Supporting Information

Protein destabilization in ionic liquids: The role of preferential interactions towards denaturation

Angelo Miguel Figueiredo[a*], Joao Sardinha[a], Geoffrey R. Moore[b], Eurico J. Cabrita[a*]

[a] REQUIMTE, CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Monte de Caparica, Portugal

[b] Centre for Structural and Molecular Biochemistry, School of Chemistry, University of East Anglia, Norwich, NR4 7TJ, UK

E-mail: am.figueiredo@fct.unl.pt; ejc@fct.unl.pt

MATERIALS AND METHODS

Sample Preparation

^{15}N-labelled Im7 was overexpressed and purified as previously described [1]. *E. coli* JM109 competent cells were transformed with Im7 plasmid and growth in M9 media enriched with $^{15}NH_4Cl$ (1 g/L), 100x MEM vitamin solution (10ml/L), 0.01 M FeCl3, 100 mg/ml carbenicillin (1ml/L), 1% thiamine solution (0.2 ml/L), 1 M $MgSO_4$ (2ml/L), 100 mM $CaCl_2$ (2 ml/L) and glucose (4 g/L). Cells were sonicated and debris removed by centrifugation for further purification. Ammonium sulphate precipitation was performed before anion exchange chromatography followed by S75 gel-filtration chromatography. Samples were then dialyzed against purified water. Lyophilized samples for NMR experiments were prepared by re-suspending in 50 mM phosphate buffer, pH 7, 10% (v/v) ^2H_2O/90% H_2O into the different ILs to a final protein concentration of 0.5 mM.

ILs: 1-butyl-3-methylimidazolium chloride ([C_4mim][Cl]), 1-butyl-3-methylimmidazolium dicyanamide ([C_4mim][dca]), 1-ethyl-3-methylimidazolium chloride ([C_2mim][Cl]) and 1-ethyl-3-methylimidazolium dicyanamide ([C_2mim][dca]) employed in this work were purchased from Iolitec (Denzlingen, Germany). All ILs were at least 99% pure, and before solution preparation were dried for 24 h under vacuum at 60 °C. pH values are direct meter readings uncorrected for any isotope effects. Buffer solutions were used to ensure pH stabilisation during the course of the experiment.

Differential Scanning Calorimetry

Thermal transitions were characterized by differential scanning calorimetry (DSC) using a VP-DSC instrument (Microcal, Northampton, USA). Protein samples and reference solutions were properly degassed to avoid bubble formation during the temperature scan, and pressure in the measuring cell was kept at approximately 2 atm. Exhaustive cleaning of the cells was undertaken before each experiment. Thermograms were obtained with freshly prepared 0.09 mM Im7 dissolved in 50 mM phosphate buffer pH 7, with different types and concentrations of ionic liquids: [C_2mim][Cl] (1 M and 240 mM), [C_4mim][Cl] (1 M and 240 mM), [C_2mim][dca] (1 M and 240 mM) and [C_4mim][dca] (1 M and 240 mM). A scan rate of 90 °C/h, from 25 to 90 °C was used, and 20 baselines were acquired before each experiment. To guarantee thermodynamic equilibrium at the outset of each experiment all protein samples were pre-equilibrated overnight in the proper ionic liquid under study. T_m values were determined in duplicate for each sample.

The thermodynamic functions (ΔH, ΔS and ΔG) for protein unfolding were estimated from the heat capacity changes upon protein unfolding. Following, a two-state transition model, i.e., $\Delta H_{exp}(T_m) \approx \Delta H_{vH}(T_m)$ at the transition temperature (T_m) the population of native and unfolded states is equal [2]. This means that at T_m the equilibrium constant of the reaction $K_{eq} = F_N/F_U$ is equal to 1, and thus the Gibbs free energy difference between the native and unfolded states at T_m is equal to zero and $\Delta G(T_m) = \Delta H(T_m) - T_m \cdot \Delta S(T_m) = 0$. Thus the entropy change upon protein unfolding at the transition temperature, $\Delta S(T_m)$, is equal to:

$$\Delta S(T_m) = \frac{\Delta H(T_m)}{T_m}$$

A complete thermodynamic description of the system at any other temperature can then be calculated as follows:

$$\Delta G(T) = \Delta H(T_m) - T\,\frac{\Delta H(T_m)}{T_m} + \int_{T_m}^{T} \Delta C_p\, dT - \int_{T_m}^{T} \Delta C_p\, d\, lnT$$

where the heat capacity change upon protein unfolding, ΔC_p is a function of temperature itself. In practice, however, if no extrapolations over a wide temperature range are necessary, as a first approximation a temperature-independent ΔC_p is a reasonable assumption [2].

