Supporting Information For:

Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S

*Aaron Fafarman and Steven G. Boxer**

Department of Chemistry

Stanford University

Stanford, California 94305-5080

* sboxer@stanford.edu

SUPPLEMENTAL INFORMATION TABLE OF CONTENTS

I. Nitrile-containing Small Molecules and Unstructured Peptides.

The spectroscopic responses to environmental changes (e.g. different solvents or temperatures) of the nitrile stretching transition of small molecule-nitriles or unstructured peptides were explored to better understand the observations of the nitrile-modified RNase complexes. In general, when solvated by a polar solvent, it is expected that the large ground state dipole moment of nitrile-containing molecules will exert an orienting force on the nearby solvent dipoles. These oriented dipoles will exert an electrostatic force on ground state dipole

moment, which is parallel[1] to $\Delta\dot{\mu}_{CN}$, leading to a lowering of the transition energy described to first order approximation by:

$$hc\Delta\overline{\nu}_{CN}^{obs} = -\Delta\dot{\mu}_{CN} \cdot \Delta\dot{F}_{solvent} \qquad (1)$$

where $\Delta\overline{\nu}_{CN}^{obs}$ is the frequency shift between two given solvents and $\Delta\dot{F}_{solvent}$ is the difference in the average electrostatic field exerted on the nitrile by the two different solvents. The size of the dipole and the degree of thermal disorder will affect the magnitude of the shift due to this mechanism. $\Delta\overline{\nu}_{CN}^{obs}$ is also subject to effects not described by this model, some of which can be solvent-specific such as those due to direct solvent-nitrile hydrogen bonding (see main text).

Small-Molecule Nitriles in Simple Solvents. The solvent-induced frequency shifts of benzonitrile (BenzCN) and methyl thiocyanate (MeSCN) were compared with those of acetonitrile in a range of different solvents to test the generality of the solvent-nitrile interaction reported by the vibration frequency. As seen in Fig S1, all three nitriles have very similar solvatochromic responses despite their large differences in size and shape.

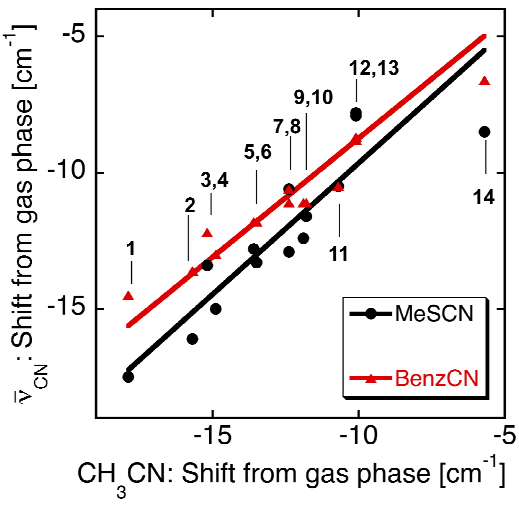

Figure S1. Gas-phase to solvent frequency shift for MeSCN and BenzCN relative to CH_3CN in identical solvents. Spectral shifts for the nitrile stretch of methylthiocyanate (MeSCN) and benzonitrile (BenzCN) relative to the gas phase, plotted against the same for acetonitrile (acetonitrile data from Reimers and Hall[2]). Solvents are: DMSO (1); DMF (2); THF (3); pyridine (4); 2Me-THF (5); acetone (6); formamide (7); Et_2O (8); NO_2CH_3 (9); CH_2Cl_2 (10); $CHCl_3$ (11); 3Me-pentane (12); hexanes (13); water (14). Least squares fitting yields slopes of 0.87 [$\Delta\overline{\nu}_{BenzCN}$]/[$\Delta\overline{\nu}_{MeCN}$] for BenzCN and 0.96 [$\Delta\overline{\nu}_{MeSCN}$]/[$\Delta\overline{\nu}_{MeCN}$] for MeSCN. Published gas-phase frequencies for MeSCN[3] and BenzCN[4] are 2171 cm^{-1} and 2242 cm^{-1} respectively. MeSCN and BenzCN were 100 mM in each solvent.

