Protein loop compaction and the origin of the effect of arginine and glutamic acid mixtures on solubility, stability and transient oligomerization of proteins

Jascha Blobel[1], Ulrika Brath[2], Pau Bernadó[1], Carl Diehl[2], Lidia Ballester[1], Alejandra Sornosa[1], Mikael Akke[2] & Miquel Pons[1,3]

1. Laboratory of Biomolecular NMR, Institute for Research in Biomedicine. Parc Científic de Barcelona, Baldiri Reixac, 10, 08028 Barcelona, Spain.

2. Division of Biophysical Chemistry, Center for Molecular Protein Science, Lund University, Lund, Sweden.

3. Departament de Química Orgànica, Universitat de Barcelona (UB), Martí i Franquès, 1-11, 08028 Barcelona, Spain.

*** Corresponding author:** Miquel Pons Laboratory of Biomolecular NMR, Institute for Research in Biomedicine. Parc Científic de Barcelona, Baldiri Reixac, 10, 08028 Barcelona, Spain. Tel +34934034683. Fax +34934039976. e_mail: mpons@ub.edu

Abstract

Addition of a 50 mM mixture of L-arginine and L-glutamic acid (RE) is extensively used to improve protein solubility and stability although the origin of the effect is not well understood. We present Small Angle X-ray Scattering (SAXS) and Nuclear Magnetic Resonance (NMR) results showing that RE induces protein

compaction by collapsing of flexible loops on the protein core. This is suggested to be a general mechanism preventing aggregation and improving stability to proteases and to originate from the polyelectrolyte nature of RE.

Molecular polyelectrolyte mixtures are expected to display long range correlation effects according to dressed interaction site theory. We hypothesize that perturbation of the RE solution by dissolved proteins is proportional to the volume occupied by the protein. As a consequence, loop collapse, minimizing the effective protein volume, is favored in the presence of RE.

Keywords: Protein compaction, crowding, molecular polyelectrolytes, loop dynamics, protein stability, co-solutes

Abbreviations:

FKBP12	FK506 binding protein
lmwPTP	Bovine low molecular weight protein tyrosine phosphatase
MBP	Maltose Binding Protein
PIC	Polyelectrolyte Induced Compaction
RE	An equimolar mixture of L-arginine and L-glutamic acid
R_g	Radius of gyration
SAXS	Small Angle X-ray scattering

Introduction

Weak protein-protein interactions can be modulated by the presence of co-solutes. Non-specific interactions leading to aggregation of partially unfolded proteins is a physiological relevant process that is minimized by the action of chaperones and possibly by increasing the concentration of small molecular weight co-solutes under stress conditions. Protein aggregation is a common cause of failure of structural biology projects either using NMR or X-ray. The addition of an equimolar mixture of arginine and glutamic acid (RE) was reported by Golovanov et al. (2004) to prevent aggregation and to improve the solubility of a large number of proteins while preserving biologically relevant macromolecular interactions. The same publication reported an increased resistance to degradation by proteases in the presence of 50 mM RE. A large scale study by Vedadi et al (2006) of the influence of added co-solutes on structural proteomics projects showed that RE and n-dodecyl-β-D-maltoside were the two most universal co-solutes in stabilizing a wide variety of different proteins by more than 4°C. Our group have studied the structural effects of RE mixture with bovine low molecular weight protein tyrosine phosphatase (lmwPTP), a protein that forms dimers and higher oligomers in solution. Using ^{129}Xe-NMR we showed that RE strongly suppresses non-specific interactions of Xe atoms with lmwPTP (Blobel et al 2007). In the same work it was shown that RE increases the population of lmwPTP dimers. Thus, while RE prevents non-specific aggregation of a very diverse set of proteins, it preserves and even increases the affinity of specific

interactions. The origin of these apparently conflicting effects of RE is not well understood.

Here we have studied the effect of RE on three different proteins by Small Angle X-ray Scattering (SAXS) and NMR. Bovine lmwPTP, had been previously studied in our laboratories by a number of different techniques (Åkerud et al. 2002, Bernadó et al. 2003; Blobel et al. 2007; Blobel et al. 2009). Dimerization is a conserved feature of lmwPTP from bacteria to mammals (Blobel et al. 2009b). Dimerization involves loops which are highly mobile in the monomer (Åkerud et al. 2002). Applying the Multivariate Curve Resolution (MCR) approach to SAXS data, as recently described (Blobel et al. 2009), we show that RE causes a selective reduction of the radius of gyration of lmwPTP dimers determined by SAXS. NMR data on the same system map RE induced perturbations to the dimerization loop and the N- and C- terminal extremes of lmwPTP. FK-506 binding protein (FKBP12) is a monomeric, highly stable protein that has a loop, comprising residues Ala81-Thr96, with restricted conformational dynamics. This loop is part of the binding surface for immunosuppressant and peptide substrates (Brath et al. 2006, Brath & Akke 2009). While SAXS data do not detect RE induced changes in FKBP12, NMR chemical shift and relaxation time measurements show that the effect of RE is concentrated in the loop regions of FKBP12. Chymotrypsinogen A contains large loop regions that are cleaved by proteases to generate different forms of chymotrypsin. SAXS results show a clear compaction of chymotrypsinogen and lmwPTP dimer caused by the addition of RE. NMR results identify in lmwPTP and FKBP12 that RE induces perturbations

mainly localized in the loops. RE also induces an increase in the denaturation temperature of chymotrypsinogen but not of maltose binding protein which does not have exposed loops.

Together, these results are consistent with a model explaining the RE-induced promiscuous solubility and stability enhancement by the compaction of loops and other poorly structured regions onto the surface of the folded protein core. This compaction effect would efficiently reduce aggregation involving unstructured regions and would also explain the RE-enhanced stability to proteases, which preferentially hydrolyze locally unfolded regions. We propose that RE, and other polyelectrolytes, induce protein compaction to minimize the perturbation by the protein of long-range electrostatic interactions between molecular polyelectrolyte co-solutes.

