## Supplementary Methods

*Protein Engineering*

All experimental studies were performed on the highly stable Δ+PHS variant of SNase [1,2] and single-site variants of Δ+PHS. The variant plasmids were prepared with QuikChange site-directed mutagenesis on a pET24a+ vector as described previously[1,2]. Protein purification was performed as described previously[3].

*Stability Measurements*

Stability measurements were performed with guanidine hydrochloride titrations using an Aviv Automated Titration Fluorimeter 105 as described previously [4].

*Crystallization and X-ray diffraction data collection*

Δ+PHS variants were crystallized using the hanging drop vapor diffusion method. The protein was pre-incubated with calcium chloride and pdTp in a 1:3:2 molar equivalent ratio prior to mixing with the reservoir solution. In each case, the reservoir solution contained 25 mM potassium phosphate and 20-29% w/v 2-methyl-2,4-pentanediol (MPD) (Sigma-Aldrich) in the pH 6-9 range. The proteins were mixed in a 1:1 ratio with reservoir solutions prior to suspension over the reservoir solution and incubation at 277K. Single crystals were harvested using a nylon loop mounted on a copper base (Hampton Research) and flash-cooled in liquid nitrogen. Diffraction data were collected at 100K from a single crystal of each variant using the X25 beamline at Brookhaven National Laboratory (BNL) at wavelength 1.100Å for variants L36A, L38A, V66A, V74A, I92A and L103A, 1.000Å for V23A, L25A, T62A and L125A, and 1.008Å for F34A. The data were indexed, integrated, scaled and merged using the HKL2000 software package[5] to yield a set of unique reflections.

*Crystallographic structure determination and analysis*

Initial phases for each structure were obtained by maximum likelihood-based molecular replacement method with Phaser software[6] within the CCP4i [7] suite version 6.0.7, using 3BDC.pdb as the search model. Prior to molecular replacement, 3BDC.pdb coordinates were modified by truncating the substituted amino acid for the appropriate variant to Ala, removing all water molecules, and setting all B-factors to 20.00 $Å^2$. Residues H8, T13, Q30, K64, V74, T82, Y113, V114 and Y115 were also truncated to Ala in the search model. Iterative model building and refinement were performed using COOT [8] and Refmac5 [9] to yield the final models. Water molecules were added in the protein model during model building in regions where spherical electron density in in $2F_o$-$F_c$ and $F_o$-$F_c$ maps within 3.5 Å of a hydrogen bonding partner were observed. Structure factors, geometry and Ramachandran statistics were evaluated using the SFCHECK [10] and PROCHECK [11] programs and the MolProbity server [12]. For structures V23A, L25A, L36A, V66A, V74A and L125A the TLSMD server [13] was used for the final combined TLS and restrained refinement cycle. Each structure reports Ramachandran statistics with 113 of 114 (99.1%) of the non-glycine, non-proline and non-end residues in the allowed regions. Calculations for RMS deviation of protein atoms of the variant structure and 3BDC.pdb were performed using the VMD program [14]. For each RMSD calculation, the 3BDC.pdb coordinates were modified to reflect the alanine truncation in the variant structure. In addition, for side chains adopting multiple conformations, only the dominant conformer was included.

HP Fluorescence experiments

High-pressure fluorescence experiments were carried out as previously described using an ISS steady-state fluorometer (Champaign, IL) [15], with 50μM protein samples dissolved in 50mM Tris buffer at pH 7. The temperature was maintained stable at 293K. For each experiment, a tryptophan emission spectrum was

collected at equilibrium, from 320 to 450 nm, using an excitation wavelength of 290 nm. At equilibrium tryptophan emission spectrum was collected from 320 to 450 nm using an excitation wavelength of 290 nm. At each pressure, the intensity-weighted average wavelength $\langle\lambda\rangle$ was calculated using the ISS software:

$$\langle\lambda\rangle_j = \frac{\sum_j F_j \lambda_j}{\sum_j F_j}$$

Where j=320, 321…450 nm.

Data were fitted to a two-state unfolding equilibrium as function of pressure for values of $\Delta G_u^0$ and $\Delta V_u^0$ using the BioEQS software [16], assuming a linear evolution of the free energy of unfolding with the pressure p:

$$\Delta G_u(p) = \Delta G_u^0 + p\Delta V_u^0$$

where

$$\Delta G_u(p) = -RT\ln K_u(p)$$

and

$$K_u(p) = \frac{\langle\lambda\rangle_f - \langle\lambda\rangle_p}{\langle\lambda\rangle_p - \langle\lambda\rangle_u}$$

The ensemble of curves at different GuHCl concentrations was fitted globally and the uncertainty in the recovered $\Delta V_u^0$ parameter was evaluated by rigorous confidence limit testing in which the tested parameter is reanalyzed at multiple values while the other parameters are allowed to float. Values of average emission wavelength were quantum yield weighted to correct for the decrease in intensity upon unfolding.

HP NMR experiments

Uniformly $^{15}$N-labeled protein samples were dissolved at 1mM concentration in 200 µL (3 mm tube) or 300 µL (high pressure ceramic tube) 10mM Tris buffer at pH7. 10% of $D_2O$ was added for the lock procedure. In all experiments, the $^1$H carrier was centered on the water resonance and a WATERGATE sequence [17,18] was incorporated to suppress the solvent resonance. All NMR spectra were processed and analyzed with GIFA [19].

Amide resonances were assigned at atmospheric pressure from 3D [$^1$H,$^{15}$N] NOESY-HSQC (mixing time 100 ms) and 3D [$^1$H,$^{15}$N] TOCSY-HSQC (isotopic mixing 60 ms) double-resonance experiments [20,21] recorded on a Bruker AVANCE 700 MHz spectrometer equipped with a 5 mm Z-gradient $^1$H-$^{13}$C-$^{15}$N cryogenic probe, using the standard sequential procedure. $^1$H chemical shifts were directly referenced to the methyl resonance of DSS, while 15N chemical shifts were referenced indirectly to the absolute frequency ratios $^{15}$N/$^1$H = 0.101329118.

High-pressure heteronuclear 2D $^{15}$N-$^1$H HQSC spectra [22] were recorded at 293K on a 600MHz Bruker Avance III spectrometer equipped with a 5 mm Z-gradient $^1$H-X double-resonance broadband inverse (BBI) probe. Commercial ceramic high-pressure NMR cell and an automatic pump system (Daedalus Innovations, Philadelphia, PA) were used to vary the pressure in the 0-2.5 kbar range. Under equilibrium conditions, native cross-peak intensities were integrated from the corresponding HSQC spectrum and the resulting

intensity *versus* pressure data points were individually fitted for each resonance. The fitting procedure was equivalent to the one used for the high-pressure fluorescence experiments described above, except no correction for quantum yield was applied.