In either of the strong hydrogen bond donors, water or formamide[5], MeSCN, benzonitrile and acetonitrile exhibit similar shifts relative to the gas-phase, indicating similar solvent-solute interactions in each case. In both cases the magnitude of the solvent-shift is much smaller than that for similar polar solvents like dimethylsulfoxide and dimethylformamide, indicating the likely complication of a hydrogen bond, which contributes a blue shift to the frequency (see main text). This reinforces previous studies which have demonstrated that small-molecule alkyl thiocyanates[6] and aryl nitriles[7] engage in hydrogen bonds when exposed to water. This suggests an explanation for the blue-shift observed in Fig. 3 for all three nitrile-modified S-peptides when free in solution compared to when bound in the complex: the presence of a hydrogen bond to the nitrile of the peptide in water that does not exist in the protein. This suggestion is further tested based on the comparison of the thermochromism of the peptides compared to that of small molecules.

Thermochromism. The effects of temperature on the IR spectra of small molecule nitriles were also studied in a subset of the solvents described above with an eye to finding a spectroscopic signature unique to the nitrile-solvent hydrogen bonded case. As shown in Fig. S2, in non-hydrogen bond-donating solvents, the nitrile stretching frequency shifts to lower energy with decreasing temperature (positive thermochromism), while water and formamide (both examples of good hydrogen-bond donating solvents) display negative thermochromism. For nitriles in alcohol solvents, two overlapping peaks, which become better resolved at low temperature, have been previously observed and assigned[8] to the presence and absence of a hydrogen bond between the nitrile-nitrogen and hydroxyl-hydrogen. For methylthiocyanate in ethanol, the two peaks have opposite thermochromic trends (Fig. S2), with the hydrogen-bonded

peak frequency displaying a negative slope with temperature, and the free nitrile peak displaying a positive slope.

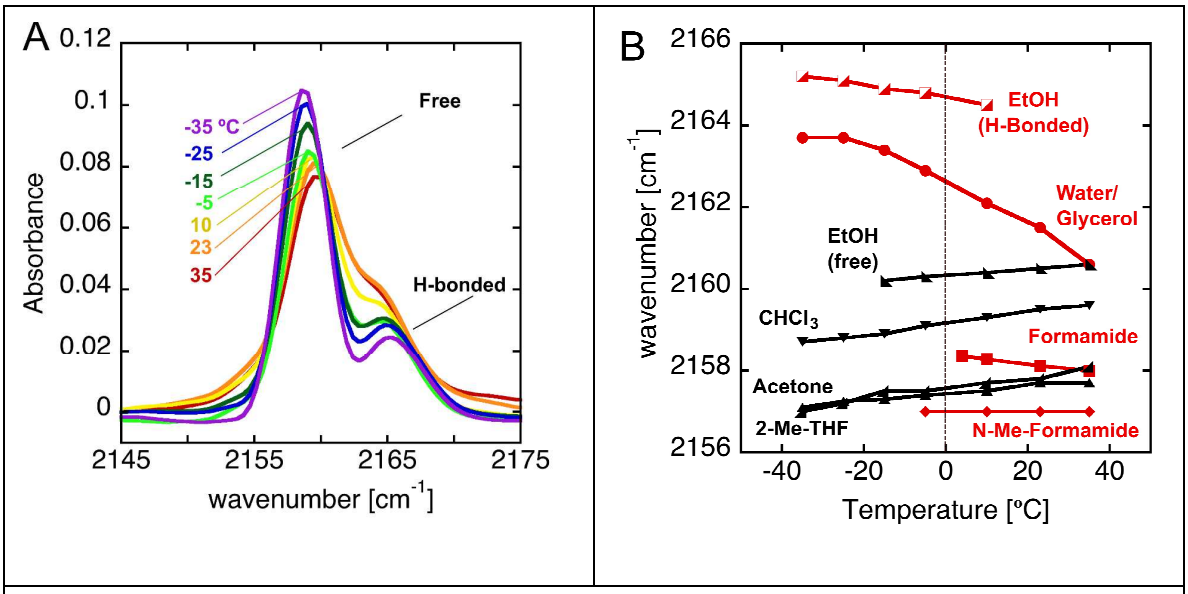

Figure S2. Thermochromic behavior of methylthiocyanate in different solvents. (A) Temperature dependent spectra of MeSCN in ethanol, with Fourier Self-Deconvolution[9] (3.7 cm^{-1} envelope function) applied to better resolve the high-energy shoulder; (B) MeSCN peak frequency *vs.* temperature for solvents containing oxygen- or nitrogen-hydrogen bonds, (red) or carbon-hydrogen bonds (black). Data in EtOH are separated into two populations as seen in (A). Water/glycerol is 1:1 (v/v).