MATERIALS AND METHODS

RE buffer preparation

A stock solution of 1 M RE was prepared by dissolving a mixture of pure L-arginine (Sigma) and L-glutamic acid (Sigma) in water. The solubility of the mixture is higher than that of L-glutamic acid alone. The stock solution (pH 6.5) was used to prepare samples in different buffers.

Sample preparation.

Chymotrypsinogen A (M_r = 20.4 kDa) was bought from Sigma. Recombinant human FKBP12 (M_r = 11.8 kDa) and lmwPTP (M_r = 18.1 kDa) were expressed and purified as explained elsewhere (Standaert et al. 1990; Wo et al. 1992). From the same batches of chymotrypsinogen A and FKBP12, pairs of samples, one with the maximum RE concentration and the other without RE, were prepared in 25 mM phosphate (pH = 6.5) and 1mM TCEP•HCl. Samples with intermediate RE concentrations were obtained by mixing of the two standard samples. The lmwPTP buffer consisted of 200 mM potassium phosphate, 3 mM sodium azide and 10 mM TCEP•HCl at a pH = 6.0. In all cases, total protein concentrations were determined by UV absorption.

Small Angle X-ray scattering.

Synchrotron radiation X-ray scattering data was collected following standard procedures on the X33 beamline at the European Molecular Biology Laboratory (EMBL) on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) (Roessle et al. 2007). Scattering curves were recorded on a MAR345 image plate detector. Chymotrypsinogen A was measured at a concentration of 0.36 mM (7.3 mg/ml) and FKBP12 at 0.76 mM (9 mg/ml) at a temperature of 20°C. Both proteins were measured in the presence of 0, 50, 100 and 200 mM RE. lmwPTP was measured in zero and 50 mM RE at ten different potrein concentrations (1.00, 0.80, 0.60, 0.55, 0.48, 0.41, 0.34, 0.25, 0.17 and 0.056 mM; 18 to 1.01 mg/ml) at 37°C. lmwPTP samples were prepared in both non-RE and RE buffer by successive dilution of a 1 mM lmwPTP sample with the corresponding buffer. Measurements up to 0.60 mM lmwPTP in non-RE buffer

were already reported in an earlier publication of our group (Blobel et al. 2009). Scattering curves of the buffer were collected before and after each acquisition of a protein sample to avoid systematic error. The scattering curves covered a momentum transfer range of $0.0956 < s < 5.0455$ nm^{-1}. The scattering due to buffer was subtracted by averaging the buffer measurements enclosing the actual protein measurement. All data manipulations were performed with the program PRIMUS (Konarev et al. 2003).

The forward scattering $I(0)$ and the radius of gyration R_g were evaluated with the Guinier approximation assuming that, at very small angles ($s < 1.3/R_g$), the intensity is represented as $I(s)=I(0)\exp(-(sR_g)^2/3)$ (Guinier 1939). The actual concentration of the lowest lmwPTP concentration sample (0.056 mM) was determined from its $I(0)$ value by extrapolation of the $I(0)$ obtained at 0.17 mM lmwPTP.

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) analysis of the SAXS curves.

The SAXS curves of 0.17, 0.25, 0.34 and 0.41 mM lmwPTP ranging from s = 0.185 – 1.666 nm^{-1}, consisting of 659 points, were jointly analyzed by MCR-ALS as described elsewhere (Blobel et al. 2009). The maximum s-value was chosen to ensure that only positive intensities for all scattering curves were present up to 1.666 nm^{-1}. The large divergences in the intensities up to 0.0185 Å^{-1} of the lowest used concentration curve (0.17 mM) restricted the analysis to lower angle measures. The error due to the concentration matrix, resulting from its deviation

158 from the dissociation constant, was scaled in respect to the previous publication
159 to make up 5% of the total error, giving a scaling factor ε = 500.
160 *NMR measurements.*
161 ^{15}N HSQC spectra at different RE concentrations were measured using 500 MHz
162 Varian or Bruker spectrometers. HSQC spectra of 1.0 mM FKBP12 were
163 recorded at 20ºC in 25 mM phosphate and 1mM TCEP•HCl at a pH = 6.5 in the
164 presence of 0, 50, 100, 200, 300 or 400 mM RE. HSQC spectra of 0.1 mM
165 lmwPTP were measured at 37ºC in 100 mM potassium phosphate, 1 mM sodium
166 azide and 1 mM TCEP•HCl at a pH = 6.0 with 0, 50, 100, 200, 400 or 600 mM
167 RE. ^{15}N relaxation measurements of 0.94 and 1.4 mM FKBP12 were
168 accumulated on a 600 MHz Varian spectrometer. Spectra were processed with
169 NMRPipe (Delaglio et al. 1995). Further data processing was done using in
170 house written programs running in Matlab©.
171 $\langle R_2/R_1 \rangle$ in the presence of 50 mM RE were scaled using the core residues not
172 affected by chemical exchange to correct for the different viscosity of the two
173 solvents. The scaling factor (1.077) would correspond to a ratio of viscosities of
174 1.038, in good agreement with the measured value of 1.041 (Blobel et al. 2007).
175 *Denaturation measurements.*
176 Denaturation experiments were performed in 96-well plates on a FluoDia T70
177 fluorimeter by following the increase of fluorescence of SYPRO Orange (Sigma
178 Aldrich) in the presence of 10 μM protein in 25 mM phosphate buffer containing
179 150 mM NaCl at pH 6.5 for chymotrypsinogen A and pH 7.5 for maltose binding

protein. The melting temperature was calculated by fitting a two state model to the unfolding curve using an in-house program written in Matlab©.

