Biophysical Journal, Volume *111*

Supplemental Information

Molecular Effects of Concentrated Solutes on Protein Hydration, Dynamics, and Electrostatics

Luciano A. Abriata, Enrico Spiga, and Matteo Dal Peraro

Supporting Material for

Molecular Effects of Concentrated Solutes on Protein Hydration, Dynamics and Electrostatics

L. A. Abriata,[*] E. Spiga and M. Dal Peraro[*]

Additional Methods for NMR Experiments

Titrations were monitored through 2D phase-sensitive, ^{1}H-detected ^{1}H-^{15}N HSQC spectra using water flip-back pulses and gradient selection with decoupling during acquisition (pulse program hsqcetf3gpsi from the Bruker library, the closest literature description is that by Schleucher, Schwendinger, Sattler, Schmidt, Schedletzky, Glaser, Sorensen and Griesinger *J. Biomol. NMR* 1994). The resolution at acquisition was 2k points in the direct ^{1}H dimension and 128 points in the indirect ^{15}N dimension, on sweep widths of 12 and 32 ppm, respectively, and centered at 4.699 and 115 ppm, respectively. Spectra were processed with 2k points in ^{1}H and 256 points in ^{15}N, using linear prediction through 32 coefficients and a sine bell function in the ^{15}N dimension. A baseline of polynomial degree 5 was applied. At each point of the titration, we optimized the lock phase, the shim currents and the hard pulse length for ^{1}H. For all solutes but arginine and glutamate, the pulse length remained close to its value in buffer, *i.e.* around 11 μs. For arginine and glutamate, the 90º pulse for ^{1}H increased from ~11 μs to ~17 μs (glutamate) or ~18 μs (arginine) at the highest concentrations. We further point out that these titrations are reproducible and that the differences observed for CSPs among residues within spectra are well beyond the uncertainties in the chemical shift measurements (except, as indicated in the text, for the Ficoll crowders and for several residues in the titrations with proteins).

^{15}N relaxation data (T_1 and T_2) were acquired through pseudo-3D (direct ^{1}H dimension and indirect ^{15}N dimension, plus an extra dimension for time delays) phase-sensitive spectra with water flip-back pulses, gradient selection and decoupling during acquisition (Bruker pulse programs hsqct1etf3gpsi3d which uses inversion recovery to measure T_1, and hsqct2etf3gpsitc3d which uses interleaved temperature-compensated Carr–Purcell–Meiboom–Gill blocks to measure T_2). The relaxation delays for T_1 measurement were 8, 64, 136, 232, 336, 472, 664, 800 and 1600 ms for the determination of longitudinal relaxation times; and those for T_2 measurement were 33.9, 67.8, 101.8, 135.7, 169.6, 203.5, 237.4 and 271.4 ms. These values were chosen based on previous work (Tjandra, Feller, Pastor and Bax *J. Am. Chem. Soc.* 1995) with the addition of the 1600 ms delay for T_1 measurement which was especially important at high solute concentrations of small molecules. A recycle delay of 5 seconds was employed in all experiments given the long T_1 values operative under some of the high-concentration conditions. Acquisition and processing parameters of the ^{1}H-^{15}N HSQC planes in these pseudo-3D spectra was as described above for HSQC spectra. We further notice that we only measured ^{15}N relaxation on selected systems containing high-solute concentrations but with low ionic strength, where hard pulses remain as low as in buffer.

^{15}N-resolved NOESY (or NOESY-HSQC) spectra were acquired with a 3D (direct ^{1}H dimension, indirect ^{15}N dimension and indirect ^{1}H NOE dimension) ^{1}H-detected phase-sensitive spectrum with gradient selection and decoupling during acquisition (pulse program noesyhsqcetgp3d of the Bruker library, built up from Marion, Kay, Sparks, Torchia and Bax, *J. Am. Chem. Soc.* 1989). The sweep widths were of 14, 32 and 14 ppm centered at 4.699, 115 and 4.699 ppm, with 2k, 48 or 64 detected points and 2k, 64 and 128 points for processing, all respectively for the direct ^{1}H dimension, the indirect ^{15}N dimension and the indirect ^{1}H dimension. The mixing time (during which both NOEs and exchange processes can build crosspeak intensity) was 40 ms as used in similar works testing ubiquitin hydration dynamics (Nucci, Pometun and Wand *Nat. Struct. Mol. Biol.* 2011).

Additional Methods for Molecular Dynamics Simulations

The equilibration phase of each simulation involved 5 steps. First, conjugate gradient energy minimization with constrained Cα atoms. Second, warm-up from 10 K to 300 K in 30 K increments along 10000 steps followed by 20000 further steps at 300 K, with a timestep of 1 fs and constraints on the Cα atoms to their initial positions. Third, equilibration at 300K and 1 atm with constrained Cα atoms with a time step of 1 fs for 2 ns. Constraints on Cα atoms were then removed and the systems warmed up again from 10 to 300 K in 30 K increments along 10000 steps, followed by 20000 steps additional 1 fs steps at 300 K. This was then held at 300 K for 20,000 additional 2 fs steps, continuing into the production runs. The temperature was controlled using a Langevin thermostat with a damping coefficient of 5/ps. The pressure was kept at 1 atm using the Berendsen barostat with a relaxation time constant of 1 ps (warm-ups) or with a Langevin-Nosé-Hoover barostat with barostat oscillation time of 200 fs and damping time of 100 fs (constant temperature equilibrations and production runs). Non-bonded interactions (van der Waals and Coulombic) were treated with a cut-off of 12 Å with the construction of a neighbor list having pair list distance of 13.5 Å. Outside the cut-off the Coulombic interactions are calculated using the Particle Mesh Ewald method as implemented in NAMD.