Rather than developing a detailed model for the thermochromisim of nitriles, we note that the attractive potentials due to permanent and induced solvent dipoles should increase in magnitude with a decrease in temperature due both to a decrease in thermal disorder and an increase in material density[10,11]. Thus for simple, aprotic solvents a red shift is expected with decreasing temperature, i.e. a positive thermochromism; Fig S2B confirms that this trend is general for all *aprotic* solvents studied. This implies that the largest organizing force present in the solvation sphere is the torque that drives the alignment of the nitrile ground state dipole and the dipole moments of the solvent.

In contrast to aprotic solvents, in strongly hydrogen bond donating solvents the nitrile exhibits negative thermochromism. This contrary trend of hydrogen-bonding solvents can be understood by considering that hydrogen bond strengths increase with decreasing temperature as seen in gas-phase IR[12] and in solution phase NMR[13] measurements, and that for the reasons given in the main text, stronger hydrogen bonds equate to larger blue-shifts. The most dramatic example of the contrast of the two thermochromic trends is observed in the temperature dependent spectra of MeSCN in ethanol in Fig S2A. Here, with the aid of Fourier Self-Deconvolution techniques, the hydrogen bonded and non-hydrogen bonded peak centers are seen clearly moving away from each other with decreasing temperature (also see the trend-lines for the deconvolved peaks in Fig S2B). Note that chloroform is a very weak hydrogen bond donor[5] and for acetonitrile in chloroform the absolute frequency of the nitrile is not perturbed to the blue relative to aprotic solvents of similar polarity and polarizability as would be expected for nitrile-solvent hydrogen bond formation[2]; consistent with these pieces of evidence that argue against the presence of a nitrile-chloroform hydrogen bond, MeSCN exhibits positive thermochromism.

For the nitrile-modified S-peptides free in glycerol/water buffer, the thermochromic behavior provides further evidence that nitrile-modified S-peptides are participating in hydrogen bonds to water when free as a random coil in aqueous buffer. As seen in Fig. S3 for the case of pCN-S-peptide, they display the negative thermochromism typical of small-molecules solvated in hydrogen bond-donating solvents (Fig S2).

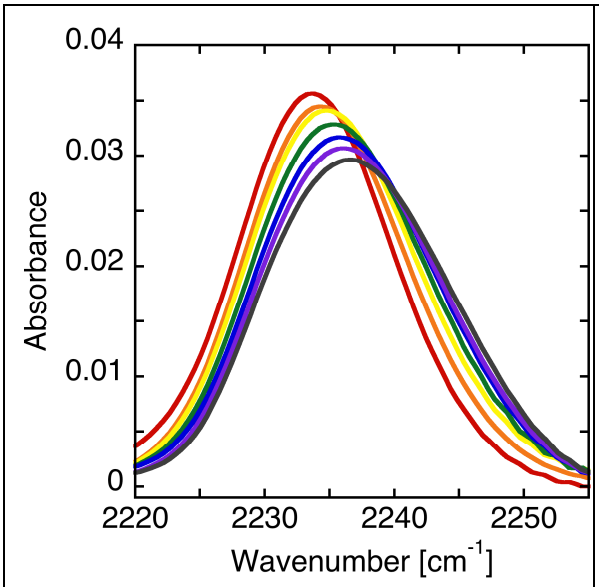

Figure S3. IR absorption spectra for pCN-S peptide free in solution at different temperatures. Buffer is 20 mM HEPES, 50% vol/vol glycerol. Temperatures from red to grey are 35, 23, 10, -5, -15, -25, -35 ºC.

