Thermodynamic relationships between protein-solvent and protein-protein interactions

L. Costenaro[†] and C. Ebel[*]

Laboratoire de Biophysique Moléculaire, Institut de Biologie Structurale UMR 5075:CEA-CNRS-UJF, 41 rue Jules Horowitz, 38027 Grenoble, France. E-mail: Christine.Ebel@ibs.fr

How the solvent modulates the weak inter-particle interactions in solution and affects macromolecule solubility is not yet understood. Well-established thermodynamic relationships link second virial coefficient and preferential solute binding parameter. We present the meaning of these thermodynamic parameters and the way to measure them. When a solvation shell has a composition different from the bulk solvent, a negative contribution is found in the second virial coefficient corresponding to an effective attraction between the macromolecules in solution. A quantitative evaluation using simple models of solvated particles in solution suggests that solvation could induce, at high or low concentration of a small molecule solute, attractive inter-particle interactions corresponding to favorable crystallization conditions.

Keywords: solvent; interactions; thermodynamics; hydration; solubility; binding; second virial coefficient.

1. Introduction

Weak interactions, as excluded-volume forces, coulombic repulsion, van der Waals attraction or solvation effects, determine the tendency for a suspension of particles to remain in solution, to aggregate, to overcome phase separation or to form crystals. Crystallization still represents a crucial and frequently very difficult step for the determination of three-dimensional structures at high resolution of biological macromolecules. Their weak inter-particle interactions in under-saturated or super-saturated solutions have been the object of numerous studies with the aim of understanding their capacity to form crystals. For a number of proteins in solvents known to promote crystallization, the values of the second virial coefficient A_2, reflecting the net inter-particle interactions, are moderately negative and lie within a fairly narrow range, about -1.0 to -8.0 10^{-4} mol ml g^{-2} (George & Wilson, 1994; George *et al.*, 1997; Guo *et al.*, 1999; Bonneté *et al.*, 1999).

Inter-particle potentials have been modeled from the concentration dependencies of the static and transport, i.e. equilibrium and hydrodynamic, properties of proteins in solution (Pusey & Tough, 1985; Malfois *et al.*, 1996; Ducruix *et al.*, 1996; Georgalis & Saenger, 1999). It is clear that the inter-particle interactions and protein solubility depend on the solvent conditions. For example, the Hofmeister series allow to classify salts with respect to their general effect on macromolecule solubility (Von Hippel & Schleich, 1969). The anion order was found however to be reversed in the case of basic proteins (Riès-Kautt & Ducruix, 1991; 1997). It is thus obvious that there are connections between macromolecule-solvent interactions and macromolecule-macromolecule interactions.

A_2 can be expressed rigorously as a function of $\int(1 - g(r))r^2dr$ with g(r) the protein pair distribution function and r the inter-particle distance. g(r) depends implicitly on the protein-solvent and solvent-solvent interactions. For spherical particles, g(r) is related to the potential of mean force W(r) (Hansen, 1986): g(r) = exp(-W(r)/kT). W(r) has a precise definition: it is the difference in free energy between the solution with two proteins at separation r and the same at infinite dilution. Statistical thermodynamics allows to derive g(r) and W(r) from all the direct potentials between all the solution molecules: macromolecules and solvent components (for a recent review see: Belloni, 2000).

In the present paper, we consider a less detailed description of the macromolecule-macromolecule interactions. Two thermodynamic parameters can be easily measured: the second virial coefficient A_2 and the preferential solute binding parameter. We present their significance and the way to measure them. We recall the relationships between them, using well-established thermodynamics. It indicates that a solvation shell with a composition different from bulk corresponds to an effective attractive term between the macromolecules. This contribution is quantitatively and qualitatively discussed.

2. Theoretical background

2.1. The "experimental" device: osmotic pressure

The temperature T is considered to be constant, and will not be mentioned further. In a first stage, we consider two solutions *A* and *B* separated by a dialysis membrane (Fig. 1a). All the solvent components, water (w) and small solute (s), are able to be redistributed through the membrane (in the present manuscript, we will use "solute" only for the both solvent components and not for the macromolecule). *A* contains only the solvent components. *B* contains in addition the macromolecule, for example a protein (p). The macromolecule causes solvent redistribution in order to equilibrate the chemical potentials of the diffusible components in *A* and *B* and reach the dialysis equilibrium condition. There are two causes for the solvent redistribution.

The first one is related to a simple dilution effect of the solvent components due to the presence of the macromolecule. The chemical potential μ_i of the component i (i = p, w or s) can be expressed as a function of its chemical potential in a standard state $\mu_i°$ and activity a_i or activity coefficient γ_i in the molal scale and molality m_i (mol/kg of water), with R the gas constant. In this case, component w is not apparent in the equations. Alternatively, μi can be expressed in a mole fraction scale, using $\mu'_i°$, A_i, the concentration in mole fraction f_i and activity coefficient in the mole fraction scale χ_i:

$$\begin{aligned}\mu_i &= \mu_i^o + RT\ln a_i = \mu_i^o + RT\ln m_i\gamma_i \\ &= \mu_i'^o + RT\ln A_i = \mu_i'^o + RT\ln f_i\chi_i.\end{aligned} \qquad (1)$$

Since in *A*, the mole fractions of the solvent components (w) and (s) are larger than in *B* because of the macromolecule, the chemical potentials of the solvent components are larger in the bulk solvent *A*. The presence of the macromolecule, without considering any effect of neither solvation nor interactions, induces a flux of solvent from *A* to *B*. If the size of the compartment *B* is restricted, the solvent flux will cause an excess of pressure, which is the osmotic pressure Π, P being the reference pressure in compartment *A*, usually the atmospheric pressure. The equilibrium condition thus corresponds to:

$$\mu_{w\,A(P)} = \mu_{w\,B(P+\Pi)} \qquad \mu_{s\,A(P)} = \mu_{s\,B(P+\Pi)}. \qquad (2)$$

[†] Present address: Biological Chemistry Dept., John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

Another cause for solvent redistribution is related to the non-ideal behavior of the macromolecule. We will emphasize below the macromolecule solvation effect (part 2.4).