RESULTS

lmwPTP dimer compaction and stabilization

The dimerization and further oligomerization of lmwPTP was studied in the presence of 50 mM RE and in RE-free buffer by measuring SAXS curves at ten different lmwPTP concentrations. Figure 1 shows the apparent radius of gyration (R_g) and maximum dimensions (D_{max}) of the particles present in solution as a function of the total protein concentration.

Insert Figure 1 here

The measured values are intermediate between those expected on the basis of the pure monomer and pure dimer X-ray structures (Zhang et al. 1994, Tabernero et al. 1999), consistent with an oligomerization equilibrium described with earlier NMR measurements (Åkerud et al. 2002, Bernado et al. 2003, Blobel et al. 2007). We have previously demonstrated that the presence of RE increased the stability and therefore the population of lmwPTP dimer in the equilibrium mixture (Blobel et al. 2007). In apparent contrast with these observations, samples in RE buffer consistently show smaller apparent R_g

values. The difference is greater at intermediate concentrations where the proportion of dimer is the highest.

The apparent D_{max} values increases with the total protein concentration up to a plateau value at 0.17 and 0.34 mM in the presence and in the absence of RE, respectively. In RE-free buffer the plateau value of D_{max} value agrees with the dimensions of the dimer observed in crystals. In contrast, D_{max} is approximately 5 Å smaller in the presence than in the absence of RE. At higher concentrations D_{max} become larger than expected for a dimer indicating the presence of higher oligomers.

Pure SAXS curves were extracted for eachof the two main components in the mixture by simultaneous analysis of all curves in the concentration range from 0.17 to 0.41 mM using the MCR-ALS algorithm (Blobel et al. 2009). At higher concentrations at least one more component contributes significantly to the experimental curves preventing further analysis. Figure 2 shows the extracted pure curves for the monomer and dimer of lmwPTP in the presence of RE when fitted to the crystallographic models of both species with CRYSOL (Svergun et al. 1995).

Insert Figure 2 here

The extracted curve for the monomer is in good agreement with the one expected from the crystal structure (χ^2 = 2.4; R_g crystal = 16.45 Å, R_g exp = 15.91 Å) but the curve for the dimer shows a lower agreement (χ^2 = 3.4). The R_g value obtained from the extracted curve is 21.1 Å which is smaller than the one

predicted from the crystal structure (R_g crystal = 23.3 Å). As a reference, the R_g values of the individual species extracted from the data measured in the absence of RE were 16.4 Å and 23.9 Å for the monomer and dimer respectively (Blobel et al. 2009). Thus, both R_g values extracted from the pure curves and the evolution of D_{max} values suggest that RE is causing a compaction of the lmwPTP dimer. The compaction is also manifested in the poor agreement of the crystallographic structure of the dimer to the MCR-ALS derived SAXS curve (χ^2 = 3.4, see above) that indicates that RE perturbs its structure. The MCR-ALS method provides estimates of the dimer dissociation constant of 0.79 ±0 .06 mM in the presence of RE and 1.62 ± 0.12 mM in RE free buffer. These results confirm the stabilization of lmwPTP dimers by RE observed with different techniques (Blobel et al. 2007). The MCR-ALS analysis of the concentration dependent SAXS dataset has the capacity to deconvolute the thermodynamic changes and the compaction effects exerted by RE.

lmwPTP NMR studies map the sites perturbed by RE

In order to further characterize the structural effects caused by the addition of RE at a residue level, we measured changes in the NMR chemical shifts of 100 μM lmwPTP as a function of RE concentration in the range of 0-600 mM. At this protein concentration the population of dimer is negligible. Chemical shift differences between 0 and 50 mM RE along the protein sequence are shown in Figure 3. The solvent accessibility of each residue and the location of secondary

structure elements are also shown. The most perturbed regions correspond to loops. The large loops ranging from residues 46 to 55 and from residues 120 to 135 that form the dimerization interface of lmwPTP show large perturbations. The most perturbed residues include charged, polar uncharged as well as hydrophobic groups. Although the perturbed residues in general are solvent exposed there are other residues with higher solvent accessibility that are unaffected.

Insert Figure 3 here

FKBP12 NMR studies also map RE perturbations to loops

In order to check if the effect of RE on loop residues is a general feature, we studied the effect of different RE concentrations on the ^{1}H chemical shift and ^{15}N relaxation rates of FKBP12.

Figure 4 shows the ^{1}H chemical shift changes of FKBP12 induced by the addition of 50 mM of RE, the solvent accessibility of the different residues and the location of secondary structure elements. The mobility of different regions of FKBP12 in the presence and absence of 50 mM RE was characterized by ^{15}N relaxation. The residue specific R_2/R_1 values, corrected for the viscosity differences between the two buffers, are shown in Figure 5.

Insert Figure 4 here

Insert Figure 5 here

RE induced chemical shift perturbations of FKBP12 are mainly associated to loops and do not correlate with solvent accessibility or residue type. Relaxation data confirm that the residues that are more perturbed by the addition of RE are located in loops and probably reflect chemical exchange effects. The largest R_2/R_1 differences are observed for Gly86 and are clearly outside the experimental error. Smaller differences are observed for residues Gly83, His87, Ile91 and Thr96 in the same loop. While the differences are subtle, the result was confirmed at two different FKBP12 concentrations. At the higher concentration Arg42, located in a second mobile region, shows a significant decrease of R_2/R_1 values in the presence of 50 mM RE. The fact that this residue is only affected at high concentrations suggests that the effect may result from minor aggregation being suppressed by the addition of RE.

Chymotrypsinogen A loops are collapsed on the folded core

SAXS data reveal the compaction of the lmwPTP dimer but do not detect any changes in the lmwPTP monomer. This suggests that the effect of RE on the conformational space sampled by small loops is too modest to cause a detectable change by SAXS. To directly characterize the effect of RE on a monomeric protein using SAXS we selected chymotrypsinogen A that contains

296 large disordered loops not observed in the crystal structure of the protein. As a
297 control we measured SAXS data of FKBP12 under the same conditions.