The results from the thermochromism of simple solvents cannot be generalized to that of nitrile-modified RNase, because in a protein the polar groups are much more constrained and lack the same freedom to re-orient around the nitrile dipole. The minimal perturbation to the RNase structure evidenced by X-ray crystallography (performed at 77 K) supports the idea that protein polar groups do little to minimize the electrostatic potential of their interaction with the nitrile group. Furthermore, the relationship between temperature and material density is not as simple in a protein as it is in a solvent and may vary from site-to-site based on limited structures that have been studied as a function of temperature[14]. Thus the sign of the thermochromic effect for nitriles in proteins is not necessarily expected to be positive, and in the case of SCN-RNase and pCN-RNase it is negative, while for mCN-RNase, two opposing trends are present.

As an aside, we note that the free nitrile is more thermodynamically favored than the hydrogen-bonded form in ethanol, evidenced by the fact that the hydrogen bonded peak is a small fraction of the total absorption spectrum (Fig. S2A), despite the fact that hydrogen bonded

nitriles generally have much higher oscillator strengths[8]. At cryogenic temperatures, this peak is undetectable, presumably because the energy difference relative to the non-hydrogen-bonded state is much greater than k_BT at 77 Kelvin (data not shown).

II. mCN-RNase Peak Fitting.

The room temperature absorption spectrum in the nitrile stretching region of mCN-RNase is subtly asymmetric, and as the temperature is lowered the asymmetry becomes increasingly pronounced until a distinct shoulder is apparent (Fig 4 of the main text and Fig S4 below). In the main text it is speculated that these two peaks arise from the pair of possible orientations, α and β, with respect to the surrounding protein structure, of the nitrile substituent of the phenyl ring (see Fig. 2 of the main text). The β conformer places the nitrile near the partial positive charge of an aryl hydrogen of phenylalanine 120 (main text, Fig. 2B), and the α conformer places it very near the partial negative charge of the sulfur of methionine 13 (main text, Fig. 2A).

MD simulations started in either the α and β conformer do not interchange and the former is higher in energy than the latter at both simulated pH values (see Table 4 of the main text). Although the average simulated frequencies are separated by several wavenumbers in energy in the initial few 100 ps of sampling (comparable to the observed splitting at low temperature), they relax to a much smaller average splitting as the structures drift from the initial X-ray coordinates, particularly for the pH 8 simulation where the difference is less than a wavenumber by the end of the simulation. However, since the two populations are not experimentally resolved at room temperature, a standard absorption experiment alone cannot assess whether the MD is accurately representing the experiment at room temperature. It is

possible that the experimental energy difference actually is a function of temperature due to temperature-dependent structural changes, and that MD sampling captures these changes, more accurately reflecting the room temperature structure, and therefore calculating only a small splitting. It is equally plausible that the small splitting at pH 8 is an artifact due to the MD force field leading to a subtly inaccurate structural ensemble upon equilibration or an accurate structural ensemble and misrepresented electric fields.

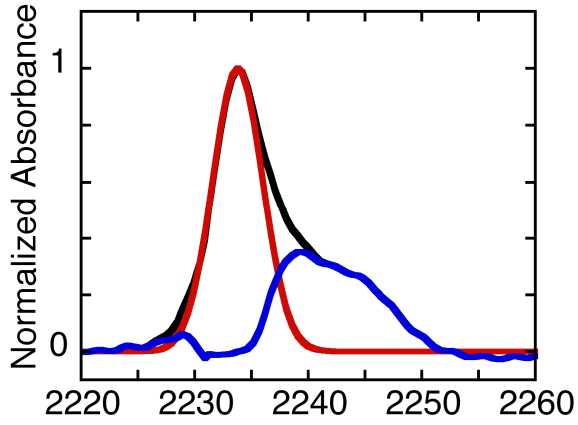

Figure S4. IR absorption spectrum of mCN-RNase at 80 Kelvin with calculated fit. 10 mM mCN-RNase in 50% glycerol water (self-buffered at pH 7) shown in black with a Gaussian peak fitted to the main band (5.2 cm^{-1} FWHM, 2233.8 cm^{-1} peak max) shown in red and the residuals to the fit shown in blue, indicating at least one additional population 8-9 cm^{-1} to the blue.

