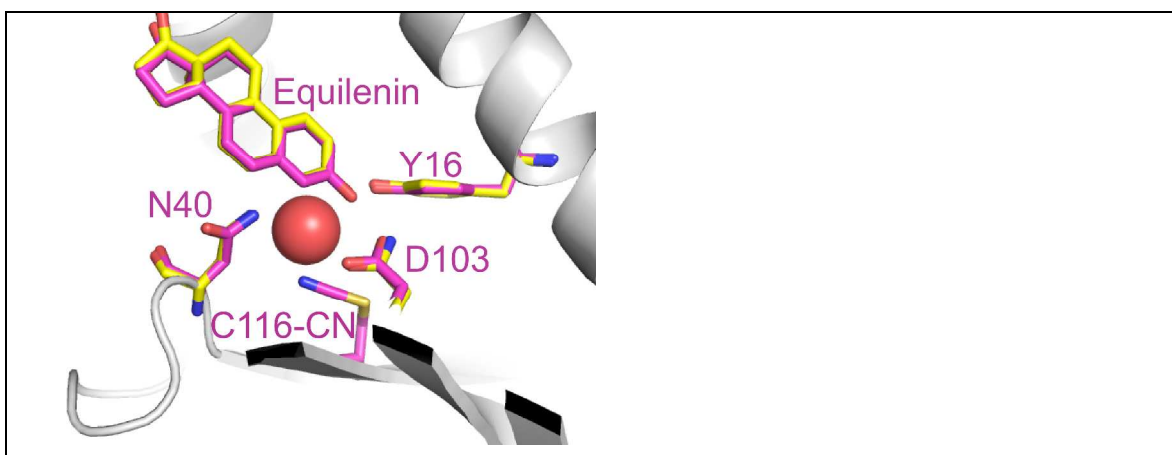in cyan.

Figure S4. Alignment of M116C-CN/D40N•Equilenin structure with D40N/D103N•Equilenin structure. M116C-CN/D40N•Equilenin structure (manuscript in preparation[35]) shown with backbone represented in grey ribbon, select side-chains and bound equilenin in magenta. One of the two independently refined monomers of the crystallographic dimer of D40N/D103N•Equilenin (3FZW[37]) was aligned with the nitrile-modifed structure, with select side-chains in yellow, water shown as a red sphere. The distance between the water from chain A (Water 162) to the nitrile nitrogen, is 2.4 Å in this alignment. In the second monomer of the asymmetric unit, a water molecule (Water 189) was refined with partial occupancy at two sites, one within 0.1 Å of Water 162, the other displaced by 1.7 Å; the second site is still within 3.3 Å of the nitrile nitrogen of the aligned M116C-CN structure.

Table S1: Vibrational frequencies and NMR chemical shifts for nitrile-modified KSI D40N at pH 7 in 40 mM potassium phosphate

Probe Position	Ligand	ν_{CN} (FWHM) [cm^{-1}]	$\delta(^{13}CN)$ [ppm]
L61C-CN	Apo	2158.6 (11)	117.10
	equilenin	2154.0 (8)	115.10
M116C-CN	Apo	2159.6 [a] (8)	115.62
	Equilenin	2162.4 [a] (8)	114.07
	19-nortest.	2160.0 [a] (8)	115.45
	2-naphthol	2162.5 [a] (8)	114.26
	4-F-3-Me-ph	2162.3 [a] (9)	114.38
M105C-CN	Apo	2154.0	113.60

[a] Reproduced from [5]

	ν_{CN} [cm^{-1}]	$\delta(^{13}CN)^{b}$ [ppm]
Table S2: Vibrational frequencies and NMR chemical shifts for Ethyl thiocyanate in various solvents[a]		
cyclohexane-d$_6$	2161.1	109.10
toluene-d$_8$	2157.1	111.40
CDCl$_3$	2158.2	112.30
tetrahydro-furan-d$_8$	2155.8	112.10
CD$_2$Cl$_2$	2157.2	112.60
acetone-d$_6$	2155.9	113.20
dimethylform-amide-d$_7$	2153.2	113.70
dimethyl-sulfoxide-d$_6$	2151.7	113.10
D$_2$O	2160.3	115.40
trifluoro-ethanol[c]	2168.5	115.70
formamide[c]	2156.1	114.40

[a] 1% v/v EtSCN in all cases but cyclohexane where 0.5% was required to avoid self association as indicated by concentration dependence of the IR peak frequency.
[b] referenced relative to 1% TMS, except D$_2$O, referenced to sodium 3-trimethylsilylpropionate-22,33-d_4 (TSP, 0 ppm)
[c] non-deuterated solvents

S.I. REFERNCES

(40) Lide, D. R. *CRC Handbook of Chemistry and Physics, 90th Edition*; 90th ed.; CRC Press, 2009

(41) Macomber, R. S. *A Complete Introduction to Modern NMR Spectroscopy*; 1st ed.; Wiley-Interscience, 1997.

(42)   Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. *J Comput Chem* **2004**, *25*, 1157-1174.

(43)   Sorin, E.; Pande, V. *Biophysical Journal* **2005**, *88*, 2472-2493.

(44)   Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Computer Physics Communications* **1995**, *91*, 43-56.

(45)   Lindahl, E.; Hess, B.; van der Spoel, D. *Journal of Molecular Modeling* **2001**, *7*, 306-317.

(46)   Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Computer Physics Communications* **1995**, *91*, 1-41