All-atom Molecular Dynamics simulations

As previously[15], molecular dynamics (MD) simulations were performed using the GROMACS 4.4 package [23] and OPLS-AA [24] protein force field. A full-length structure of Δ+PHS was constructed from the crystallographic reference structure 3BDC.pdb and MODELLER software [25]. The protein was inserted in a cubic box with a minimal distance of 1 nm from any protein atom to the box boundary. The box was hydrated with 15719 SPC/E [26] water molecules and 5 $Cl^-$ were added to neutralize the system. After steepest descent energy minimization and 10 ps equilibration, the resulting conformation was used as starting point for 10 ns long MD simulation with 2 fs time step. Conformations were saved every 10 ps. H-bonds for protein and water were constrained respectively with LINCS [27] and SETTLE [28]. Long-range electrostatic interactions were calculated using particle-mesh Ewald[29] with a grid spacing of 0.12 nm and cubic interpolation. van der Waals interactions were cut off at 1 nm. The non-bonded list was updated every 10 integration steps. The temperature was controlled using a Nosé-Hoover thermostat [30,31] and the pressure was controlled using a Parrinello-Rahman barostat [32]. Both were used with a 5 ps coupling time. Simulations were performed at 293K and 1 bar with system compressibility set to $4.6.10^{-5}$ $bar^{-1}$.

.

Estimation of the cavities location in the folded states ensemble structures was done as previously. Briefly the 1000 generated configurations were analyzed using the McVol algorithm[33] and the void density, defined as the mean number of MC cavity points found within a 0.5 nm radius sphere around each Cα atom, was computed. Correspondingly, the water density was defined as the mean number of water oxygen atoms found within a 0.5 nm radius sphere around each Cα and normalized to the largest number of oxygen atoms found. A 0.5 nm probe radius was found after testing several values to appropriately minimize overlap and maximize coverage.

Data based Go-model simulations

Full-length structures of Δ+PHS and L125A variant were constructed from crystallographic reference structures 3BDC and 3OSO respectively and MODELLER software. A Cα model of these proteins and the corresponding Go-model parameters were generated using the SMOG@ctbp web server [34]. A complete thermodynamic integration of the Δ+PHS protein using a WHAM algorithm[35,36] was performed to determine the folding temperature of this model. A temperature of $0.85*T_f$ was then used for both Δ+PHS and L125A simulations. The pressure dependence was introduced through the experimentally derived fractional contact maps with a simple three steps procedure:

1. Based on the HSQC spectra recorded at each pressure, the probability to find a residue *i* in a folded states at a pressure p: $p(i)_p$ is given by the corresponding normalized resonance intensity.

2. The probability to form a specific native contact between residues i and j is then simply calculated as: $p(i,j)_p = p(i)_p.p(j)_p$

3. A list of native contacts is established through a random number generator by individually testing each native contact (i.e. a native contact is accepted in the list if : rand() < $p(i,j)_p$). For example, if the random number between 0 and 1 is 0.4 and the contact probability is 0.6, then this contact is counted in the list. If however, the random number generated for that contact in that list is 0.8 then the contact is not counted. This procedure was repeated 100 times, generating 100 different lists of native contacts for both models. A 100 ns long Cα Go-model simulation was finally performed independently from each list of native contacts and the resulting conformations (400,000) were collectively analyzed based on the fraction of native contact (Q). Free energy profiles at several pressures were therefore reconstructed from these simulations.

| | Δ+PHS V23A | Δ+PHS L25A | Δ+PHS F34A |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ |
| Cell dimensions | | | |
| *a, b, c (Å)* | 31.09, 60,19, 37.99 | 31.14, 60.25, 38.01 | 30.58, 61.16, 38.62 |
| *α, β, γ (°)* | 90.00, 94.69, 90.00 | 90.00, 94.94, 90.00 | 90.00, 92.52, 90.00 |
| Resolution (Å) * | 50.00-1.60 (1.63-1.60) | 50.00-1.60 (1.63-1.60) | 50.00-1.72 (1.75-1.72) |
| $R_{merge}$ | 0.060 (0.231) | 0.044 (0.260) | 0.058 (0.280) |
| Average I/σ(I) | 15.8 (7.4) | 19.5 (8.2) | 16.8 (7.1) |
| Completeness (%) | 0.969 (0.934) | 0.993 (0.987) | 0.997 (0.999) |
| Redundancy | 6.2 (5.6) | 7.3 (7.2) | 7.3 (6.9) |
| **Refinement** | | | |
| Resolution (Å) | 32.00-1.60 (1.64-1.60) | 30.13-1.60 (1.64-1.60) | 38.58-1.72 (1.77-1.72) |
| No. of reflections | 17911 (1140) | 18427 (1368) | 15120 (1109) |
| $R_{work}$ / $R_{free}$ | 0.174 / 0.213 | 0.167 / 0.206 | 0.182 / 0.230 |
| No. atoms | | | |
| Protein | 1064 | 1062 | 1133 |
| Solvent | 96 | 100 | 79 |
| Ligand | 25 | 25 | 25 |
| Ion | 1 | 1 | 1 |
| B-factors ($Å^2$) | | | |
| Protein | 18.0 | 19.9 | 22.3 |
| Solvent | 22.6 | 25.7 | 27.7 |
| Ligand | 13.5 | 15.0 | 18.1 |
| Ion | 15.2 | 18.0 | 34.4 |
| RMS deviations | | | |
| Bonds (Å) | 0.019 | 0.018 | 0.016 |
| Angles (°) | 1.72 | 1.78 | 1.58 |
| PDB accession code | 3PMF | 3OSO | 3MVV |
| **RMS deviation from Δ+PHS** | | | |
| Backbone (Å) | 0.21 | 0.26 | 0.33 |
| Backbone and sidechain (Å) | 0.74 | 0.88 | 0.78 |

* Values in parentheses are for highest-resolution shell.

| | Δ+PHS L36A | Δ+PHS L38A | Δ+PHS V66A |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ |
| Cell dimensions | | | |
| *a, b, c (Å)* | 31.01, 60.40, 38.28 | 31.04, 60.58, 38.17 | 30.96, 60.65, 38.55 |
| *α, β, γ (°)* | 90.00, 93.77, 90.00 | 90.00, 93.29, 90.00 | 90,00 93.30, 90.00 |
| Resolution (Å) * | 50.00-1.70 (1.73-1.70) | 50.00-1.55 (1.58-1.55) | 50.00-1.90 (1.93-1.90) |
| $R_{merge}$ | 0.040 (0.253) | 0.052 (0.282) | 0.061 (0.303) |
| Average I/σ(I) | 21.9 (6.1) | 17.2 (5.0) | 16.8 (6.5) |
| Completeness (%) | 0.989 (0.926) | 0.992 (0.929) | 0.995 (1.000) |
| Redundancy | 7.1 (6.1) | 6.9 (5.8) | 7.1 (6.9) |
| **Refinement** | | | |
| Resolution (Å) | 30.94-1.70 (1.74-1.70) | 30.99-1.55 (1.59-1.55) | 30.91-1.90 (1.95-1.90) |
| No. of reflections | 15410 (1072) | 20355 (1290) | 11246 (737) |
| $R_{work}$ / $R_{free}$ | 0.189 / 0.230 | 0.188 / 0.238 | 0.186 / 0.250 |
| No. atoms | | | |
| Protein | 1047 | 1084 | 1039 |
| Solvent | 45 | 90 | 40 |
| Ligand | 25 | 25 | 25 |
| Ion | 1 | 1 | 1 |
| B-factors ($Å^2$) | | | |
| Protein | 29.5 | 22.3 | 31.1 |
| Solvent | 31.0 | 27.4 | 33.4 |
| Ligand | 25.5 | 18.8 | 27.4 |
| Ion | 25.8 | 21.5 | 34.8 |
| RMS deviations | | | |
| Bonds (Å) | 0.017 | 0.024 | 0.013 |
| Angles (°) | 1.68 | 2.22 | 1.45 |
| PDB accession code | 3NP8 | 3MHB | 3NQT |
| **RMS deviation from Δ+PHS** | | | |
| Backbone (Å) | 0.22 | 0.21 | 0.23 |
| Backbone and sidechain (Å) | 0.60 | 0.72 | 0.63 |