298 SAXS curves of chymotrypsinogen A and FKBP12 were measured under
299 equivalent conditions in the presence of 0, 50, 100 and 200 mM RE. The R_g
300 values extracted from the curves are plotted vs. RE concentration in Figure 6. As
301 a reference, the R_g expected from the published crystal structures (1chg Freer et
302 al. 1979, 1d7h Burkhard et al. 2000) are indicated.

303 The R_g of chymotrypsinogen A in the absence of RE is 22.3 Å and
304 decreases to 17.8 Å in 200 mM RE. Most of the compaction takes place below 50
305 mM RE. The presence of RE does no affect de R_g value of FKBP12 and this
306 value is in excellent agreement with the one predicted from its crystal structure
307 confirming that the RE buffer does not perturb the SAXS measurements.

308 The R_g value calculated from the crystallographic structure of
309 chymotrypsinogen A is 17.3 Å. Since the mobile loops are not present in the
310 structure, this value reflects the dimensions of the folded core. The compact form
311 of chymotrypsinogen A in the presence of 200 mM has a R_g value of only 17.8 Å,
312 just slightly larger than that of the protein core observed in the crystal.
313 Comparison of the complete SAXS curve with the ordered part of
314 chymotrypsinogen A using CRYSOL shows a not very good agreement in the
315 absence of RE that dramatically improves in the presence of 200 mM RE (Figure
316 7). These results suggest that the mobile loops of chymotrypsinogen A collapse
317 onto the surface of the folded core in the presence of RE.

318

319 **Insert Figure 7 here**

The additional interactions between the loops and the protein core are expected to increase the proteins thermodynamic stability. Indeed, chymotrypsinogen A melting temperature increases by nearly 4K in the presence of 200 mM RE (Table 1). As a comparison, we measured also the stability of maltose binding protein (MBP), a highly soluble periplasmic protein that does not have large disordered loops on its surface. RE had no effected on the stability of MBP, although the stabilization by sodium sulphate and the destabilization by sodium thiocianate, two classical Hoffmeister type cosolutes (Tadeo et al. 2007), and also the stabilization by its ligand, maltose, could be clearly observed.

Insert Table 1 here

DISCUSSION

Reported effects of 50 mM RE on the solution properties of proteins include: a) Increase in protein solubility (Golovanov et al. 2004); b) Increase in the second virial coefficient (Valente et al. 2005); c) Decrease of non-specific interactions of Xe atoms with proteins (Blobel et al. 2007); d) Preservation of specific intermolecular interactions (Golovanov et al. 2004); e) Increase of the stability of oligomers (Blobel et al. 2007); f) Increase in protein thermal stability (Vedadi et al. 2006); g) Increase in the resistance to proteolytic degradation (Golovanov et al. 2004).

Some of the observed effects are clearly related and suggest that RE causes a reduction in the exposure of hydrophobic regions of the protein that are

responsible for non-specific protein-protein interactions measured by the second virial coefficient and favouring aggregation, and also for the reduced interaction with a non-specific hydrophobic probe like Xe atoms. On the other hand, the preservation of specific interactions with other macromolecules suggests that RE is not affecting equally all exposed regions of the protein. The observed enhancement of the stability of the dimers in the case of lmwPTP indicates an effect of RE in oligomerization processes. Moreover, the increase in thermal stability observed with many, but not all, of the proteins whose solubility is enhanced by the addition of RE indicates also suggests an effect on the unfolding equilibria. The entropic cost of rigidifying loops upon compaction could reduce or compensate the energetically positive intramolecular interactions explaining the negligible thermal stabilization of some proteins in RE containing buffers.

Finally, the surprising observation of the increased stability to proteases of proteins that are not significantly thermally stabilized by RE remains unexplained. It has been speculated that it has a completely different origin from other RE effects, and is related to competitive inhibition of contaminating proteases by the added amino acids (Golovanov et al. 2004).

The effects of RE on three unrelated proteins, all of which display disordered loops on their surface (Figure 8), have been measured by NMR and SAXS. While NMR provides information at an atomic scale, SAXS provides a low resolution view of large amplitude changes induced by the co-solutes. NMR chemical shifts and relaxation data on FKBP12 and lmwPTP presented in the results section and

367 elsewhere (Blobel et al 2007) show that the residues most perturbed by RE are
368 located in loop regions. Conversely, exposed residues in structured regions are
369 typically not perturbed by RE suggesting that the effect of RE is associated with
370 changes in the conformation of flexible regions, rather than accessible to co-
371 solutes per se.
372 SAXS measurements provide complementary information with respect to NMR.
373 In the published crystallographic structure of chymotripsinogen A 23 of the 245
374 residues have atoms that could not be located in the electron density. The portion
375 of the protein that is observed in the crystal structure roughly corresponds to the
376 folded protein core. The R_g calculated from this structure is much smaller than
377 the experimentally observed one in the absence of RE in which the larger volume
378 occupied by the disordered loops is also detected. However, the SAXS curve of
379 chymotrypsinogen A in 200 mM RE is in good agreement with the one calculated
380 using only the folded core. The solution R_g value is only slightly larger than the
381 one calculated for the ordered core strongly suggesting that the disordered loops
382 have collapsed onto the protein core in 200 mM RE.
383 Despite the effects of RE on the chemical shifts of FKBP12 and monomeric
384 lmwPTP, RE does not have any significant effect on the measured R_g values for
385 these two proteins, suggesting that the volume sampled by the loops of these two
386 proteins in the absence of RE is smaller than in the case of chymotrypsinogen A.
387 However, a significant compaction was observed for the lmwPTP dimer. It is
388 likely that compaction of the dimer interface, which is formed by the interaction of
389 symmetry-related loops from both protomers, causes changes in the dimer

dimensions that are large enough to be observed by SAXS. Thus, the compaction of lmwPTP dimer and that of chymotrypsinogen A can be seen as two examples of minimization of the effective protein volume in the presence of RE.