NMR SPECTROSCOPY

NMR experiments were performed at 283 K in a Bruker Avance II+ 600-MHz spectrometer equipped with 5-mm TCI cryoprobe. Proton chemical shifts were referenced against external DSS while nitrogen chemical shifts were referenced indirectly to DSS using the absolute frequency ratio. Data were processed using the NMRPipe/NMRDraw package [3] and analysed with CCPNMR [4]. Im7 assignments were taken from the BMRB databank under the accession code 7188. A 0.2 mM solution of ^{15}N-labeled Im7 in 50 mM phosphate buffer (10% $^{2}H_2O$, pH 7) was titrated individually with different ILs at the following concentrations: 0, 30, 60, 120, 240, 500 and 1000 mM, in order to maintain the protein concentration constant throughout the titration procedure. One ^{1}H-^{15}N HSQC spectrum was acquired for each IL concentration sample. The changes in ^{1}H and ^{15}N chemical shifts are referred to as chemical shift perturbations (CSP) and were calculated as:

$$CSP = \sqrt{\Delta\delta_{HN}^2 + \frac{\Delta\delta_N^2}{6^2}}$$

where, $\Delta\delta_{HN}$ is the proton chemical shift difference of Im7 in the presence of added ionic liquid minus the same resonance in the absence of added species, and $\Delta\delta_N$ is the same term for the ^{15}N chemical shift.

In order to decide whether a given residue belongs to the class of interacting or non-interacting residues, we have calculated a cut-off value based on the corrected standard deviation to zero, according to the method developed by Kalbitzer and co-workers [5]. For residues above the cut-off the CSPs values were fitted to a binding isotherm, according to:

$$CSP = \frac{CSP_{max}\,[IL]}{K_D + [IL]}$$

The binding isotherm contains the apparent dissociation constant, K_D, as well as the constant CSP_{max}, which represent the maximum increase in the CSP caused by the binding of ions.

N^1H/N^2H exchange

Before hydrogen exchange experiments, the spectrometer was first shimmed with a protein sample dissolved in 50 mM phosphate buffer, with the different ILs studied, pH 7, at 10 °C. The sample for the exchange experiment was prepared by dissolving lyophilised ^{15}N-labelled Im7 protein in 600 µl of ^2H_2O (> 99.9% isotopic purity)

containing 50 mM phosphate buffer (pH 7), in the corresponding IL ([C_4mim][dca] 240mM, [C_2mim][dca] 240mM, [C_4mim][Cl] 1M and [C_2mim][Cl] 1M) to a final protein concentration of 1 mM. All buffers were pre-equilibrated at 10 °C. ^{1}H-^{15}N HSQC spectra were acquired at 600 MHz, 10 °C at a ^{1}H sweep width of 9615 Hz (centre on water) and a ^{15}N sweep width of 2311 Hz (centre at 118 ppm). The first spectrum finished acquiring within 3 min of dissolution using two scans and 512 complex points in t_2 and 64 complex points in t_1 and a recycle time of 0.8 s. Consecutive spectra were collected with the following number of scans for each t_1 increment: 4 spectra with 2 scans, 7 with 4 scans, 16 with 8 scans, 7 with 16 scans and 4 with 32 scans. In this way, acceptable signal/noise ratios were maintained as peak intensities decreased with amide hydrogen exchange for deuterium. In total, 38 spectra were acquired over a 14-hour period. Processing was performed with NMRPipe [3]; cross-peak volumes were extracted from CCPNMR software [4] and normalized over the number of scans of each ^{1}H-^{15}N HSQC spectrum. Rates of exchange, k_{ex}, were obtained from plots of peak intensity against exchange time fitted to a single exponential function. The exchange time was defined as taken to be the time from dissolution of the sample in $^{2}H_2O$ to the midpoint of the time taken to acquire each spectrum. Free energies of exchange (ΔG_{ex}) values were calculated as described:

According to the literature [6] the exchange reaction is,

$$(NH)_{closed} \underset{kop}{\overset{kcl}{\rightleftarrows}} (NH)_{open\ states} \xrightarrow{kint} N - D$$

in which an amide group undergoes structural opening (k_{op}) and closing (k_{cl}) to an exchanged form, $(N\text{-}H)_{open}$. Exchange occurs in the solvent-exposed open state with an intrinsic rate constant for exchange, k_{int}. According to Gorski *et al* [7] deuterium exchange process on Im7 (in buffered solution pH 7 at 10 °C – the same conditions as used on this study) occurs by an EX2 mechanism [8,9], in which $k_{cl} >> k_{int}$. In other words, the reclosing is faster than the exchange ($k_{cl} >> k_{int}$), the structural opening reaction appears as a pre-equilibrium step prior to the rate-limiting chemical exchange, and the observed rate constant (k_{ex}) is $k_{ex} = K_{op}\ k_{int}$ where K_{op} is the equilibrium constant for structural opening ($K_{op} = k_{op} / k_{cl}$). With the values from the exchange rate from the open state available (k_{int}), stability is measured as the free energy of exchange or opening ΔG_{HX};

$$\Delta G_{HX} = -RT \ln\left(k_{ex} / k_{int}\right)$$

where R is the gas constant, T is temperature in Kelvin, k_{ex} is the hydrogen exchange rate obtained from the exponential decay of peak intensity with time and k_{int} is the intrinsic exchange rate constant. Values for k_{int} were determined by using the web program SPHERE [10] which is based on published rates, with default activation energies: E_{acid} = 15 kcal/mol, E_{base} = 2.6 kcal/mol. The higher the value of ΔG_{HX} the more protected is the residue for exchange.

Relaxation Experiments

A freshly prepared 1 mM ^{15}N-labeled Im7 sample dissolved in 50 mM phosphate buffer, at pH 7, 10% D_2O was prepared with ([C_4mim][dca] 240 mM, [C_2mim][dca] 240 mM, [C_4mim][Cl] 1 M and [C_2mim][Cl] 1 M) and without IL, in order to collect ^{15}N-R_1, ^{15}N-

R_2 and {^{1}H}-^{15}N NOE data using the procedures described by Kay *et al*. [11] and Farrow *et al*. [12]. Spectra of Im7 were recorded at 283 K in an Bruker Avance II+ 600 MHz, as matrices of 1024 x 128 complex data points with spectral widths of 7945 Hz (^{1}H) and 1845 Hz (^{15}N) using 8 scans per t_1 increment. ^{15}N-R_1 data were acquired with relaxation delays of 5, 10 (2x, duplicated), 20, 30 (2x), 50, 80, 100, 200 (2x), 500 (2x), 750, 850, 900, 950, 1000, 1500 and 2000 msec. ^{15}N-R_2 experiments were measured with relaxation delays of 16 (2x, duplicated), 32 (2x), 48, 64 (2x), 80, 96, 112, 128 (2x), 144, 160 (2x, duplicated), 192, 210, 242, 322, 404 msec. Recycle delays of 4 s were used for the measurement of R_1 and R_2 data and acquired as single scan interleaving manner in order to minimize the effects of sample heating. For steady-state {^{1}H}-^{15}N NOE determination, two spectra were acquired, with and without ^{1}H saturation. Proton saturation was achieved by a train of 120° pulses separated by 5 msec during the 5 sec relaxation delay. Triplicate sets of interleaved saturated/unsaturated experiments were measured for Im7 in the different solution media to determine uncertainties of the NOE values, taken to be the standard deviation of the average NOE determined from the three repeat experiments. Resonances were picked and fitted simultaneously on all spectra using the NlinLS procedure in NMRPipe [3]. Uncertainties in the relaxation times were taken to be the standard errors of the fitted parameters. Heteronuclear NOE ratio was determined directly after fitting of the resonances and extraction of the intensities in each spectrum (with and without saturation). Residues were excluded from analysis where resonance overlap was too severe or peak intensity too weak to reliably determine the relaxation parameters. Order parameters and rotational correlation times were obtained from a Lipari-Szabo model-free analysis conducted under Tensor2 [13] after exclusion of residues with NOE

values lower than 0.65, and residues that exhibit conformational exchange according to the criteria proposed by Tjandra and coworkers [14];

$$\left(\frac{\langle T_2\rangle - T_{2,n}}{\langle T_2\rangle} - \frac{\langle T_1\rangle - T_{1,n}}{\langle T_1\rangle}\right) > 1.5\,SD$$

where $T_{2,n}$ is the T_2 value ($1/R_2$) of residue n, and $\langle T_1\rangle$ and $\langle T_2\rangle$ are the averages over the residues that show NOE > 0.65. SD is the standard deviation of the calculation.