* Values in parentheses are for highest-resolution shell.

| | Δ+PHS V74A | Δ+PHS I92A | Δ+PHS L103A |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ |
| Cell dimensions | | | |
| *a, b, c (Å)* | 31.08, 60.44, 38.13 | 31.09, 60.14, 38.19 | 31.07, 60.50, 38.03 |
| *α, β, γ (°)* | 90.00, 93.30, 90.00 | 90.00, 93.57, 90.00 | 90.00, 93.65, 90.00 |
| Resolution (Å) * | 50.00-1.65 (1.68-1.65) | 50.00-1.50 (1.53-1.50) | 50.00-1.58 (1.61-1.58) |
| $R_{merge}$ | 0.046 (0.183) | 0.060 (0.397) | 0.044 (0.193) |
| Average I/σ(I) | 22.3 (9.8) | 15.2 (5.2) | 18.9 (7.7) |
| Completeness (%) | 0.988 (0.960) | 0.991 (0.925) | 0.978 (0.902) |
| Redundancy | 7.2 (6.5) | 12.2 (8.9) | 7.1 (6.1) |
| **Refinement** | | | |
| Resolution (Å) | 30.22-1.65 (1.69-1.65) | 31.00-1.50 (1.54-1.50) | 30.25-1.58 (1.62-1.58) |
| No. of reflections | 16787 (1086) | 22299 (2233) | 18883 (1133) |
| $R_{work}$ / $R_{free}$ | 0.196 / 0.244 | 0.186 / 0.234 | 0.190 / 0.238 |
| No. atoms | | | |
| Protein | 1054 | 1103 | 1069 |
| Solvent | 76 | 111 | 75 |
| Ligand | 25 | 25 | 25 |
| Ion | 1 | 1 | 1 |
| B-factors ($Å^2$) | | | |
| Protein | 27.8 | 23.9 | 23.3 |
| Solvent | 31.0 | 29.8 | 27.2 |
| Ligand | 22.8 | 20.4 | 20.0 |
| Ion | 24.7 | 22.6 | 21.8 |
| RMS deviations | | | |
| Bonds (Å) | 0.015 | 0.018 | 0.016 |
| Angles (°) | 1.59 | 1.83 | 1.62 |
| PDB accession code | 3NK9 | 3MEH | 3MZ5 |
| **RMS deviation from Δ+PHS** | | | |
| Backbone (Å) | 0.20 | 0.21 | 0.21 |
| Backbone and sidechain (Å) | 0.72 | 0.86 | 0.70 |

* Values in parentheses are for highest-resolution shell.

| | Δ+PHS L125A |
|---|---|
| **Data collection** | |
| Space group | $P2_1$ |
| Cell dimensions | |
| *a, b, c (Å)* | 30.95, 60.64, 38.18 |
| *α, β, γ (°)* | 90.00, 92.54, 90.00 |
| Resolution (Å) * | 50.00-1.65 (1.68-1.65) |
| $R_{merge}$ | 0.057 (0.207) |
| Average I/σ(I) | 20.3 (7.6) |
| Completeness (%) | 0.994 (0.939) |
| Redundancy | 8.2 (5.2) |
| **Refinement** | |
| Resolution (Å) | 30.92-1.65 (1.69-1.65) |
| No. of reflections | 16938 (1068) |
| $R_{work}$ / $R_{free}$ | 0.188 / 0.238 |
| No. atoms | |
| Protein | 1050 |
| Solvent | 56 |
| Ligand | 25 |
| Ion | 1 |
| B-factors (Å$^2$) | |
| Protein | 27.9 |
| Solvent | 31.6 |
| Ligand | 24.3 |
| Ion | 28.9 |
| RMS deviations | |
| Bonds (Å) | 0.017 |
| Angles (°) | 1.63 |
| PDB accession code | 3NXW |
| **RMS deviation from Δ+PHS** | |
| Backbone (Å) | 0.23 |
| Backbone and sidechain (Å) | 0.84 |

* Values in parentheses are for highest-resolution shell.

| Snase Variant | $T_m$ ($^oC$) | $\Delta V_u^0$ ($T_m$) (ml/mol) | $<\Delta V_u^0>$ ($20^oC$, NMR) (ml/mol) |
|---|---|---|---|
| I92A | 62.8 | -26.8 | -104 |
| L103A | 63.3 | -9.3 | -85 |
| L125A | 62.0 | -7.7 | -69 |
| V66A | 68.4 | 2.7 | -71 |
| □+PHS | 73.8 | 4.0 | -55 |
| WT Snase | 51.5 | -19 | -75 |

**Table S2**. **Thermodynamic values for $\Delta V_u$**

Values are for $\Delta V_u(T_m)$ from pressure perturbation calorimetry, carried out as described in [37]. The cavity variants are all in the □+PHS background. Values can only be compared directly for proteins that exhibit similar values for $T_m$. This is because the thermal expansivity for unfolded states of proteins is larger than for their folded state, hence $\Delta V_u$ is a strong function of temperature, becoming less negative with increasing temperature and even changing sign to become positive at higher temperature. However, consistent with the apparent $\Delta V_u$ values obtained from fluorescence, that for the I92A variant is significantly larger in magnitude than that observed for either the L103A or L125A variants, which, as in the case of the fluorescence data, exhibit similar values for $\Delta V_u$. The reference protein, $\Delta$+PHS, unfolds at much higher temperature due to it stability, and hence cannot be compared to the cavity variants. On the other hand, the true wild type nuclease, which unfolds at a lower temperature, exhibits a smaller, rather than larger, absolute value for $\Delta V_u$ compared to the I92A variant, consistent with a much greater effect of pressure on the cavity bearing mutants. In PPC, the heat absorbed or released upon a small (5 bar) change in pressure, at each temperature in a thermal scan, is related to the variation in volume with respect to the variation in temperature, or the thermal expansivity:

$$(dQrev/dp)_T = -T(dV/dT)_T = -TVa$$

For a transition from one state to another, integrating a over the transition temperature range yields the change in volume $\Delta V$ for the transition at Tm.

$$\Delta V/V = \int_{T_O}^{T_f} \alpha dT$$

Reference List


[1] J. J. Dwyer*, et al.*, "High apparent dielectric constants in the interior of a protein reflect water penetration," Biophys. J **79**(3), 1610 (2000).

[2] D. A. Karp*, et al.*, "High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp," Biophys. J **92**(6), 2041 (2007).

[3] D. Shortle and A. K. Meeker, "Residual structure in large fragments of staphylococcal nuclease: effects of amino acid substitutions," Biochemistry **28**(3), 936 (1989).