The thermodynamic stability of chymotrypsinogen A is increased in the presence of RE. In principle this could be caused by stabilization of the folded form or destabilization of the unfolded state. To check if RE can destabilize unfolded states we measured the melting temperature of MBP in different conditions. MBP is larger than chymotrypsinogen A, but it does not have any large disordered loops. If RE had a general destabilizing effect on unfolded proteins, denaturation temperature of MBP would be expected to increase in the presence of RE. The results in Table 1 show that RE has no effect on the stability of MBP in contrast to classical osmolites like Na_2SO_4 or NaSCN which show a clear stabilization and destabilization, respectively. The stabilization resulting from maltose binding could also be clearly observed for MBP. These results are consistent with the hypothesis that chymotrypsinogen A stabilization occurs in the folded state and is a consequence of the additional interactions between the core and the collapsed loops.

A unified explanation for RE effects on proteins

The results presented in this work show a common theme in the three systems studied: the addition of RE selectively affects disordered loops on the protein

surface leading to compaction, probably by forcing their collapse onto the stably folded protein core. This observation suggests a unified and general explanation for all the effects caused by RE on many different proteins. Disordered regions protruding from the protein surface of an otherwise well folded protein are often mediating non-specific interactions and protein aggregation and are also especially susceptible to proteolytic degradation. By collapsing these regions onto the protein core, non-specific interactions and proteolytic degradation are reduced.

The additional intramolecular interactions between the collapsed loops and the core may stabilize the folded state and contribute to the increased thermal stability in some cases. The enhanced stability would be a consequence of the collapse of the loops and not the other way around (i.e. the lower exposure of disordered regions may be the primary effect and not a consequence of a higher thermal stability of the protein core). Thus, in some cases it may be possible that the collapse of small loops contributes to reduce aggregation and increase protein solubility but has no measurable influence in the thermal stability of the protein core, as has been observed by Golovanov et al. (2004).

Protein compaction, as observed in the presence of RE has been previously observed with much larger concentration of sugars (Kim et al. 2003), polyols (Kaushik and Bhat 1998) and trimethylamine N-oxide (TMAO, Stanley et al. 2008). Reduction of protein aggregation by these co-solutes has been associated to intramolecular burying of hydrophobic patches.

The origin of the RE effect: the Polyelectrolyte Induced Compaction model

Which is the mechanism by which RE causes the collapse of disordered loops on the protein surface? Arginine is one of the most common additives used to prevent aggregation during refolding of denatured proteins (De Bernardez Clark, 2001) and has been shown to prevent heat-induced aggregation (Shiraki et al, 2002). Either arginine or glutamic acid separately reduce aggregation induced by partial thermal unfolding at 50 mM concentration although higher concentrations of arginine have larger effects (Shiraki et al. 2002). Arginine prevents aggregation during refolding of lysozyme by dilution from 8M urea but glutamic acid induces aggregation under these conditions. Protection from heat-induced aggregation by arginine and glutamic acid shows no correlation with the isoelectric point of the protein (Shiraki et al. 2002). Very high concentrations of arginine (0.5-2 M) decrease the thermal stability of proteins (Arakawa and Tsumoto, 2003). The aggregation suppressing effect of arginine has been related to its capacity to bind to both hydrophobic and hydrophilic amino acid side chains offsetting its preferential exclusion from protein surfaces suggested by surface tension measurements (Arakawa et al. 2007). An alternative explanation based on the formation of arginine clusters has been suggested by Das et al. (2007).

The use of glutamic acid as co-solute is limited by its low solubility. Sodium glutamate at high concentration stabilizes both lysozyme and BSA (Arakawa and Timasheff 1984). The solubility of glutamic acid can be increased to much higher concentrations in the presence of equimolar amounts of arginine. The increase in

459 protein solubility in 50 mM RE equimolar mixture is not observed in the presence
460 of the same concentration of the individual amino acids (Golovanov et al 2004).
461 RE also has a larger effect than arginine on the second virial coefficient of
462 lysozyme (Valente et al 2005).
463 In contrast to other co-solutes, the RE mixture increases protein solubility at low
464 concentrations. The non-additive effects of the components of the RE mixture
465 and the differences from classical co-solutes like Na_2SO_4 and NaSCN on the
466 melting temperature of MBP suggests that RE acts, at least in some cases, by a
467 different mechanism. It has been speculated that charged amino acids interact
468 and mask oppositely charged groups on the protein surface while the
469 hydrophobic parts of the side chains of arginine and glutamic acid cover adjacent
470 hydrophobic groups preventing aggregation (Golovanov et al 2004). However,
471 preferential interaction would result in protein destabilization as the denatured
472 state usually has more accessible interaction sites. In the presence of 50 mM RE
473 a significant number of proteins are stabilized by more than 4ºC (Vedadi et al
474 2006) or not affected.
475 We would like to suggest an alternative model, the Polyelectrolyte Induced
476 Compaction (PIC), that could explain the loop compaction effect caused by RE.
477 The model is based on the following premises:
478 a) Arginine and glutamic acid are both molecular polyelectrolytes. Each of them
479 has at neutral pH three charged interaction sites, even though the global charge
480 is +1 for Arg and -1 for Glu. The molecular polyelectrolyte nature of arginine