Values for the ^{15}N chemical shift anisotropy and N-H bond length, were set at -172 ppm and 1.02 Ang., respectively. Errors were obtained from 500 Monte Carlo simulations.

Hydrodynamic Radius

Hydrodynamic radii were measured by the approach described by Wilkins *et al.* [15]. Typically, in each experiment 32 spectra of 32K data points were collected at 283 K, with values for the duration of the magnetic field pulsed gradients (δ) of 2.0 ms, diffusion times (Δ) of 200 ms, and an eddy current delay set to 5 ms. The gradient recovery time was 200 μs. A sine shaped pulsed gradient (g) was incremented from 5 to 95 % of the maximum gradient strength (56 G/cm) in a linear ramp. To determine the diffusion coefficients, the spectra were first processed in the F2 dimension by standard Fourier transform and baseline correction. The diffusion coefficients are calculated by exponential fitting of the data belonging to individual columns of the 2D matrix using the Origin 9.0 data software program (OriginLab, Northampton, MA). Each experiment was

acquired in triplicate for error analysis. In each sample, the diffusion coefficient determined for the dioxane peak was used as the reference $R_h^{ref} = 2.12$ Å [15].

Molecular Dynamics Simulations

Molecular mechanics (MM) calculations and molecular dynamics (MD) simulations were performed with AMBER12 [16], using the ff12SB [17] and LHW-FF force fields to parameterized both protein and ILs, respectively. The ILs used were the [C_4mim][Cl], [C_4mim][dca], [C_2mim][Cl] and [C_2mim][dca]. The starting configurations were generated with the program PACKMOL filling a cube with the appropriate number water molecules and IL in order to achieve a 240 mM or 1 M concentration of IL. The MD simulations were performed using periodic boundary conditions following a five-step protocol: The first step consisted in a 20000 cycles of minimization to remove any possible unfavorable contacts between solvent and protein. The first 3000 cycles of the minimization were performed with the steepest descent method, followed by the conjugate gradient method. In this step, the solute is restrained in the Cartesian space using a harmonic potential restraint (weight 500 kcal $mol^{-1}Å^{-2}$). Subsequently, a 10000 cycles of minimization (3000 steps of steepest descent and 7000 steps of conjugate gradient method) without restraints was performed. The systems were then heated up to 283 K for 50 ps using a NVT ensemble and a weak positional restraint (10 $mol^{-1}Å^{-2}$) on the solute, to avoid wild structural fluctuations, using the Langevin thermostat with a collision frequency of 1 ps^{-1}. The positional restraints were removed and a molecular dynamics run in a NPT ensemble at 283 K for 500 ps was performed for equilibration at 1 atm with isotropic scaling and a relaxation time of 2 ps. Finally, NPT data production

runs were carried out for 50 ns and the snapshots were saved to a trajectory file every 0.2 ps. All bonds involving hydrogen atoms were constrained with the SHAKE algorithm [18] allowing the use of a 2 fs time step. The Particle Mesh Ewald method [19] was used to treat the long-range electrostatic interactions and the non-bonded van der Waals interactions were truncated with a 9 Å cut-off. The structural collected data were analysed with the PTRAJ module as implemented in the AMBER package. The MD trajectories were also clustered by RMSD similarity using the average-linkage clustering algorithm [20].