Root mean squared fluctuations (RMSF) were computed for each Cα atom as:

$$RMSF_i = \sqrt{\frac{\sum_j \left(r_{i,j} - \bar{r}_i\right)^2}{N}}$$

where r_i denotes the position of Cα atom *i*, and *j* runs through the analyzed frames of the simulation. Calculations were done after alignment to the starting structure.

Translational diffusion coefficients were computed from the mean square displacements of each protein's center of mass during the simulation:

$$MSD(t')_{protein} = \frac{1}{t_{\max}} \sum_{t_0=1}^{t_{\max}} (\bar{r}(t_0 + t') - \bar{r}(t_0))^2$$

(where t_{max} is the simulation time in frames and *r* denotes positions of the protein's center of mass) by taking the slope of this time series and dividing by 6.

Interactions between water or solute molecules and the protein were investigated by computing survival probabilities for water-protein and sugar-protein contacts. Two atoms were considered to be in contact when the distance between them was lower than 1.1 times the sum of their Van der Waals radii as optimized in recent works. The survival probability was computed as reported by Marchi's group based on Impey's original formulation for ion hydration:

$$N(t) = \frac{1}{N_t} \sum_{n=1}^{Nt} \sum_j P_j(t_n, t)$$

where $Pj(t_n,t)$ takes the values of 1 if the j^{th} water/solute molecule is in contact with the protein between time t_n and t_n+t, and zero otherwise, and N_t is the number of frames. The decay in the survival probability for water was then fitted to a stretched exponential (where the stretching parameter γ varies from 0 for non-ideal diffusion to 1 for ideal diffusion) combined with two or more simple exponentials, as described by Marchi:

$$N(t) = n_s \exp(-(t/\tau_s)^{\gamma_S}) + \sum_{i=2}^{4} n_i \exp(-(t/\tau_i))$$

where n_s and n_i are the number of water/solute molecules with residence times τ_s and τ_i on the surface of the protein on each timescale (time scales reported in Table S2).

The radial distribution functions of water and solute molecules around the protein surface were evaluated using VMD's (Humphrey et al *J. Mol. Graphics* 1996) Radial Distribution Plugin.

Table S1. Properties of the tested solutes, and main results from NMR experiments and MD simulations.

Solute at 300 g/L	Structure, molar mass and polarity	Molar conc.	Dynamic viscosity η, 300K (mPa s)	Dielectric constant ε at 300K in concentrate aqueous solution	NMR					MD	
					CSP at 300 g/L						
					Range (ppm)	Trend vs. SASA(N)	τ_c (ns)	NOESY to water	R_1R2 median ± median deviation	Average number of molecules in contact with protein	Timescale for full exploration of conform-ational space
(none)	--		0.798	78			4.2	None	13 ± 1		Few tens of ns
Glucose	Polyol, 180 g/mol	1.67	2.21	62	0.060 – 0.173	+	10.2	Strong	14 ± 1	75	> us
Glycerol	Polyol 92 g/mol	3.26	1.87	68	0.027 – 0.138	+	(N.D.)	Strong	(N.D.)		
Sodium glutamate	Negative Zwitterion 169 g/mol	1.78	N.A.	N.A., but >>78 according to MD	0.065 – 0.184	-	(N.D.)	Strong	(N.D.)	34	Hundreds of ns
Arginine hydro-chloride	Positive Zwitterion 210.7 g/mol	1.42	1.3		0.063 – 0.185	-	(N.D.)	Strong	(N.D.)	33	Hundreds of ns
Glycine	Neutral Zwitterion 75.1	3.99	1.77	166	0.241 – 0.374	-	6.6	Strong	14 ± 1		
Ficoll 70kDa	Sucrose polymer, avg. 70 kDa	0.0043	20-30	N.A.	0.012 – 0.045	+	12.3	Strong, for few residues	17 ± 2		
Ficoll 400kDa	Sucrose polymer, avg. 400 kDa	$7.5\ 10^{-4}$ (avg.)	50-60	57	0.013 – 0.034	+	11.9	Strong, for few residues	17 ± 3		
PEG 8kDa	Polymer of etylene oxide, avg. 8 kDa	0.038 (avg.)	20-50	~50	0.011 – 0.121	+	15.1	Weak, for few residues	15 ± 4		
PEG 35kDa	Polymer of etylene oxide, avg. 35 kDa	0.0086 (avg.)	N.A. but >> 50	N.A.	0.013 – 0.121	+	14.5	Weak, for few residues	15 ± 2		
Lysozyme (200 g/L)	14.3 kDa	0.011 (avg.)	1.6	N.A. but << 78	0.001 – 0.051	?	N.D. but likely high	Weak, for few residues	(N.D.)		
Bovine serum albumin	66.5 kDa	0.0045 (avg.)	4.8	N.A. but ~90 at 70 g/L	0.008 – 0.053	?	N.D. but likely high	Weak, for few residues	(N.D.)		