481 could be further enhanced by the formation of clusters as suggested by mass
482 spectrometry (Das et al 2007, Julian et al 2001).
483 b) The interaction between sites of different molecules will introduce a spatial
484 correlation between them. The asymptotic behaviour of the correlation function
485 will be either monotonic or oscillatory depending on the concentration of the
486 electrolyte. Dressed interaction site theory (DIST) predicts a shifting to lower ionic
487 concentrations, of the point at which the leading asymptotic term turns from a
488 strictly monotonic decay to a damped oscillatory behaviour in the case of
489 molecular electrolytes (Gonzalez-Moruelos et al 2005). Model calculations for a
490 20 Å rigid rod with four equally charged sites and spherical, single site, counter
491 ions predict that the onset of the transition would take place at a concentration of
492 25 μM. At this point the effective decay length would be 262 Å. This distance is
493 the effective decay length of the effective potential between oppositely charged
494 species and it has the same order of magnitude as the dimensions of a large
495 macromolecule. While equivalent calculations for a real arginine-glutamic acid
496 system are not feasible at this point we assume that, due to their molecular
497 polyelectrolyte nature, the interaction length will be significantly longer than
498 expected for simple ions.
499 c) In the absence of protein, RE solutions will reach a minimum energy state
500 which will include the effect of the long range correlation between arginine and
501 glutamate.
502 d) The presence of the protein is expected to perturb the electrolyte solution and
503 increase the free energy of the system. If the protein can adopt different

conformations, those that minimize the occupied volume will also minimize the perturbation and will be favoured, leading to compaction.

e) PIC is reminiscent of other phenomena favouring protein states that occupy the lowest possible volume, such as steric macromolecular crowding or surface tension effects opposing the formation of the solvent cavity needed to accommodate the protein. However, it has a different origin and dependency on structural and dynamic factors. In particular, according to our hypothesis, PIC is associated to a *mixture* of interacting co-solutes and originates from the *dynamic correlation* between different chemical species. Thus it may not be quantified by surface tension measurements and is expected to depend on the translational diffusion rates of the electrolytes and the perturbing protein.

f) The correlation between molecular electrolytes affects their relative motions taking place at a velocity related to their translational diffusion which is large in the case of small molecules, such as arginine and glutamic acid. The perturbation induced by a protein in the solution will be different depending on its mobility. For a denatured or unfolded protein, local diffusion is fast and the perturbation will be minimal. Thus, in contrast to macromolecular steric crowding, small polyelectrolytes are not expected to stabilize proteins by destabilizing the denatured state. On the other hand, loops protruding from the surface of a large protein and exploring a considerable volume will cause a large perturbation on the electrolyte system since their translational motion is restricted to that of the complete protein. Thus, RE favours the reduction of protein volume by favouring intramolecular interactions of loops with the folded core. The burying of exposed

sites would be finally responsible for the observed reduction of non-specific interactions and concomitant increase in solubility and stability to proteases.

Although the PIC model explains qualitatively the observed behaviour of RE in different systems, it is still speculative and it is difficult to derive quantitative predictions that could be tested. In addition, it is likely that other mechanisms may be simultaneously contributing to the observed effects. However, the PIC model provides a unified explanation for the different effects of RE observed and can be generalized to other systems, leading to qualitative predictions, some of which have been verified.

If the polyelectrolyte nature of RE is responsible for the observed effects, other polyelectrolyte systems should show similar behaviour. We have tested over 50 different mixtures of organic polyelectrolytes and observed stabilization of chymotrypsinogen A by more than 4ºC in 46 mixtures and in half of the cases the stabilization was higher than 10ºC at co-solute concentrations below 600 mM. These results will be reported in detail elsewhere (Sornosa et al. in preparation). A large number of small biological molecules such as amino-acids or polyamines can also be considered as polyelectrolites and may exert similar effects to proteins in cellular environments.

The PIC model could also explain the chaperone activity of intrinsically disordered proteins ERD10 and ERD14 produced by plants under stress conditions (Kovacs et al 2008). Their chaperone activity is associated to the presence of charged repetitive regions that can be described as polyelectrolytes.

549 We suggest that the PIC model provides a new conceptual framework to
550 understand the modulation of protein solution properties by multiple charged
551 molecules. This framework can direct the search for new stabilizing additives for
552 biotechnological applications and help to understand biologically relevant
553 transient interactions involving intrinsically disordered chaperones.
554
555 **Acknowledgements**
556
557 We thank Dmitri I. Svergun from EMBL Hamburg for providing SAXS measuring
558 time. The work was partially supported by funds from the Spanish MICINN
559 (BIO2007-63458 and BIO2010-15683 to MP) the Generalitat de Catalunya
560 (2009SGR-1352) and the EU 7th FP BIO-NMR (contract 261863). J.B. was a
561 recipient of a predoctoral fellowship from the Spanish Ministerio de Education y
562 Ciencia. P.B. holds a Ramón y Cajal contract that is partially financed by the
563 Spanish Ministry of Education and by funds provided to the IRB by the
564 Generalitat de Catalunya.
565
566 **References**
567
568 Åkerud T, Thulin E, Van Etten RL, Akke M. (2002) Intramolecular dynamics of
569 low molecular weight protein tyrosine phosphatase in monomer-dimer
570 equilibrium studied by NMR: a model for changes in dynamics upon target
571 binding. J Mol.Biol 322:137-152.