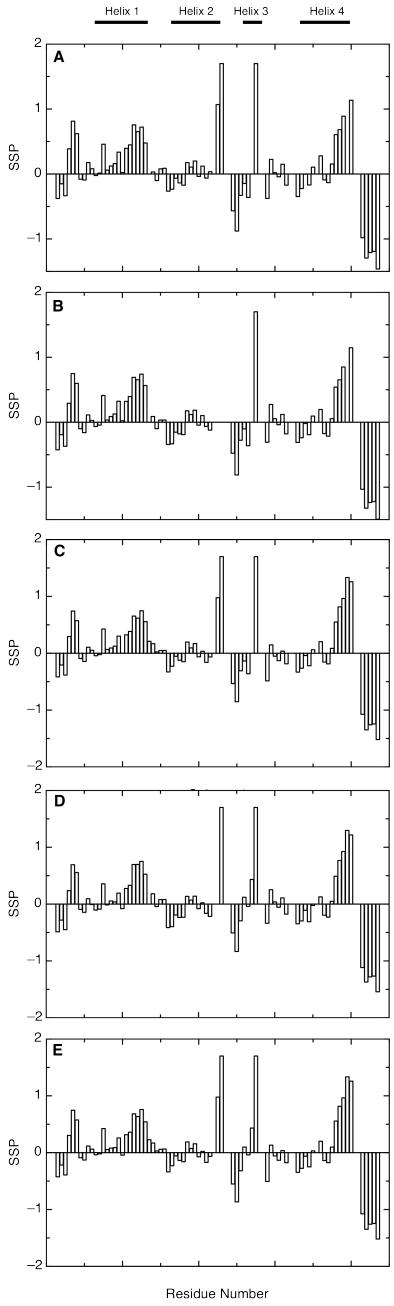

Figure S1 – Secondary structure propensity (SSP) for wild type Im7 – On top secondary structure according to the pdb file 1AYI. SSP analysis [21] of H^N and ^{15}N at 283 K, pH 7.0: A - 50 mM phosphate buffer; B - $[C_4mim][Cl]$ at 1 M; C - $[C_4mim][dca]$ at 240 mM, D - $[C_2mim][Cl]$ at 1 M; and E - $[C_2mim][dca]$ at 240 mM.

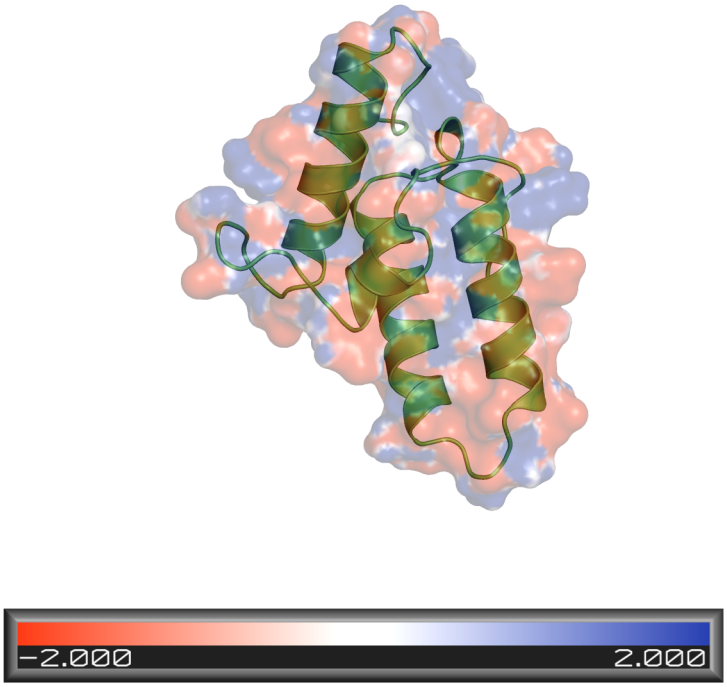

Figure S2 – Electrostatic charge distribution at the solvent accessible surface of Im7 (1AYI.pdb). Surface representation coloured by charge from red (-2 k_{cal}.mol^{-1}e^{-1}) to blue (+2 k_{cal}.mol^{-1}e^{-1}) using a dielectric constant of 80 and without taking into account ionic strength. The electrostatic potential was generated with AMBER 12 [16] package and visualised under PYMOL [22].