N.D. = Not determined; N.A. = Not Available

Table S2. Analysis of solute-protein and water-protein contacts in MD simulations.

	Water	Glucose	Arginine	Glutamate
Water exchange				
Fastest timescale (ps)	30	170	200	200
Second fastest (ns)	0.14	0.7	1	2
Third fastest (ns)	0.92	7	10	17
Slowest timescale (ns)	13	100	100	180
Average number of solutes interacting with protein surface	--	75	33	34
Dielectric constant	77.6	69.5	>> 78	>> 78
Range of rotational diffusion times	1.8 ns	> 1000 ns	23 – 31 ns	129 – 664 ns
	From Spiga et al. *J. Phys. Chem. B* 2014	(Data for solutes from simulations in this work)		

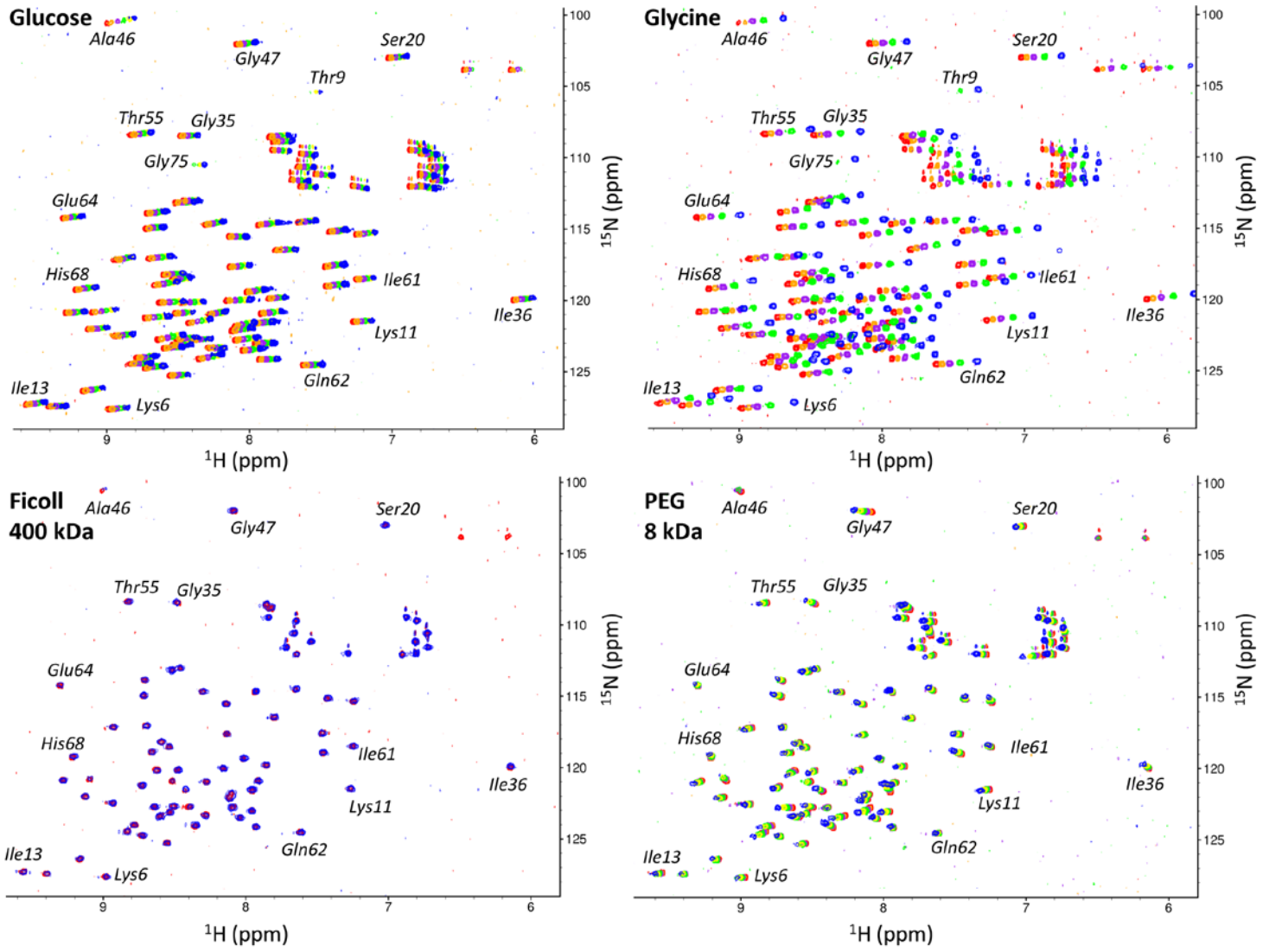

Figure S1. Effects on ^{1}H,^{15}N HSQC spectra (internally referenced to the water resonance) of stepwise additions of glucose, glycine, Ficoll 400 kDa or PEG 8 kDa to a 200 µM solution of ^{15}N-labeled ubiquitin, reaching in all cases 300 g/L final solute concentrations. In the four cases, the spectrum colored red is in pure buffer; then, color codes are:

- Glucose: orange = 50 g/L, violet = 150 g/L, green = 200 g/L, yellow = 250 g/L, blue = 300 g/L.
- Glycine: orange=45 g/L, violet=91 g/L, green=166 g/L, blue=300 G/L.
- Ficoll 400 kDa: blue is in 300 g/L.
- PEG 8 kDa: green=100 g/L, yellow=200g/L, blue=300 g/L

In the four spectra, the crosspeak for Alanine 46 is folded in the ^{15}N dimension.