572 Arakawa T, Timasheff SN (1984) The mechanism of action of Na glutamate,
573 lysine HCl, and piperazine- N,N'-bis(2-ethanesulfonic acid) in the
574 stabilization of tubulin and microtubule formation. J Biol Chem 259:4979-
575 4986.
576 Arakawa T, Tsumoto K (2003) The effects of arginine on refolding of aggregated
577 proteins: not facilitate refolding, but suppress aggregation. Biochem
578 Biophys Res Commun 304:148-152.
579 Arakawa T**,** Ejima D**,** Tsumoto K**,** Obeyama N**,** Tanaka Y, Kita Y, Timasheff SN
580 **(**2007**)** Suppression of protein inter- actions by arginine: A proposed
581 mechanism of the arginine effects. Biophys Chem 127**:**1-8
582 Bernadó P, Åkerud T, De la Torre JG, Akke M, Pons M. (2003) Combined use of
583 NMR relaxation measurements and hydrodynamic calculations to study
584 protein association. Evidence for tetramers of low molecular weight protein
585 tyrosine phosphatase in solution. *J Am Chem Soc* **125**:916-923.
586 Blobel J, Schmidl S, Vidal D, Nisius L, Bernadó P, Millet O, Brunner E, Pons M
587 (2007) Protein tyrosine phosphatase oligomerization studied by a
588 combination of ^{15}N NMR relaxation and ^{129}Xe NMR. Effect of buffer
589 containing arginine and glutamic acid. J Am Chem Soc 129:5946-5953.
590 Blobel J, Bernadó P, Svergun DI, Tauler R, Pons M (2009) Low-Resolution
591 Structures of Transient Protein-Protein Complexes using Small-Angle X-Ray
592 Scattering. *J Am Chem Soc* **131**:4378-4386.

593 Blobel J, Bernadó P, Xu H, Jin C, Pons M (2009b) Weak oligomerization of low-
594 molecular-weight proteína tyrosine phosphatase is conserved from
595 mammals to bacteria. FEBS J 276:4346–4357.
596 Brath U, Akke M. (2009) Differential responses of the backbone and side-chain
597 conformational dynamics in FKBP12 upon binding the transition-state
598 analog FK506: implications for transition-state stabilization and target
599 protein recognition. J Mol Biol 387:233-244.
600 Brath U, Akke M, Yang D, Kay LE, Mulder FAA (2006) Functional dynamics of
601 human FKBP12 revealed by methyl ^{13}C rotating frame relaxation dispersion
602 NMR spectroscopy. J Am Chem Soc 128:5718–5727.
603 Burkhard P, Taylor P, Walkinshaw MD (2000) X-ray structures of small ligand-
604 FKBP complexes provide an estimate for hydrophobic interaction energies.
605 J Mol Biol 295: 953-962.
606 Das U, Hariprasad G, Ethayathulla AS, Manral P, DasTK, Pasha S, Mann A,
607 Ganguli M, Verma AK, Bhat R, Chandrayan SK, Ahmed S, Sharma S, Kaur
608 P, Singh TP, Srinivasan A (2007) Inhibition of protein aggregation:
609 supramolecular assemblies of arginine hold the key PLoS ONE 2(11):
610 e1176. doi:10.1371/journal.pone.0001176.
611 De Bernardez Clark, E (2001) Protein folding for industrial processes. Curr Opin
612 Biotechnol 12:202-207.
613 Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a
614 multidimensional spectral processing system based on UNIX pipes. J
615 Biomol NMR 6: 277-293.

616 Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH (1970)
617 Chymotrypsinogen: 2.5-angstrom crystal structure, comparison with alpha-
618 chymotrypsin, and implications for zymogen activation. Biochemistry 9:
619 1997-2009.
620 Golovanov AP, Hautbergue GM, Wilson SA, Lian L-Y (2004) A simple method for
621 improving protein solubility and long-term stability. J Am Chem Soc
622 126:8933-8939.
623 González-Mozuelos P, Yeom MS, Olvera de la Cruz M (2005) Molecular
624 multivalent electrolytes: microstructure and screening lengths. *Eur. Phys. J.*
625 *E 16:167-178.*
626 Guinier A. (1939) La diffraction des rayons X aux tres petit angles: Application à
627 létude de phénomènes ultramicroscopique. Ann Phys (Paris) 12:161–237.
628 Julian RR, Hodyss R, Beauchamp JL (2001) Cooperative salt bridge stabilization
629 of Gas-phase zwitterions in neutral arginine clusters. J Am Chem Soc
630 123:3577–3583.
631 Kaushik JK, Bhat R (1998) Thermal stability of proteins in aqueous polyol
632 solutions: role of the surface tension of water in the stabilizing effect of
633 polyols. J Phys Chem B 102:7058–7066.
634 Kendrick BS, Carpenter JF, Cleland JL, Randolph TW (1998) A transient
635 expansion of the native state precedes aggregation of recombinant human
636 interferon-gamma. Proc Natl Acad Sci USA 95:14142–14146.
637 Kim YS, Jones LS, Dong AC, Kendrick BS, Chang BS, Manning MC, Randolph
638 TW, Carpenter JF (2003) Effects of sucrose on conformational equilibria

639 and fluctuations within the native-state ensemble of proteins. Protein Sci
640 12:1252–1261.
641 Konarev, PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS:
642 a Windows PC-based system for small-angle scattering data analysis. J
643 Appl Cryst 36: 1277-1282.
644 Kovacs D, Kalmar E, Torok Z, Tompa P (2008) Chaperone activity of ERD10
645 and ERD14, two disordered stress-related plant proteins. Plant
646 Physiology147: 381–390.
647 Roessle MW, Klaering R, Ristau U, Robrahn B, Jahn D, Gehrmann T, Konarev P,
648 Round A, Fiedler S, Hermes C, Svergun DI (2007) Upgrade of the small-
649 angle X-ray scattering beamline X33 at the European Molecular Biology
650 Laboratory, Hamburg. J Appl Cryst. 40:190-194.
651 Saunders AJ, Davis-Searles PR, Allen DL, Pielak GJ, Erie DA (2000) Osmolyte-
652 induced changes in protein conformation equilibria. Biopolymers 53:293–
653 307.
654 Shiraki K, Kudou M, Fujiwara S, Imanaka T, Takagi M (2002) Biophysical effect
655 of amino acids on the prevention of protein aggregation. J Biochem
656 132:591–595.
657 Standaert RF, Galat A, Verdine GL, Schreiber S (1990) Molecular cloning and
658 overexpression of the human FK506-binding protein FKBP. Nature 346:671-
659 674.
660 Stanley C, Krueger S, Parsegian VA, Rau DC. Protein structure and hydration
661 probed by SANS and osmotic stress. Biophys J 2008;94:2777-2789.