III. Choice of Dielectric.

The calculation of the electric field at the nitrile employs Coulombs law, and a choice must therefore be made regarding the permittivity constant to use. The atomic charges in the

Amber '99 force-field[15] used in this study (unchanged from their values in the Amber '94 set[16]) were derived from *ab initio* quantum chemical methods and were intended for use with the vacuum permittivity (i.e. the dielectric, ε=1); this choice was validated by comparing experimentally determined conformational equilibria, solvation free energies, liquid enthalpies and densities of a set of small molecules to simulations of the same set in explicit solvent[16]. It was not reported how much the agreement changed with different choices for the dielectric, or how different choices of parameters, for example, the van der Waals or torsional parameters, affected the choice for dielectric. The simulations undertaken here were performed using $\varepsilon = 1$ in the actual MD force calculations, in recognition of the interdependence of these parameters; however, the approach we have employed to output electrostatic fields from the MD trajectory (see Materials and Methods in the main text) allows the dielectric used to evaluate pair-wise forces on the probe to be varied in the post-processing of the trajectory, providing means to independently test the influence of this parameter.

The value of ε during post-processing was treated as a free parameter, ranging from 1 to 3, and the agreement between the most probable frequency in the instantaneous frequency distributions, and the experimentally determined peak frequency for the nitrile-modified proteins was determined (Fig. 5 and Table 4 of the main text). With this set of six observations and predictions (three probe sites, high and low pH), we find the best agreement to occur for $\varepsilon = 2$. This value can be compared to the dielectric directly measured for dry lysozyme at room temperature where nuclear motions of the protein are retarded but electronic polarizibility is operative[17]. Other theoretical approaches to protein electrostatics have also found that when nuclear motions are accounted for but induced dipoles are not, $\varepsilon = 2$ gives the best agreement with experiment[18,19].

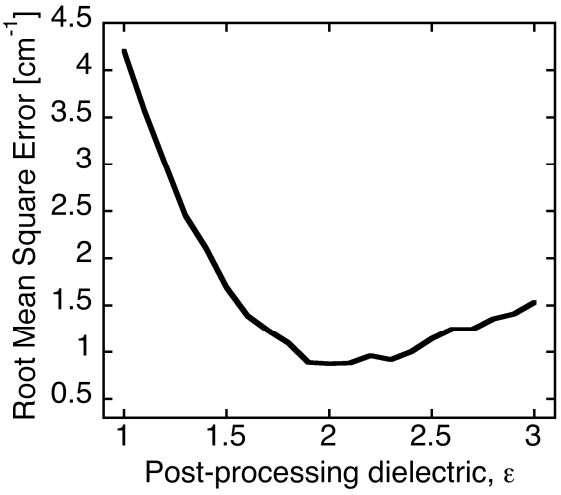

Figure S5. RMS difference between measured shifts in the peak frequency for SCN-, mCN- and pCN-RNase at pH 4.5 and pH 7.0 vs. the peak of the distribution of frequencies from a 50 ns MD trajectory (post-REMD) calculated at different values of the post-processing dielectric (see text) ; experimental and simulated shifts are calculated relative to the value for EtSCN, m-tolunitrile or p-tolunitrile in hexane respectively (see main text).

IV. Modifications to GROMACS code.

Electrostatic forces are independently calculated on all particles at every time step in an MD simulation, therefore it is a negligible increase in computational complexity to output those forces. To calculate the force at the midpoint of the CN bond, we utilize the 'virtual site' facility (included in GROMACS 3.3 and above) to introduce a mass-less virtual particle with a charge of -0.0001 electrostatic units, whose position is updated at every time-step to lie exactly halfway between the C and N atoms of the nitrile. An equal amount of positive charge is added to the total charge on the carbon atom to maintain neutrality. GROMACS will not output the forces on

virtual site particles in normal operation, so we modified the source code for the MD engine itself (md.c) to extract these forces at every time step. Dividing the resulting output by the charge on the virtual site gives the electrostatic field due to the sum of all pair-wise Coulombic forces. It does *not* account for the attractive portion of the van der Waals force which is also electrostatic in origin (or the repulsive portion), but this contribution also exists in the reference state (MeSCN, p-, or m-tolunitrile in n-hexane), and is assumed to be similar in organic solvent as in the interior of the protein, and therefore to cancel out in the calculation of differences relative to this reference.