[4] S. T. Whitten and E. B. Garcia-Moreno, "pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state," Biochemistry **39**(46), 14292 (2000).

[5] Z. Otwinowski and W. Minot, in "Macromolecular Crystallography," 1997, pp.307-326.

[6] A. J. McCoy*, et al.*, "Likelihood-enhanced fast translation functions," Acta Crystallogr. D. Biol. Crystallogr. **61**(Pt 4), 458 (2005).

[7] "Collaborative Computational Project Number 4," Acta Crystallogr. D. Biol. Crystallogr. **50**, 760 (1994).

[8] E. B. Krissinel*, et al.*, "The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography," Acta Crystallogr. D. Biol. Crystallogr. **60**(Pt 12 Pt 1), 2250 (2004).

[9] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method," Acta Crystallogr. D. Biol. Crystallogr. **53**(Pt 3), 240 (1997).

[10] A. A. Vaguine, J. Richelle, and S. J. Wodak, "SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model," Acta Crystallogr. D. Biol. Crystallogr. **55**(Pt 1), 191 (1999).

[11] R. A. Laskowski*, et al.*, "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR," J Biomol. NMR **8**(4), 477 (1996).

[12] V. B. Chen*, et al.*, "MolProbity: all-atom structure validation for macromolecular crystallography," Acta Crystallogr. D. Biol. Crystallogr. **66**(Pt 1), 12 (2010).

[13] J. Painter and E. A. Merritt, "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion," Acta Crystallogr. D. Biol. Crystallogr. **62**(Pt 4), 439 (2006).

[14] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," J Mol Graph. **14**(1), 33 (1996).

[15] J. B. Rouget*, et al.*, "Size and sequence and the volume change of unfolding," J. Am. Chem. Soc. (2011).

[16] C. A. Royer, W. R. Smith, and J. M. Beechem, "Analysis of binding in macromolecular complexes: a generalized numerical approach," Anal. Biochem. **191**(2), 287 (1990).

[17] M. Piotto, V. Saudek, and V. Sklenar, "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions," J Biomol. NMR **2**(6), 661 (1992).

[18] V. Sklenar, "Selective excitation techniques for water suppression in one- and two-dimensional NMR spectroscopy," Basic Life Sci. **56**, 63 (1990).

[19] J. L. Pons, T. E. Malliavin, and M. A. Delsuc, "Gifa V. 4: A complete package for NMR data set processing," J Biomol. NMR **8**(4), 445 (1996).

[20] A. Bax and S. S. Pochapsky, "Optimized Recording of Heteronuclear Multidimensional NMR Spectra Using Pulsed Field Gradients ," J. Magn Reson **99**, 638 (1992).

[21] D. Marion*, et al.*, "Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta," Biochemistry **28**(15), 6150 (1989).

[22] G. Baudenhausen and D. J. Ruben, "Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy," **69**, 185 (1980).

[23] SD van der Spoel*, et al.*, "GROMACS: fast, flexible, and free.," **98**, 2712 (2005).

[24] WL Jorgensen and J. Tirado-Rives, "The OPLS force field for proteins, energy minimizations for crystals of cyclic peptides and crambin. " **110**, 1657 (1988).

[25] A. Fiser and A. Sali, "Modeller: generation and refinement of homology-based protein structure models," Methods Enzymol. **374**, 461 (2003).

[26] HJC Berendsen, JR Grigera, and TP Straatsma, "The missing term in effective pair potentials," **91**, 6269 (1987).

[27] B Hess*, et al.*, "LINCS: A linear constraint solver for molecular simulations," **18**, 1463 (1997).

[28] S Miyamoto and PA Kollman, "SETTLE - An analytical version of the SHAKE and RATTLE algorithm for rigid water models," **13**, 952 (1992).

[29] T. A. Darden and L. G. Pedersen, "Particle mesh Ewald: an N.Log(N) method for ewald sums in large systems.," **98**, 10089 (1993).

[30] WG Hoover, "Canonical Dynamics: Equilibrium Phase space distributions," **31**, 1695 (1985).

[31] S Nosé, "A molecular dynamics method for simulating the canonical ensemble," **52**, 255 (1984).

[32] M Parrinello and A Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method.," **52**, 7182 (1981).

[33] M. S. Till and G. M. Ullmann, "McVol - a program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm," J Mol Model. **16**(3), 419 (2010).

[34] J. K. Noel*, et al.*, "SMOG@ctbp: simplified deployment of structure-based models in GROMACS," Nucleic Acids Res. **38**(Web Server issue), W657-W661 (2010).

[35] A. M. Ferrenberg and R. H. Swendsen, "New Monte Carlo technique for studying phase transitions," Phys Rev Lett **61**(23), 2635 (1988).

[36] A. M. Ferrenberg and R. H. Swendsen, "Optimized Monte Carlo data analysis," Phys Rev Lett **63**(12), 1195 (1989).

[37] L. Mitra*, et al.*, "Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution--experiments and theoretical interpretation," Phys. Chem. Chem. Phys. **8**(11), 1249 (2006).