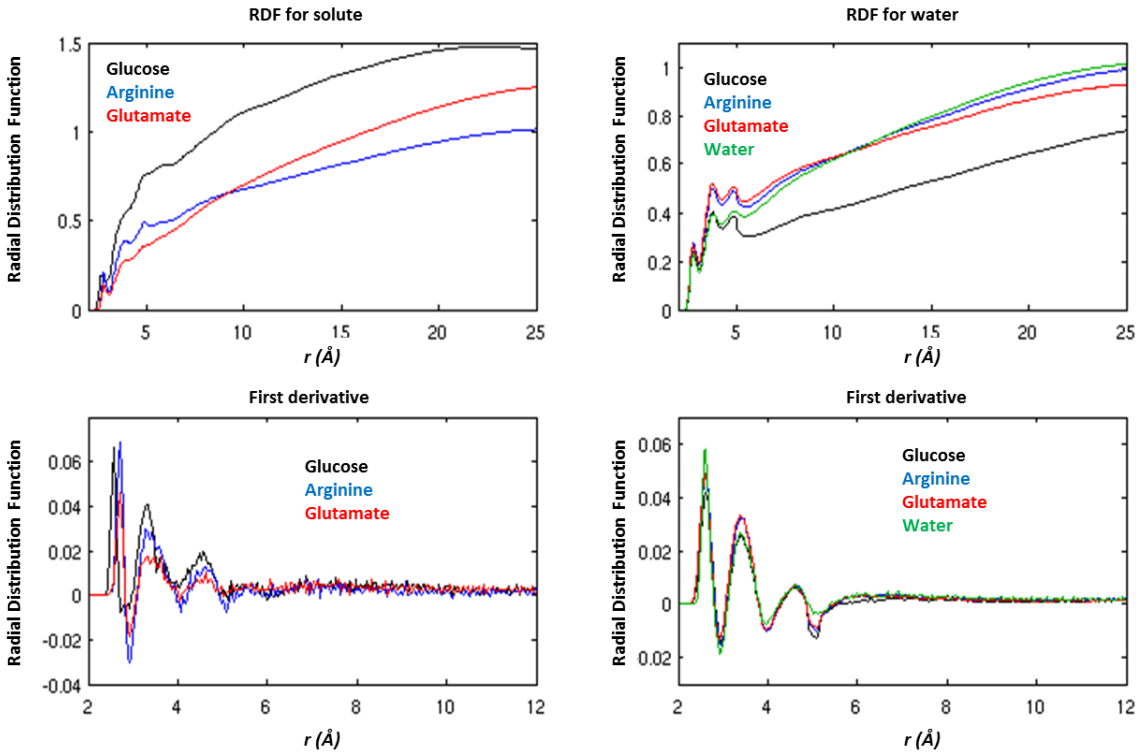

Figure S2. Radial distribution functions (top) for solutes (left) and water (right) relative to the protein surfaces, and first derivative with respect to the distance (bottom).