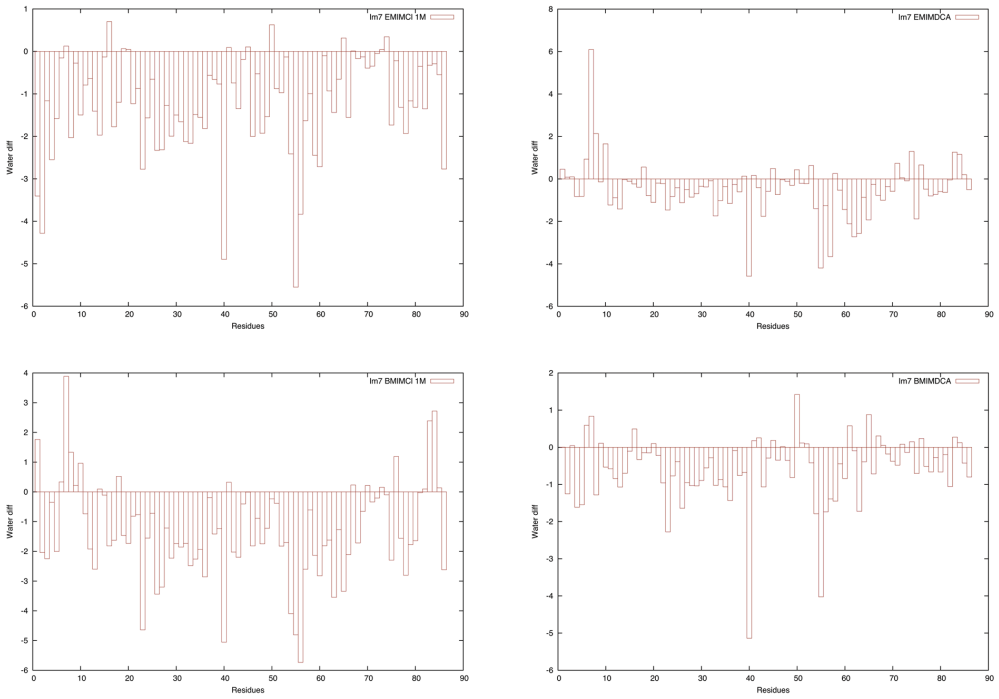

Figure S3 – Water solvation shell difference for the Im7 protein system with and without IL. A - [C_2mim][Cl] 1 M; B - [C_4mim][Cl] 1 M; C - [C_2mim][dca] 240 mM; D - [C_4mim][dca] 240 mM, determined by MD simulations using AMBER 12 package [16]. Secondary structure elements are illustrated on the top of the figure according to the 1AYI.pdb file.