[38] The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

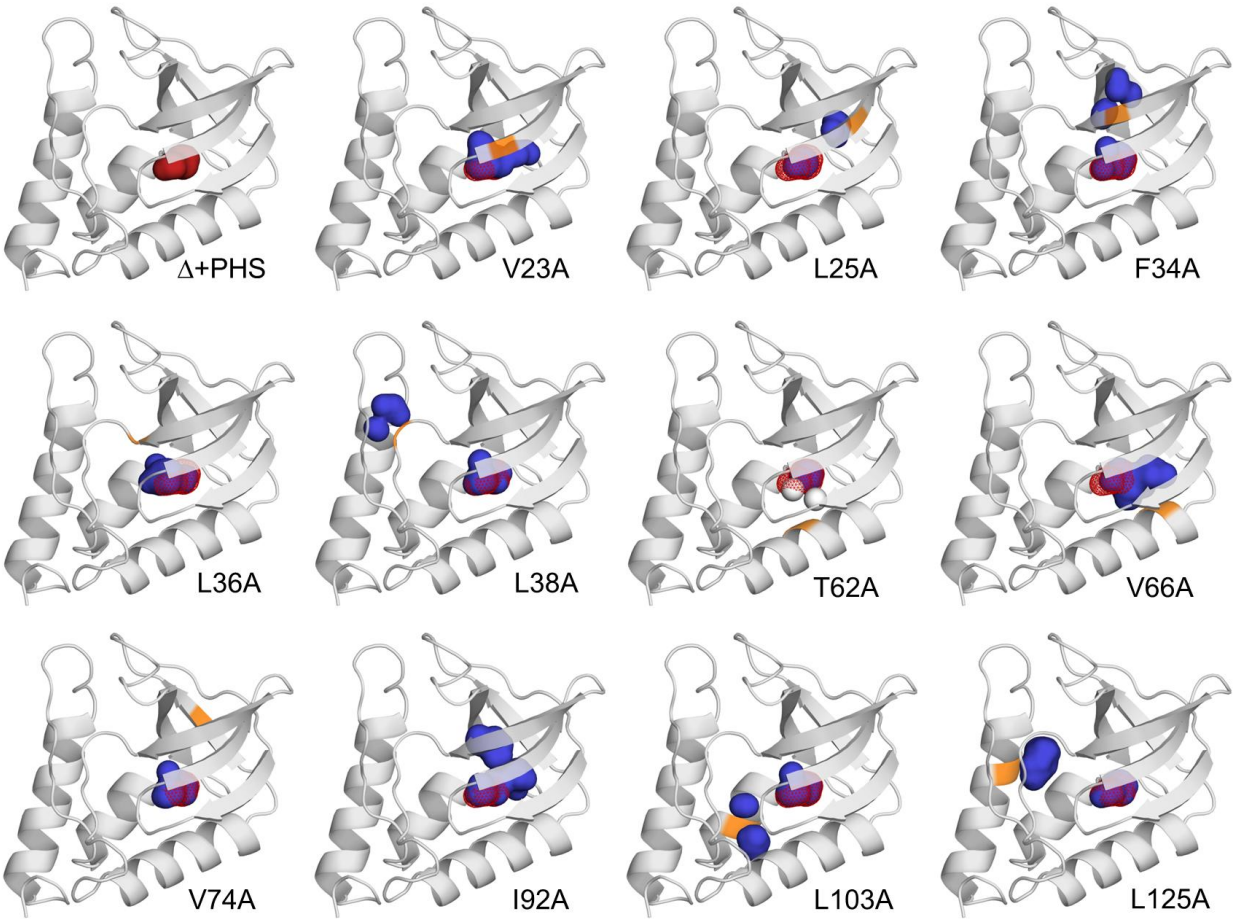


**Figure S1**: Crystal structures of Δ+PHS and of the eleven cavity-containing variants. The position of the substitution to Ala is highlighted in orange. The backbone of the Δ+PHS protein is shown in all structures. The original microcavity is shown in red. The cavity created by the substitutions is shown in blue. The T62A variant showed two internal structural water molecules (white) and the V74A variant did not display a well defined cavity near the substitution site. Surfaces drawn with a probe radius of 1.4 Å using PyMol[38] . We note that the cavities in the main text were detected using a probe of 1.4 Å.

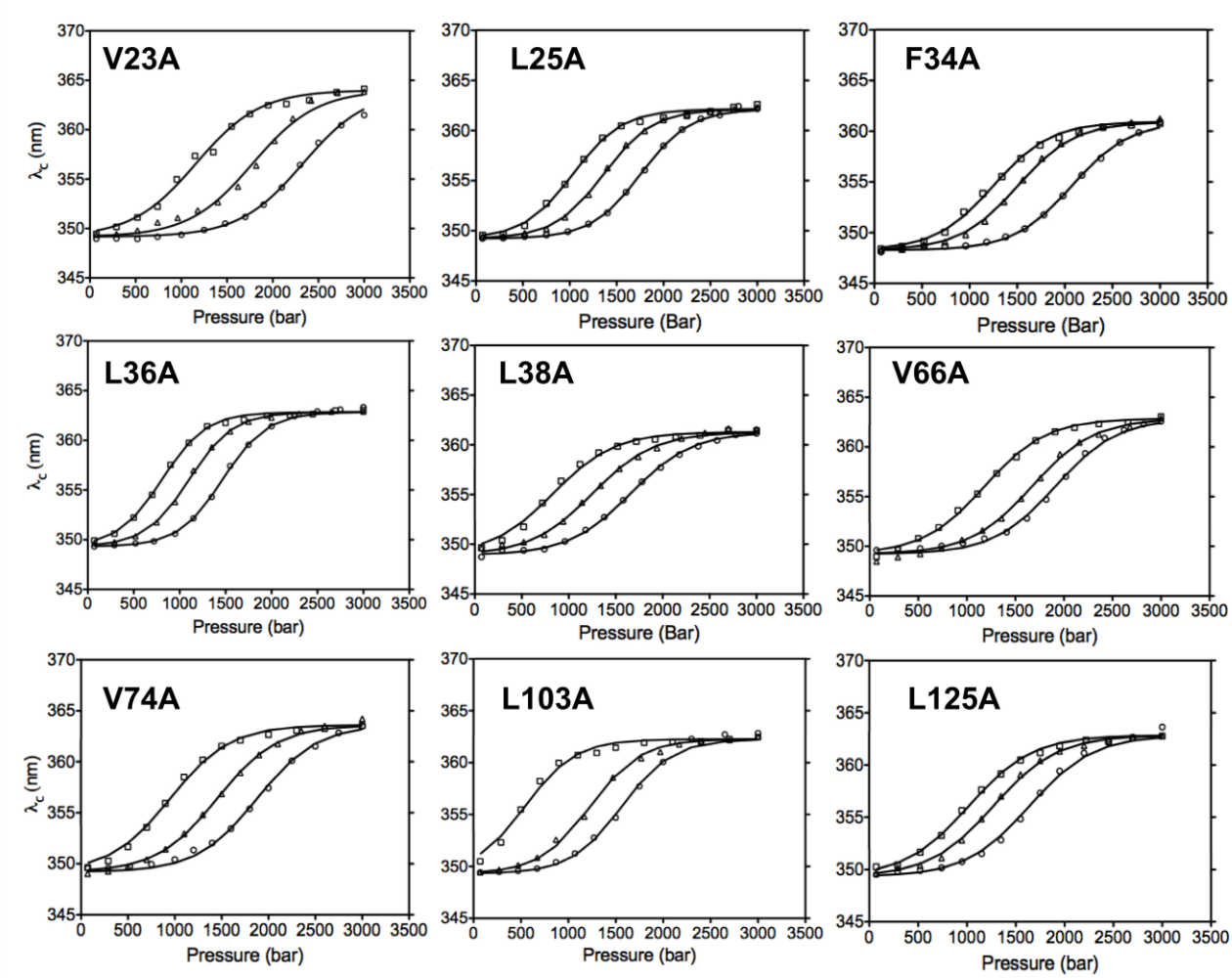


**Figure S2**: High-pressure fluorescence average emission wavelength profiles of the cavity mutants proteins. For every mutant, the three curves correspond to the three different concentrations of GuHCl used: V23A (0.8M, 1.1M, 1.4M), L25A (1.0M, 1.2M, 1.4M), F34A (0.8M, 1.0M, 1.2M), L36A (0.9M, 1.1M, 1.3M), L38A (1.6M, 1.8M, 2.0M), V66A (1.2M, 1.4M, 1.8M), V74A (1.0M, 1.2M, 1.4M), L103A (0.8M, 1.0M, 1.4M), L125A (0.8M, 1.0M, 1.2M), for circle, triangle and square symbols respectively. Lines through the points represent fits of the data to a two-state unfolding model, correcting the average emission wavelength values for the quantum yield change upon unfolding.