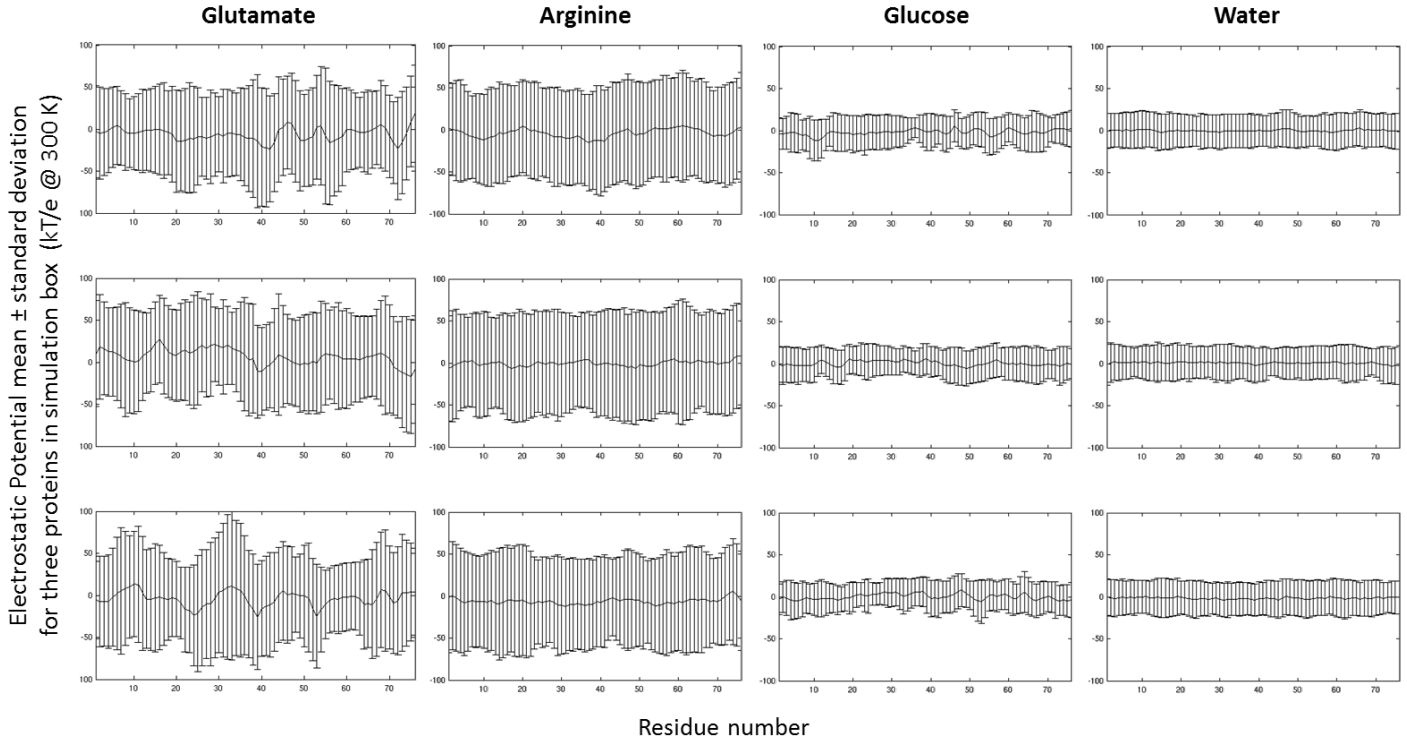

Figure S3. Trajectory average and standard deviation of the electrostatic potentials at each N atom of ubiquitin, for three proteins in the simulation boxes in water, glucose, glutamate and arginine.

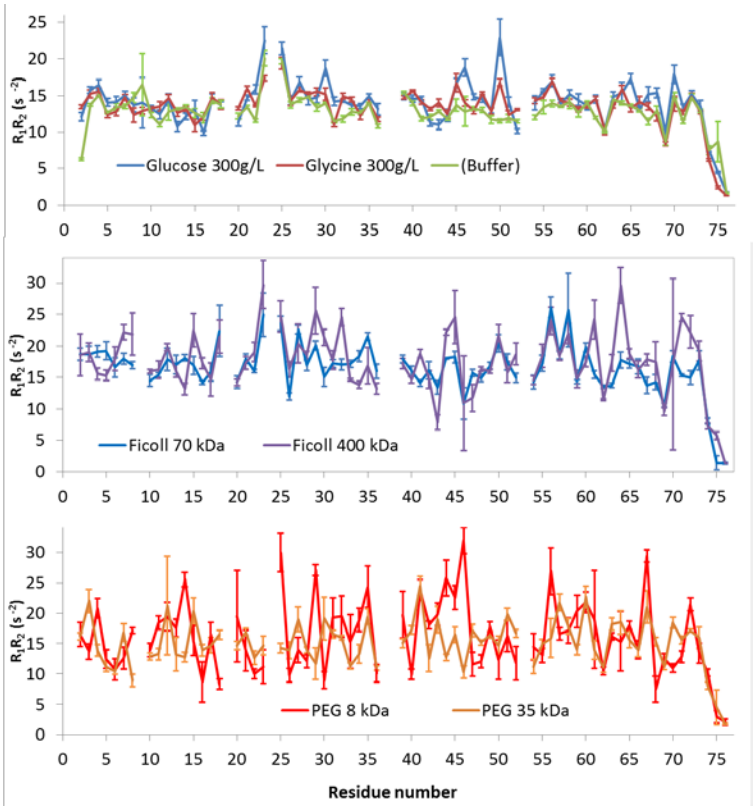

Figure S4. R_1R_2 products against ubiquitin's sequence, in buffer, concentrated glucose, glycine, Ficoll polymers and PEG polymers. Gaps correspond to weak or absent cross peaks and three prolines in ubiquitin's sequence. Error bars correspond to one standard error propagated from T1 and T2 uncertainties.