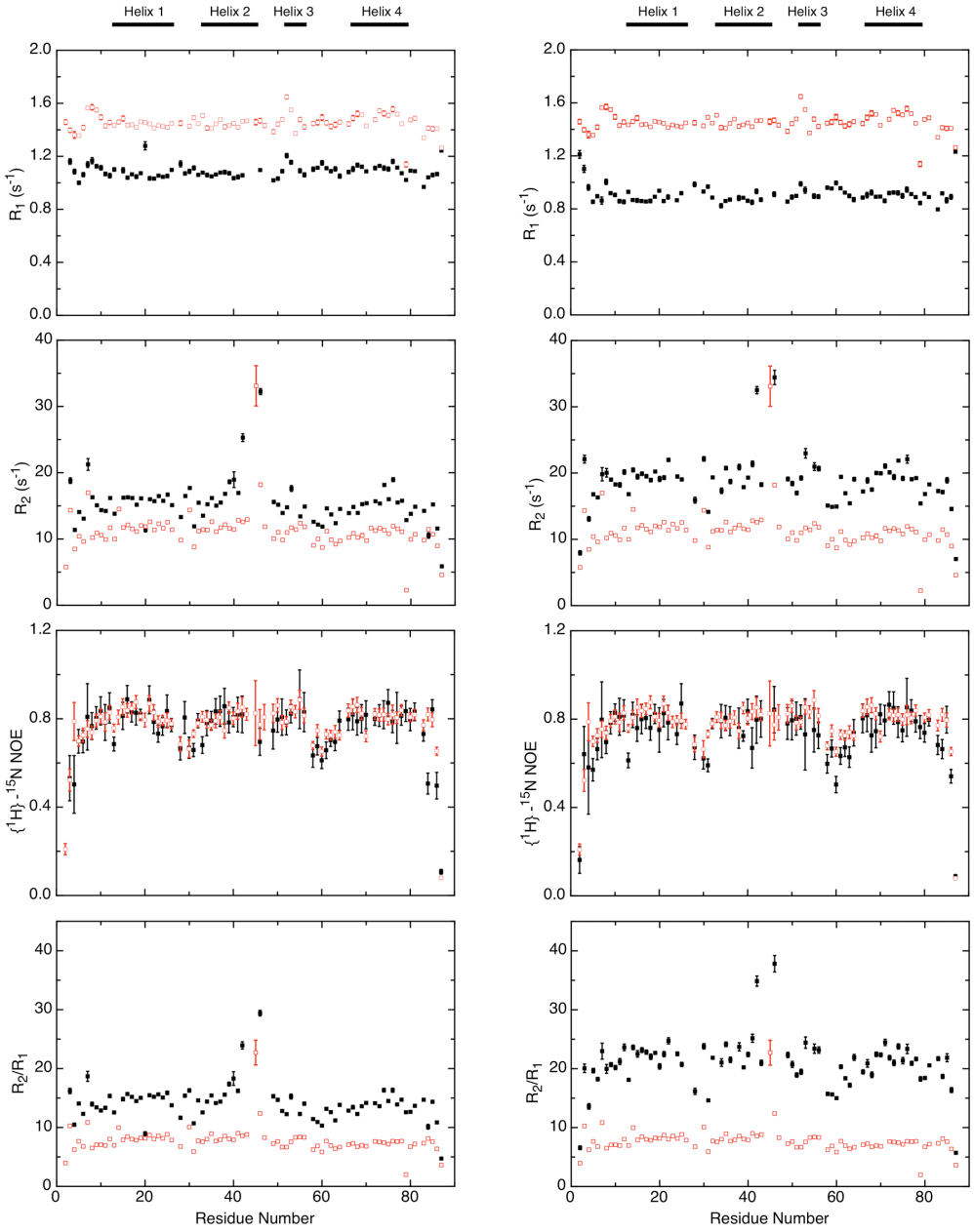

Figure S4 – ^{15}N backbone relaxation data R_1, R_2, heteronuclear {^{1}H}-^{15}N NOE values and R_2/R_1 ratios measured for Im7 in aqueous [C_2mim][Cl] (left) and [C_4mim][Cl] (right) at 1 M ionic liquid concentration (black boxes) and in aqueous buffered solution (red boxes). Black bars at the top of the figure indicate the location of the helices in the native protein. All data was acquired at proton frequency of 600 MHz, pH 7.0, 283 K.