As a practical guide for researchers who wish to employ this tool, we provide the following short description, which assumes a familiarity with GROMACS. First, select Particle Mesh Ewald electrostatics by adding the following to the .mdp file for the MD run:

```
coulombtype     = pme
```

which calculates long range electrostatics in Fourier-space and short range forces in direct space. Also in the .mdp file, the particle numbers for the C and N of the nitrile and for the virtual site (X, Y, and Z, respectively) must be added to the bottom using the syntax:

```
userint1        = X
userint2        = Y
userint3        = Z
```

The virtual site should be added to the list of particles in the .gro file and the .top files after running pdb2gmx but before solvating the protein. Add to the topology file the [virtual_sites2] directive (see GROMACS manual for details), and use the [exclusions] directive to exclude the calculation of forces between the virtual particle and the atoms of

the nitrile side-chain. The following small portion of code must be inserted below the **do_force** routine in the last third of the md.c source code (line 650 in 3.3.1):

```
/* the direct space force (f) and the reciprocal space electrical force (f_el_recip) are added together to calculate the total force
on the virtual particle (given by userint3) */
rvec_add( f[inputrec->userint3-1], fr->f_el_recip[inputrec->userint3-1], totalfieldvec );

/* the unit vector from the C (userint1) to the N (userint2) of the nitrile is calculated */
rvec_sub( state->x[inputrec->userint1-1], state->x[inputrec->userint2-1], diff );
unitv( diff, cnunitvec );

/* the projection (dot-product) of the total force on the CN unit vector is calculated and output to the log file */
efieldproj = iprod( totalfieldvec, cnunitvec );
fprintf( log, "Force projection in kJ/(mol*nm) %f\n" , efieldproj );
```

The output to the log file is equal to the projection of the force, on the virtual particle, along the CN bond, in units of kJ/(mol•nm). To get to familiar units of MV/cm, given a virtual particle with -0.0001 esu of charge, multiply the output by 1038.

Supporting References

(1) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2000**, *104*, 11853-11863.

(2) Reimers, J. R.; Hall, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 3730-3744.

(3) Sullivan, J. F.; Heusel, H. L.; Durig, J. R. *J Mol. Struct.* **1984**, *115*, 391-396.

(4) Green, J. H. S.; Harrison, D. J. *Spectrochimica Acta A* **1976**, *32*, 1279-1286.

(5) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc. Perkin Trans. 2* **1989**, 699-711.

(6) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. *In Preparation*.

(7) Aschaffenburg, D. J.; Moog, R. S. *J. Phys. Chem. B* **2009**, *113*, 12736-12743.

(8) Eaton, G.; Pena-Nunez, E. S.; Symons, M. C. R. *J. Chem. Soc. Faraday Trans. 1* **1988**, *84*, 2181-2193.

(9) Yang, W. J.; Griffiths, P. R.; Byler, D. M.; Susi, H. *Appl. Spectrosc.* **1985**, 282-287.

(10) Schweizer, K. S.; Chandler, D. *J Chem. Phys.* **1982**, *76*, 2296-2314.

(11) Bublitz, G. U.; Boxer, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 3988-3992.

(12) Wójcik, M. J. *Chem. Phys. Lett.* **1977**, *46*, 597-599.

(13) Muller, N.; Reiter, R. C. *J. Chem. Phys.* **1965**, *42*, 3265-3269.

(14) Tilton, R. F.; Dewan, J. C.; Petsko, G. A. *Biochemistry* **1992**, *31*, 2469-2481.

(15) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J Comput. Chem.* **2000**, *21*, 1049-1074.

(16) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1996**, *118*, 2309.

(17) Harvey, S. C.; Hoekstra, P. *J. Phys. Chem.* **1972**, *76*, 2987-2994.

(18) Suydam, I.; Snow, C.; Pande, V.; Boxer, S. *Science* **2006**, *313*, 200-204.

(19) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400-417.