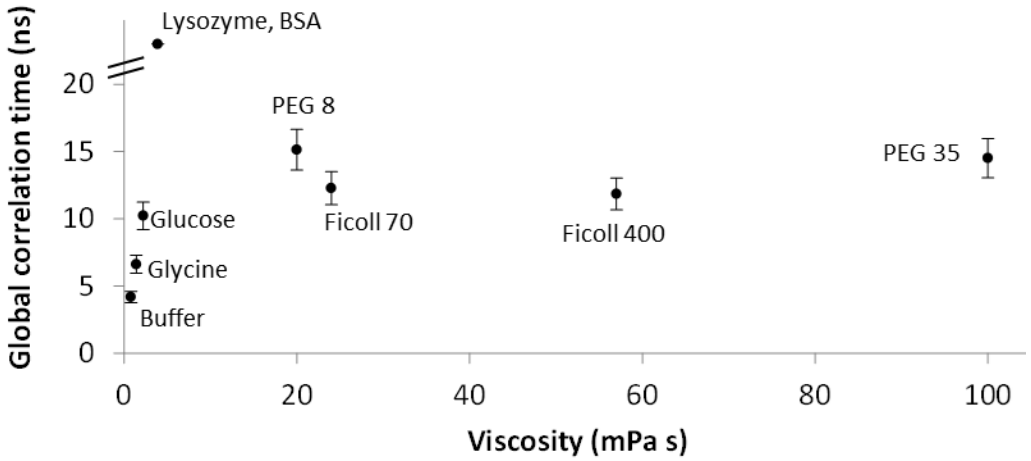

Figure S5. Global correlation time against viscosity for ubiquitin solutions studied by ^{15}N relaxation at 300 g/L of the added solutes. In lysozyme and BSA we do not report actual values of the global correlation times, we simply indicate that it will be larger than 20 ns given the spectral broadening. All viscosities are as reported in Table S1.

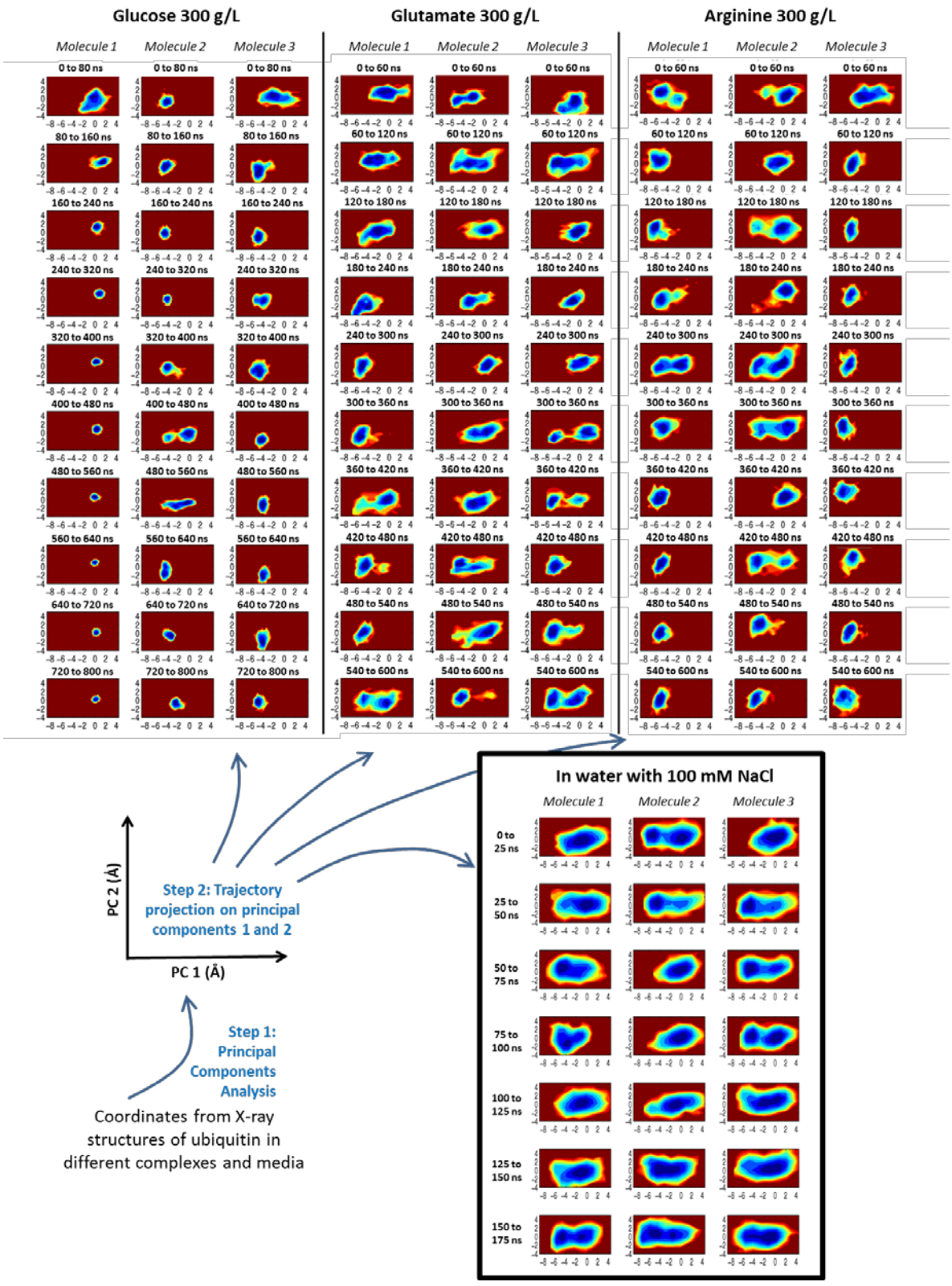

Figure S6. Projection of the conformations of each protein simulated in each system, on a plane of principal components that captures the conformational variability of ubiquitin when bound to different proteins as seen in X-ray structures. The x and y axes are principal components 1 and 2, respectively accounting for 42.6 and 12.5 % of the variability in the X-ray data, and both having units of Ångström. The diagram on the bottom left graphically explains how it was built (details can be found in Abriata et al *Phys. Biol.* 2013 or in Spiga et al *J. Phys. Chem. B* 2014). The inset shows how three ubiquitin molecules in a single box in water explore the conformational space much faster than in the presence of concentrated solutes (notice 25 ns time windows) (From Abriata and Dal Peraro *Scientific Reports* 2015, before protein aggregation).