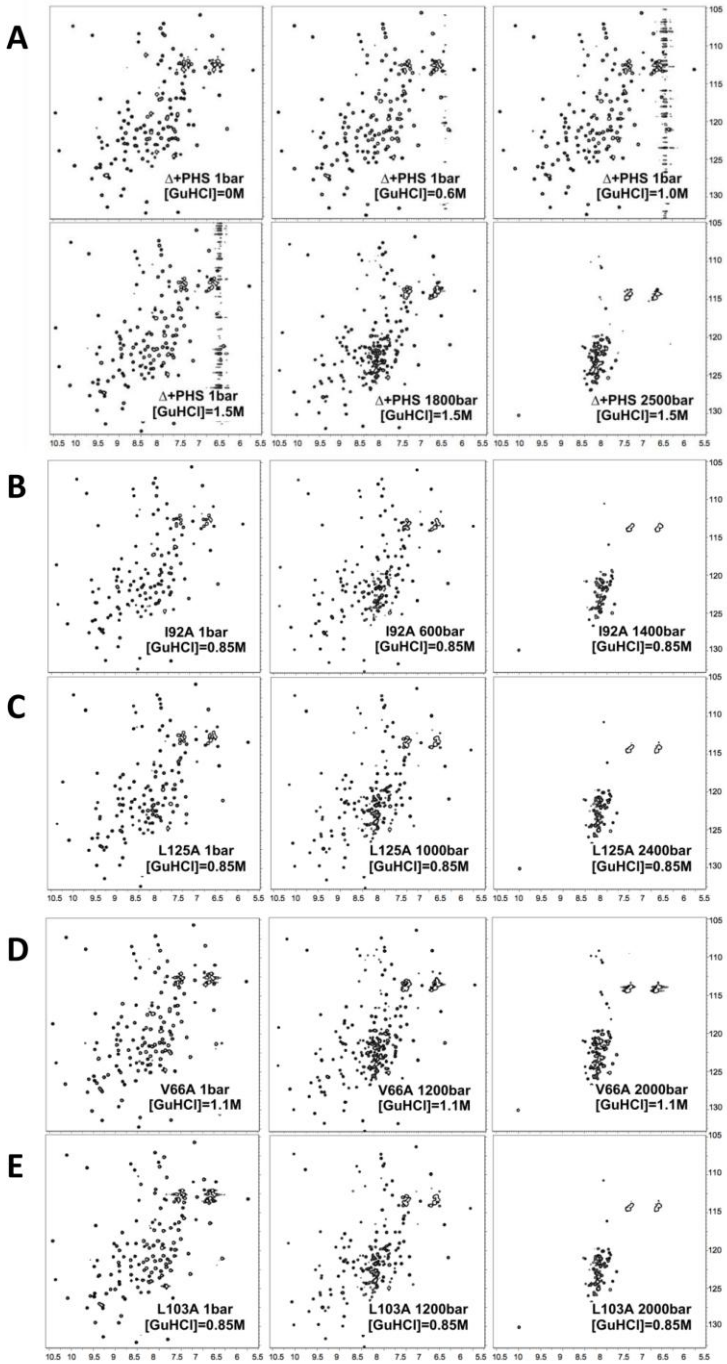


**Figure S3**: HSQC spectra of the Δ+PHS and cavity variants at 293K. A) Δ+PHS *Upper panel*: HSQC spectra recorded with, left to right, three concentrations of GuHCl. The attribution of the initial HSQC spectrum without GuHCl was done through a classical sequential attribution procedure. The GuHCl effect on the chemical shift of each cross peak was then followed with a series of HSQCs up to a final concentration of 1.5M. *Lower panel*: HSQC spectra with a GuHCl concentration of 1.5M at, left to right, 1bar (fully folded), 1800bar (the midpoint of the pressure unfolding) and 2500bar (fully unfolded). B) HSQC spectra of the I92A cavity variant with a GuHCl concentration of 0.85M at, left to right, 1bar (fully folded), 600bar (the midpoint of the pressure unfolding) and 1400bar (fully unfolded). C) HSQC spectra of L125A with a GuHCl concentration of 0.85M at, left to right, 1bar (fully folded), 1000bar (the midpoint of the pressure unfolding) and 2400bar (fully unfolded), D) HSQC spectra of V66A with a GuHCl concentration of 1.1M at, left to right, 1bar (fully folded), 1200bar (the midpoint of the pressure unfolding) and 2000bar (fully unfolded). E) HSQC spectra of L103A with a GuHCl concentration of 0.85M at, left to right 1bar (fully folded), 1200bar (the midpoint of the pressure unfolding) and 2000bar (fully unfolded).