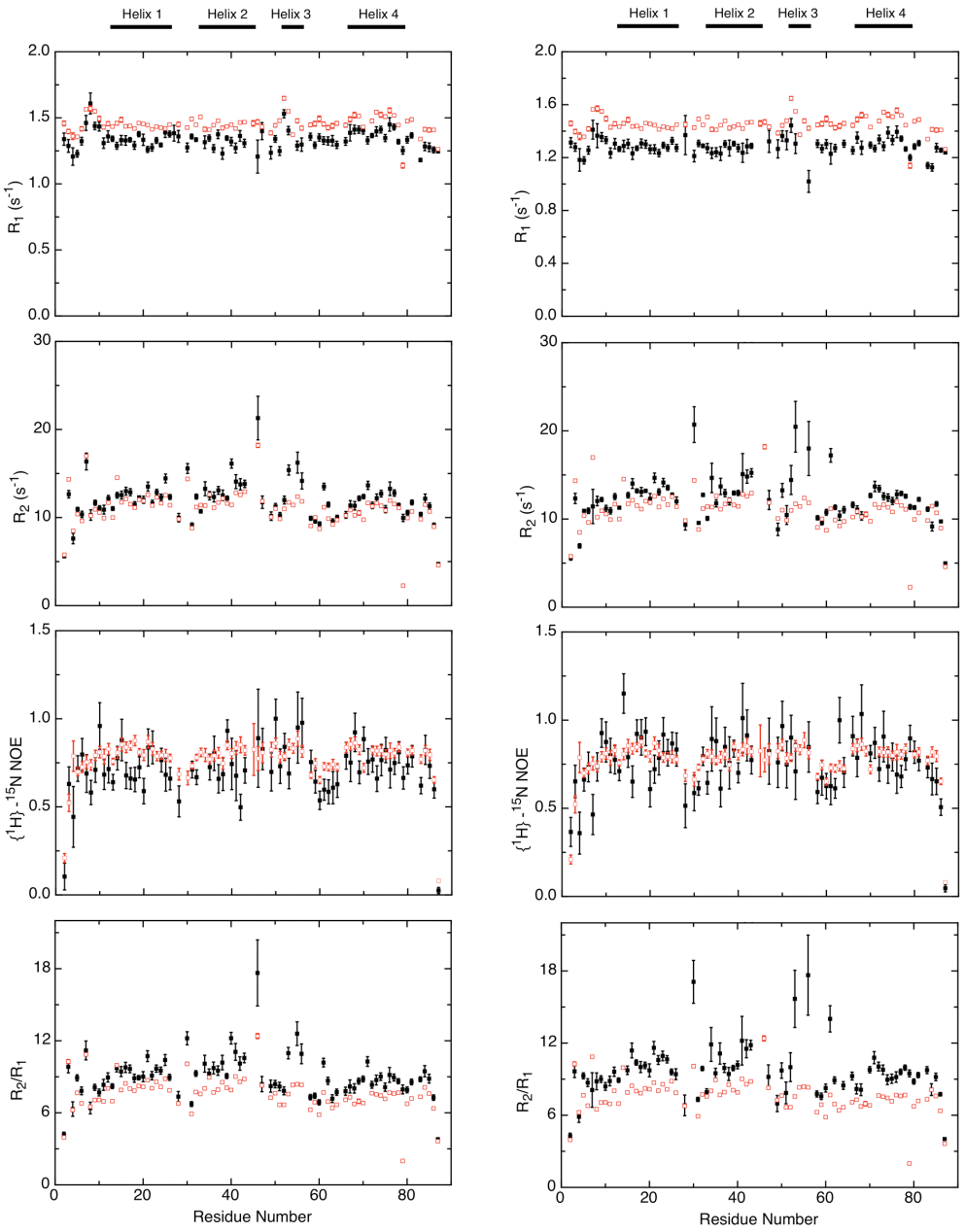

Figure S5 – ^{15}N backbone relaxation data R_1, R_2, heteronuclear $\{^1H\}$-^{15}N NOE values and R_2/R_1 ratios measured for Im7 in aqueous [C_2mim][dca] (left) and [C_4mim][dca] (right) at 240 mM ionic liquid concentration (black boxes) and in aqueous buffered solution (red boxes). Black bars at the top of the figure indicate the location of the helices in the native protein. All data was acquired at proton frequency of 600 MHz, pH 7.0, 283 K.

Additional References

1. S. A. Gorski, A. P. Capaldi, C. Kleanthous, and S. E. Radford, *J Mol Biol*, 2001, **312**, 849–863.

2. G. I. Makhatadze, *Current Protocols in Protein Science*, 2001, **Chapter 7**, Unit 7.9.

3. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *J Biomol NMR*, 1995, **6**, 277–293.

4. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, *Proteins*, 2005, **59**, 687–96.

5. F. H. Schumann, H. Riepl, T. Maurer, W. Gronwald, K.-P. Neidig, and H. R. Kalbitzer, *J Biomol NMR*, 2007, **39**, 275–89.

6. M. M. G. Krishna, L. Hoang, Y. Lin, and S. W. Englander, *Methods*, 2004, **34**, 51–64.

7. S. A. Gorski, C. S. C. S. Le Duff, A. P. Capaldi, A. P. Kalverda, G. S. Beddard, G. R. Moore, and S. E. Radford, *J Mol Biol*, 2004, **337**, 183–193.

8. Y. Bai, J. J. Englander, L. Mayne, J. S. Milne, and S. W. Englander, *Methods Enzymol*, 1995, **259**, 344–356.

9. Y. Bai, J. S. Milne, L. Mayne, and S. W. Englander, *Proteins*, 1993, **17**, 75–86.

10. Hydrogen Exchange Prediction http://www.fccc.edu/research/labs/roder/sphere/sphere.html.

11. L. E. Kay, L. K. Nicholson, F. Delaglio, A. Bax, and D. A. Torchia, *J Magn Reson*, 1992, **97**, 359–375.

12. N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay, *Biochemistry*, 1994, **33**, 5984–6003.

13. P. Dosset, J. C. Hus, M. Blackledge, and D. Marion, *J Biomol NMR*, 2000, **16**, 23–28.

14. N. Tjandra, S. E. Feller, R. W. Pastor, and A. Bax, *J Am Chem Soc*, 1995, **117**, 12562–12566.

15. D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith, *Biochemistry*, 1999, **38**, 16424–16431.

16. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, *AMBER 12*, University of California, San Francisco, 2012.

17. J. Wang, P. Cieplak, and P. A. Kollman, *J Comput Chem*, 2000, **21**, 1049–1074.

18. J.-P. Ryckaert, G. Ciccotti, and H. J. . Berendsen, *J Comput Phys*, 1977, **23**, 327–341.

19. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J Chem Phys*, 1995, **103**, 8577–93.

20. J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham, *J Chem Theory Comput*, 2007, **3**, 2312–2334.

21. J. A. Marsh, V. K. Singh, Z. Jia, and J. D. Forman-Kay, *Protein Sci*, 2006, **15**, 2795–2804.

22. W. L. Delano, 2002.