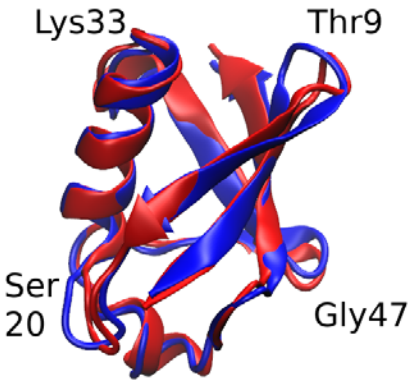

Figure S7. Structures of ubiquitin conformations at the deepest points of the basins observed in the simulations in water. The structure in red corresponds to the left-most basin, the one in blue to the right-most basin. Important amino acids from each loop are indicated. The flexible carboxi-terminal tail (residues 72-76) is not shown.

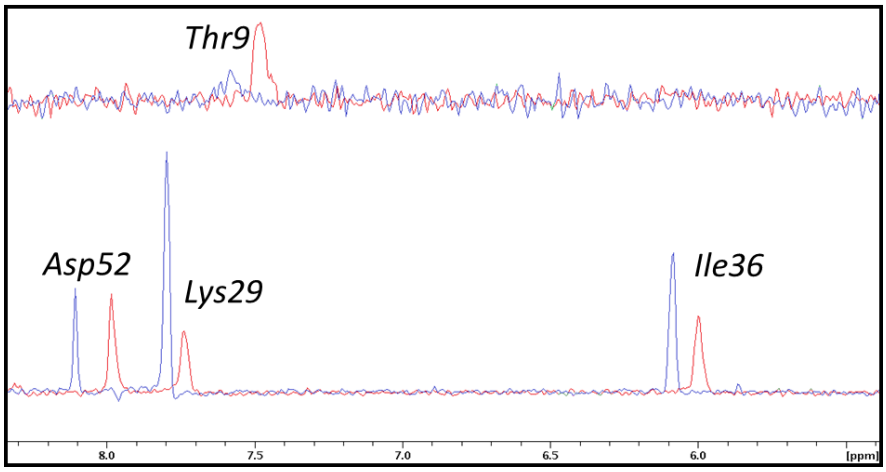

Figure S8. ^{1}H slices of ^{1}H,^{15}N HSQC spectra showing how the signal-to-noise ratio increases for Thr9's N,H cross peak in 300 g/L glucose (red) relative to buffer (blue) while the other resonances broaden (exemplified with Asp52, Lys29 and Ile36). The two spectra were acquired with the same protein concentration (200 μM) in the same buffer at controlled pH 7, with the same receiver gain, number of scans and resolution in the ^{15}N dimension, each with its own optimized tuning, matching and pulses.