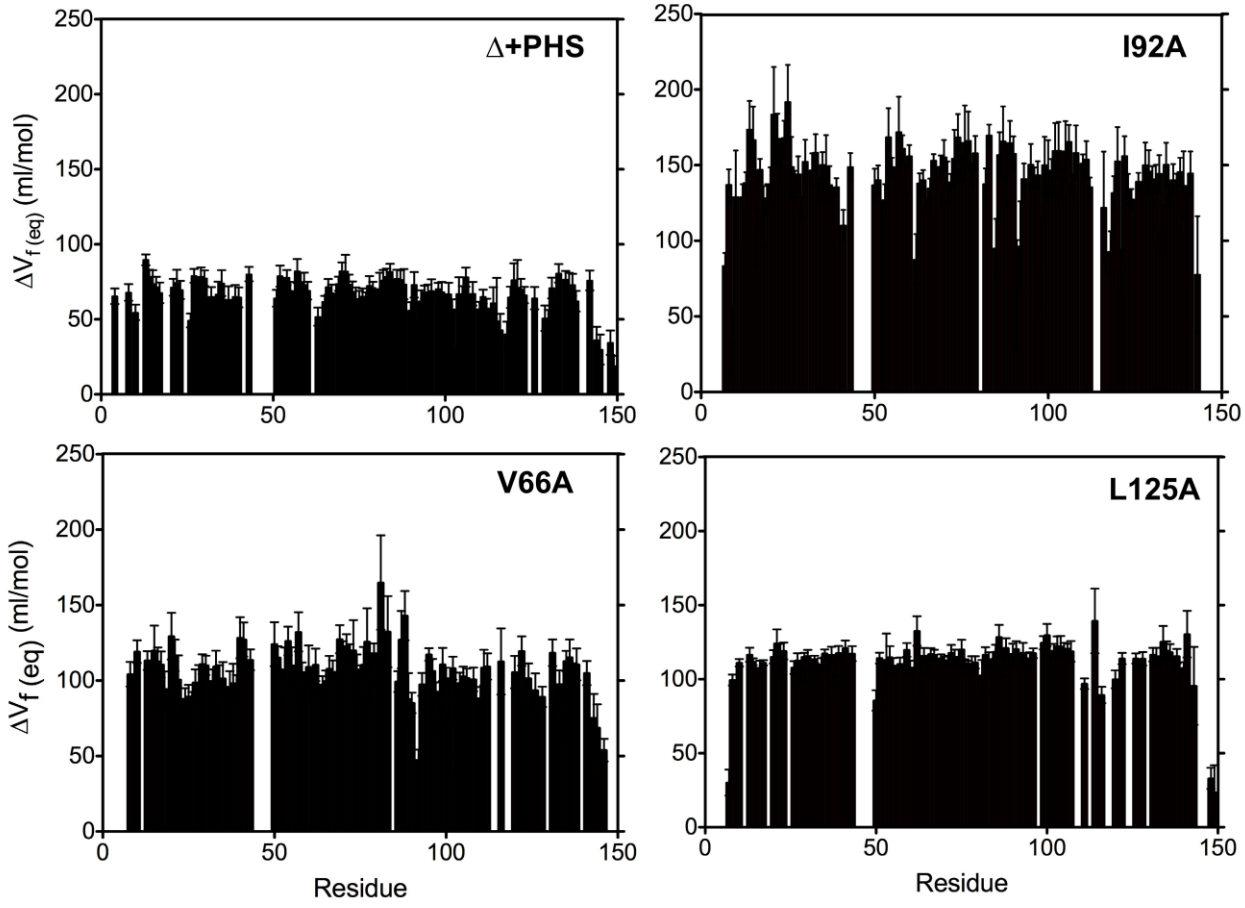


**Figure S4**: $\Delta V_f$ (=-$\Delta V_u$) values obtained from high-pressure NMR experiments as function of the protein sequence for the Δ+PHS variant and three cavity mutants (I92A at 0.85 M GuHCl, V66A at 1.1 M GuHCl and L125A at 0.85 M GuHCl). Residues from position 44 to 49 correspond to the active site loop of the wild-type SNase and are deleted in all of the Δ+PHS variant proteins. The average standard deviation on the fitted value of the $\Delta V_f$ is 6 ml/mol for all residues of Δ+PHS, 14 ml/mol for I92A, 10 ml/mol for V66A and 6 ml/mol for L125A. Not shown are data for L103A for which the average standard deviation is 8 ml/mol.

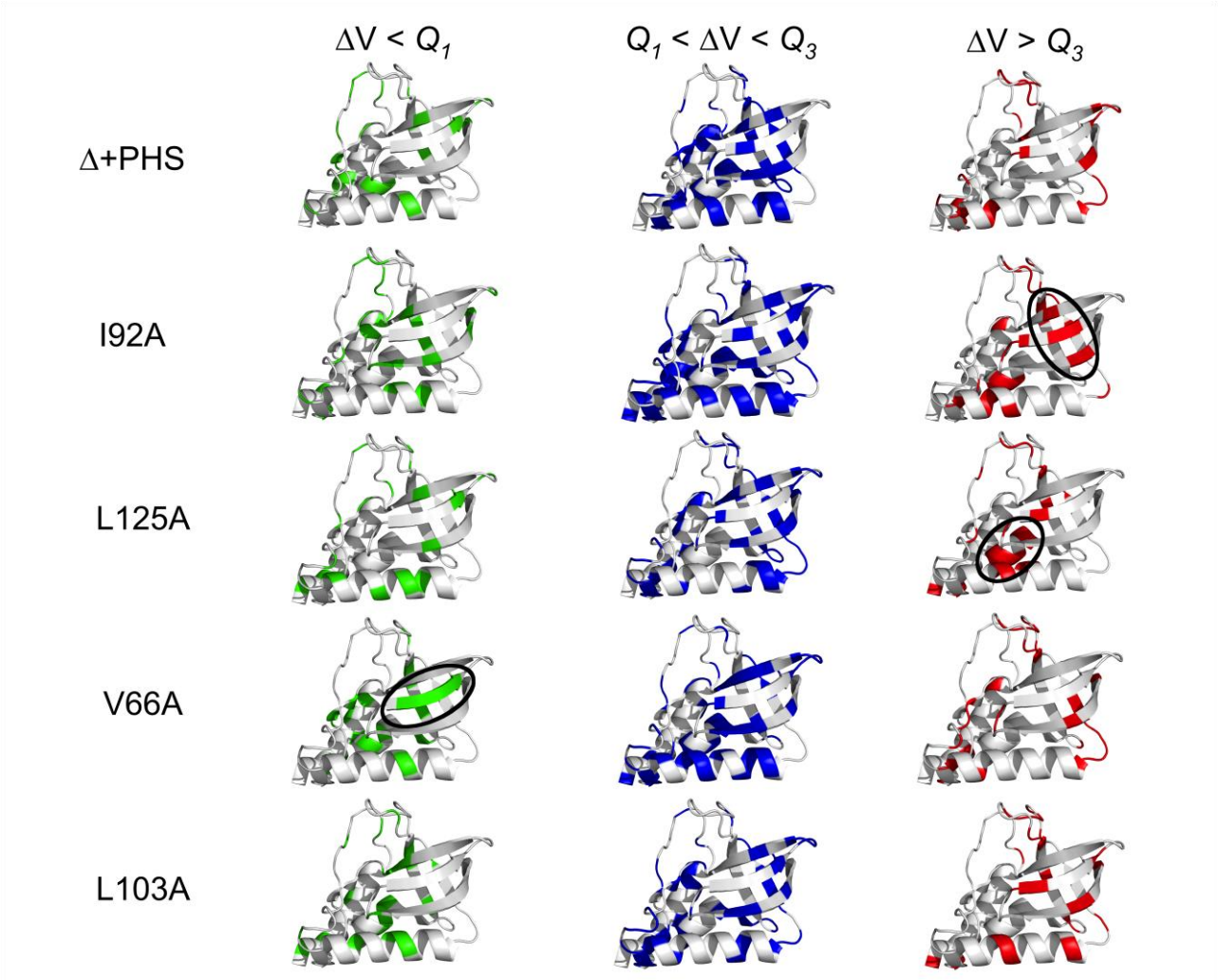


**Figure S5**: Structural mapping of the $\Delta V_f$ (=-$\Delta V_u$) values obtained from high-pressure NMR experiments for Δ+PHS (1.5M GuHCl), I92A (0.85M GuHCl), L125A (0.85M GuHCl), V66A (1.1M GuHCl) and L103A (0.85M GuHCl). Residues of the first quartile $Q_1$ (the 25% lowest values) are in green, residues of the third quartile $Q_3$ (the 25% highest values) are in red and residues lying between the first and third quartile (50 % of the values) are in blue. Comparing the cavity variants to Δ+PHS it can be seen that for I92A, residues in the core region in helix 2 and the β-barrel lining the extended cavity created by the I92A mutation exhibit very large $\Delta V_f$ values that are in the top quartile, indicating that these residues only sample unfolded-like environments when the entire ensemble is unfolded. The same is true for residues in helix 2 for L125A, as this region connects the cavity created by the L125A mutation with the naturally occurring cavity in the core. The most apparent effect of the V66A mutation is to destabilize β strand 2, which shows very low $\Delta V_f$ values compared to the average, indicating that these residues sample disrupted states whereas much of the rest of the core remains intact. For L103A, while all residues showed significantly larger $\Delta V_f$ values than for Δ+PHS, no major changes in their distribution was observed.