662 Svergun DI, Barberato C and Koch MHJ (1995) CRYSOL - a Program to
663 Evaluate X-ray Solution Scattering of Biological Macromolecules from
664 Atomic Coordinates. J Appl Cryst 28:768-773.
665 Tabernero L, Evans BN, Tishmack PA, Van Etten RL, Stauffacher CV (1999) The
666 structure of the bovine protein tyrosine phosphatase dimer reveals a
667 potential self-regulation mechanism. Biochemistry 38:11651–11658.
668 Valente JJ, Verma KS, Manning MC, Wilson WW, Henry CS (2005) Second
669 virial coefficient studies of cosolvent-induced protein self-interaction.
670 Biophysical J 89: 4211–4218
671 Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty Jr PJ, Wasney GA,
672 Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD, Nordlund
673 P, Sundstrom M, Weigel AM, Edwards AM. (2006) Chemical screening
674 methods to identify ligands that promote protein stability, protein
675 crystallization, and structure determination. Proc Natl Acad Sci USA
676 103:15835-15840.
677 Wo YY, Zhou MM, Stevis P, Davis JP, Zhang ZY, Van Etten RL (1992) Cloning,
678 expression, and catalytic mechanism of the low molecular weight
679 phosphotyrosyl protein phosphatase from bovine heart. Biochemistry
680 31:1712-21.
681 Zhang M, Van Etten RL, Stauffacher CV(1994) Crystal structure of bovine heart
682 phosphotyrosyl phosphatase at 2.2 resolution. Biochemistry 33:11097–
683 11105.
684

685

686 Table 1. Co-solute induced changes in melting temperature

687

Protein	**Buffer**[a]	**$\Delta T * K^{-1}$**
Chymotrypsinogen A[b]	50 mM RE	+1.4
Chymotrypsinogen A[b]	200 mM RE	+3.9
MBP[c]	50 mM RE	+0.5
MBP[c]	200 mM RE	+0.7
MBP[c]	200 mM Na_2SO_4	+3.1
MBP[c]	200 mM NaSCN	-7.8
MBP[c]	2 mM maltose	+6.0

688

689 [a] 25 mM phosphate, 150 mM NaCl plus the indicated cosolute. Protein

690 concentration was 10 μM. The estimated uncertainty in ΔT is 1K.

691 [b] pH 6.5

692 [c] pH 7.5

693

FIGURE CAPTIONS

Figure 1 Apparent radius of gyration (R_g) and maximum dimension (D_{max}) observed as a function of lmwPTP concentration in the absence (filled dots) and presence (empty circles) of 50 mM RE. The theoretical values for the monomer (dashed dotted line) and dimer (dashed line) are indicated.

Figure 2 MCR-ALS extracted pure SAXS curves for lmwPTP monomer (red) and dimer (green) and their fit to the corresponding crystal structures (dashed lines) by CRYSOL (Svergun et al. 1995). An equivalent analysis to a SAXS dataset measured in the absence of RE yield SAXS curves of both species that were in excellent agreement with crystallographic structures (see Figure 5A in Blobel et al. 2009). The minimization error as a function of the assumed dimer dissociation constant is shown in the inset.

Figure 3 $^{1}H^{\alpha}$ chemical shift changes induced by 50 mM RE in lmwPTP (absolute values). Dashed line is the average $^{1}H^{\alpha}$ chemical shift change measured for buried residues in the protein and therefore are not affected by RE. The localization of secondary structure elements is indicated. Filled rectangles represent α-helices and empty ones β-sheets strands. Relative solvent accessibilities (rel S_{exp}) are indicated. The residues most affected by RE are labelled and their accessibility is shown by filled bars.

Figure 4 $^1H^{\alpha}$ chemical shift changes induced by 50 mM RE in FKBP12 (absolute values). Dashed line is the average $^1H^{\alpha}$ chemical shift change measured for buried residues in the protein and therefore are not affected by RE.The localization of secondary structure elements is indicated. Filled rectangles represent α-helices and empty ones β-sheets strands. Relative solvent accessibilities (rel S_{exp}) are indicated. The residues most affected by RE are labelled and their accessibility is shown by filled bars.

Figure 5 R_2/R_1 data for 0.94 mM (A) and 1.4 mM (B) FKBP12 in RE-free buffer (filled circles) and in the same buffer plus 50 mM RE (open circles). The RE data has been corrected for viscosity. Residues showing the largest differences are indicated by downward arrows. α-helices and β-sheet strands are represented by filled and open rectangles, respectively.

Figure 6 SAXS derived radius of gyration (R_g) of chymotrypsinogen A (dashed line) and FKBP12 (continuous line) at different RE concentrations. The experimentally obtained R_gs are shown as circles. R_g values calculated from the crystal structures are shown by horizontal lines.

Figure 7 a) SAXS scattering curves of chymotrypsinogen A recorded in the presence of different concentrations of RE (0 blue, 50 mM yellow, 100 mM magenta, and 200 mM turquoise) and CRYSOL fits to the crystallographic structure 1chg. Experimental data are shown by circles and the fit by lines. The y-

axis showing the scattering intensity I(s) is presented in log scale. Curves are displaced in the y-axis for proper visualization. b) Point-by point error ($I(s)_{exp} - I(s)_{calc}$)/($\sigma(s)_{exp}$), of the fitting of each curve.

Figure 8 Crystallographic structures of a) lmwPTP monomer, b) lmwPTP dimer, c) FKBP12, and d) chymotrypsinogen A with surface loops highlighted. Disordered loops not observed in the crystal structure of chymotrypsinogen A are indicated by dashed lines.