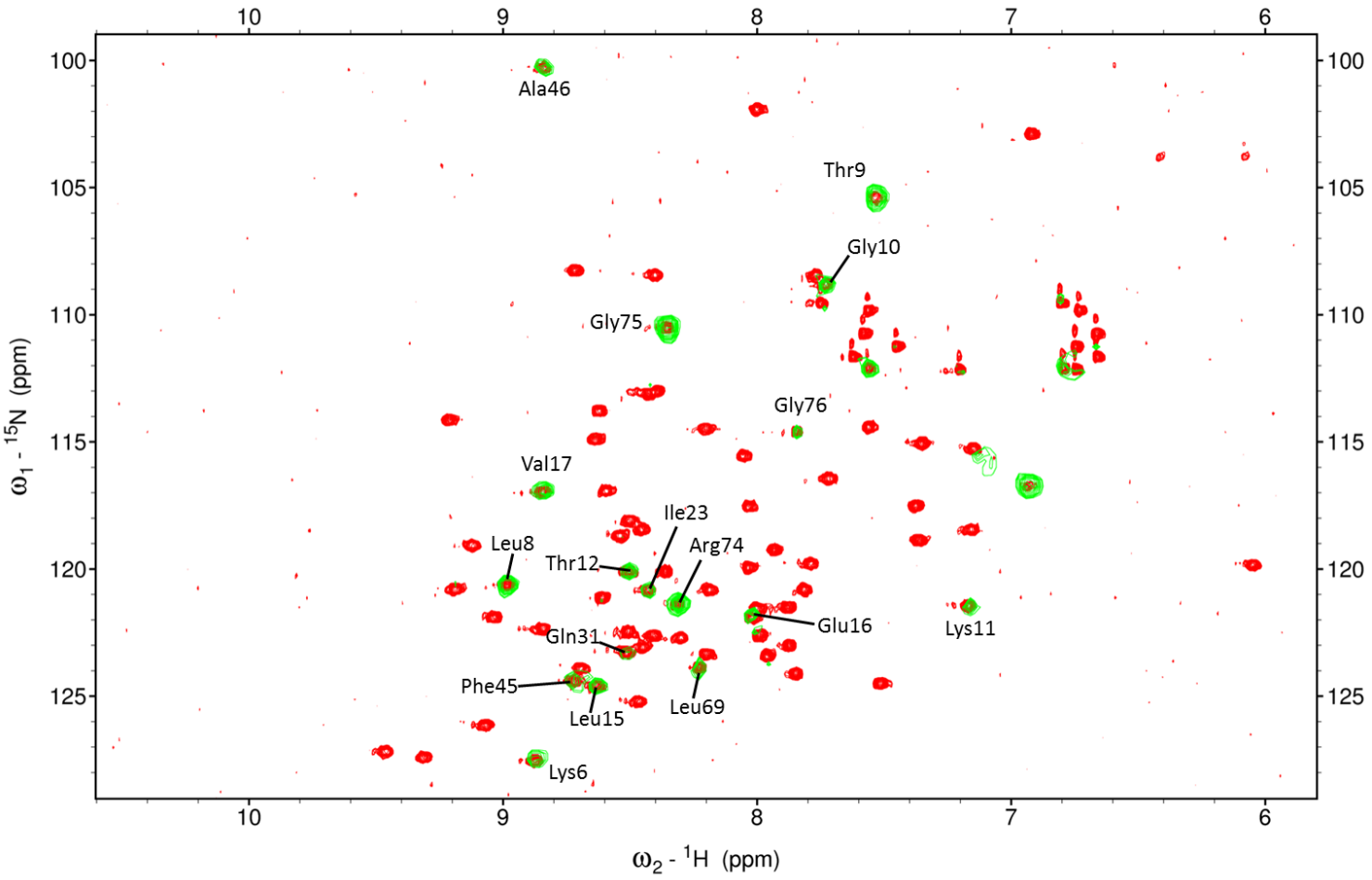

Figure S9. Plane at 4.7 ppm in the NOESY ^{1}H dimension of a 3D NOESY-HSQC spectrum of ubiquitin in 300 g/L glucose (green) overlaid on top of an HSQC spectrum of the same sample in in identical conditions. Crosspeaks from backbone resonances with enhanced NOESY intensities are labeled.

References for viscosity and static dielectric constant data in Table S1

Telis et al. Viscosity of Aqueous Carbohydrate Solutions at Different Temperatures and Concentrations, *International Journal of Food Properties* (2007) 10 (1):185-195

Mason et al. The Viscosities of Aqueous Solutions of Amino Acids at 25 and 35º, *Journal of the American Chemical Society* (1952) 74 (5):1287-1290

Saini et al. Volumetric analysis of L-arginine hydrochloride in aqueous and aqueous tetrahydrofuran solution at 303.15 K *Archives of Applied Science Research* (2012) 4 (5):2068-2076

Wyman et al. The Dielectric Constant of Solutions of Amino Acids and Peptides, *Journal of the American Chemical Society* (1933) 55:908-914

Ciu et al. The Effect of Molecular Crowding on the Stability of Human c-MYC Promoter Sequence I-Motif at Neutral pH, *Molecules* (2013) 18, 12751-12767; doi:10.3390/molecules181012751

Capuano et al. Electrostatic and Excluded Volume Effects on the Transport of Electrolytes in Poly(ethylene glycol)—Water "Mixed Solvents". *J. Phys. Chem. B* 2003, 107, 12363–12369

Czerwinski et al. The structural basis for the perturbed pKa of the catalytic base in 4-oxalocrotonate tautomerase: kinetic and structural effects of mutations of Phe-50. *Biochemistry* (2001) 40, 1984–1995

Mali et al. Dielectric relaxation of poly ethylene glycol-water mixtures using time domain technique. *Indian Journal of Pure & Applied Physics* (2007) 45, 476-481

Zaslavsky et al. Dielectric Properties of Water in the Coexisting Phases of Aqueous Polymeric Two-phase Systems. *J. Chem. Soc. Faraday Trans.* (1989) 85(9), 2857-2865

Stenger et al. Molecular Weight Dependence of the Depletion Attraction and its Effects on the Competitive Adsorption of Lung Surfactant. *Biochim. Biophys. Acta* (2008) 10:2032-2040

Cametti et al. Dielectric Relaxation Spectroscopy of Lysozyme Aqueous Solutions: Analysis of the δ-Dispersion and the Contribution of the Hydration Water. *J. Phys. Chem. B* (2011) 115, 7144-7153

Floros et al. Detailed study of the dielectric function of a lysozyme solution studied with molecular dynamics simulations. *Eur. Biophys. J.* (2015) DOI 10.1007/s00249-015-1052-7

Grant et al. The Dielectric Behavior of Aqueous Solutions of Bovine Serum Albumin from Radiowave to Microwave Frequencies. *J. Phys. Chem.* (1968) 72(13), 4373-4380