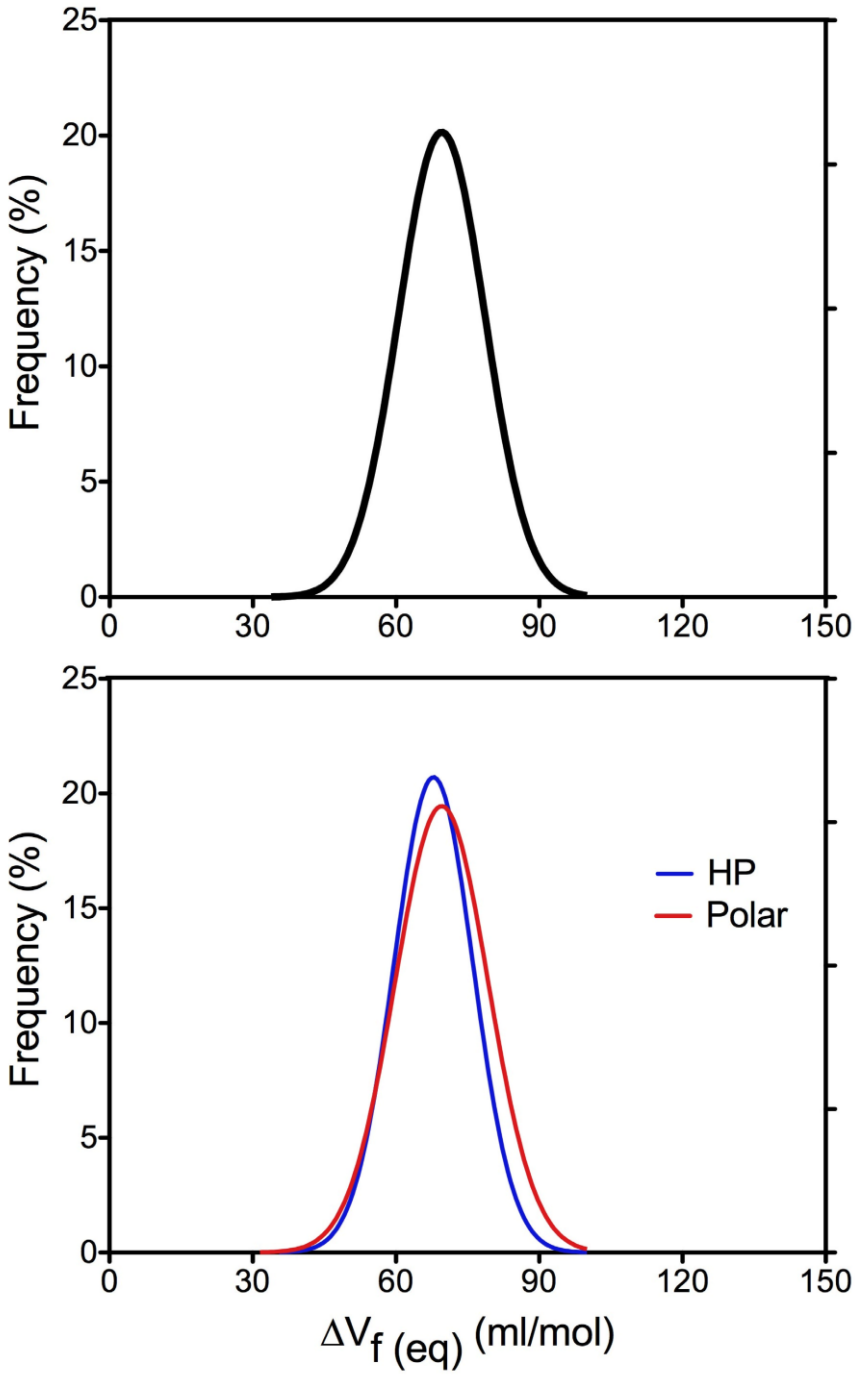


**Figure S6**: Gaussian fits of the $\Delta V_f$ (=-$\Delta V_u$) value distributions obtained from the Δ+PHS variant high-pressure NMR experiments, for the complete set of residues (*top*), for the hydrophobic residues only (*bottom*, blue curve) and for the polar residues only (*bottom*, red curve). Over a total of 101 $\Delta V_f$ values, 32 residues were classified as hydrophobic (A, L, I, V, F) and 54 as polar (R, K, D, E, N, Q, S, T, M).

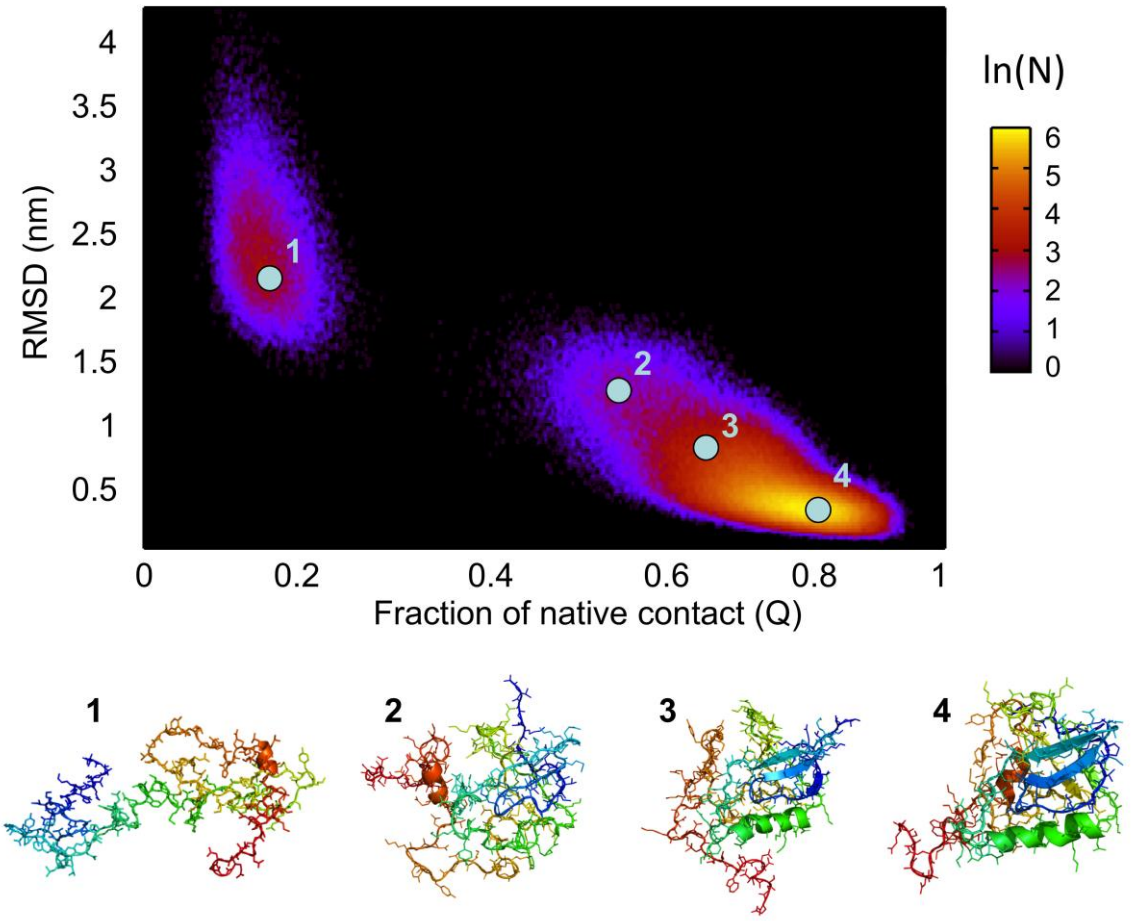


**Figure S7**: Two-dimensional (Root Mean Square Deviation, RMSD vs Q) representation of the complete set of conformations (N=400,000) obtained from 100 ns-long Go-model simulations based on the fractional contact map of Δ+PHS at 800bar. Representative conformations of the unfolded states (1), the pseudo-native basins (2 and 3) and the native states (4) are displayed at the bottom of the figure.