In a second stage (Fig. 1b), the two compartments are separated and pressure P restored. This is what the experimentalist obtains after a dialysis experiment: the content of a dialysis bag at atmospheric pressure, or after gel filtration. The composition of *A* and *B* are unchanged by this process, which however affect the chemical potentials of all components i by $\int_{P+\Pi}^{P} d\mu_i$. In the particular case of the solvent components, using in addition Eq. (2), it leads to:

$$\mu_{w\,or\,s\,B(P)} - \mu_{w\,or\,s\,B(P+\Pi)} = \mu_{w\,or\,s\,B(P)} - \mu_{w\,or\,s\,A(P)} = -\int_{P}^{P+\Pi} d\mu_{w\,or\,s\,B} \quad . \qquad (3)$$

It can be seen that Π is the excess of pressure that would have to be added to the actual pressure P of a solution *B* in order to maintain a chemical equilibrium with a phase of pure solvent *A*.

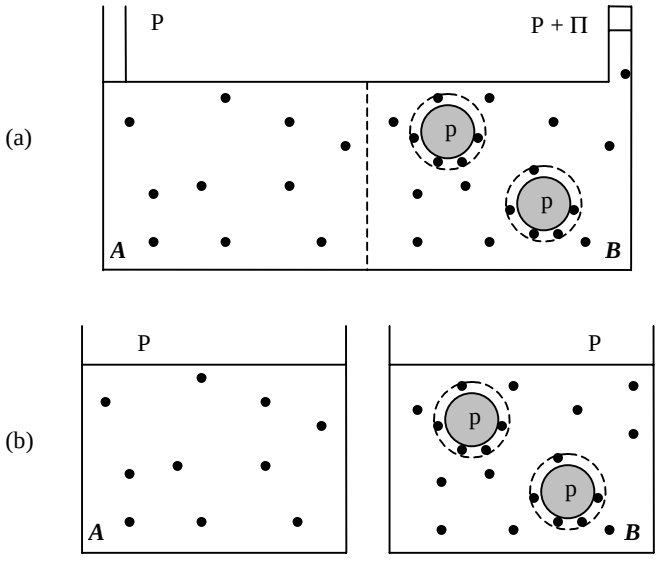

Figure 1

Dialysis experiment, osmotic pressure and chemical potentials. (a): In the dialysis equilibrium, the chemical potentials of the diffusible solvent components (water: not symbolized, and small solute: black dots) in compartment *B*, containing the protein (p), are the same as those in compartment *A*, containing the bulk solvent. The composition of the solvent components in *B* may differ from *A*, in consequence of macromolecule-solvent interactions. The dash circle around the protein symbolizes the perturbed solvent (solvation). The difference of pressure Π is related in the ideal case to the number of macromolecules in solution. Π is modulated by the intermolecular interactions, including protein-protein and protein-solvent interactions. (b): When atmospheric pressure is restored, the composition in the compartments *A* and *B* are unchanged, but the changes in the chemical potentials of the solvent components between the compartments *A* and *B* are related to Π.

2.2. Osmotic pressure and chemical potential of the solvent components

The chemical potential depends in general on the solution composition, T and P. In the described transition process from Fig. 1a to 1b, the temperature and composition are not changed: the chemical potential variations come only from Π. The partial molal volume is by definition $(\partial\mu_i/\partial P) = \overline{V}_i$. Therefore, assuming (reasonably) constant values of $\overline{V}_i$ when the pressure is increased from P to P + Π one obtains:

$$\mu_{w\,or\,s\,B(P)} - \mu_{w\,or\,s\,A(P)} = -\overline{V}_{w\,or\,s} \cdot \Pi \quad . \qquad (4)$$

It is equivalent to measure the difference of the chemical potentials of water in compartments *A* and *B* at the same pressure, or the osmotic pressure defined at constant solvent chemical potentials.

2.3. Osmotic pressure, molar mass and second virial coefficient

Using equations (1) expressed in terms of mole fraction and (4) for water, Π is related to the difference ($\ln a_{w\,B} - \ln a_{w\,A}$) and so to the ratio of the water mole fraction in the two solutions. Basic transformations allow to express Π as a function of the molality m_p of the protein (in mol/kg of water) and solution or solvent volumes, V_m or $V_m°$ (in ml/kg of water), for a diluted solution, in a form close to diluted gas systems:

$$\Pi = RT\, m_p/V_m° + \ldots. \qquad (5)$$

This shows that osmotic pressure is a colligative method that, roughly speaking, counts the number of macromolecules. Using a mass unit concentration (g/ml): $c = M_p m_p/V_m$, yields to the familiar equations that allows to determine the molar mass M_p of the macromolecule:

$$\Pi / RT = c/M_p + A_2c^2 + A_3c^3\ldots \qquad (6)$$

$$(d\Pi/dc) / RT = 1/M_p + 2A_2c + 3A_3c^2\ldots \qquad (7)$$

A_2 is the second virial coefficient and A_3 is the third, neglected at moderate concentrations.

2.4. Solvent interactions: the preferential solute binding parameter $(\partial m_s/\partial m_p)_\mu$

The protein-solvent interactions induce a rearrangement of the solvent, which will affect the solvent composition in a dialysis experiment. Preferential binding parameters are defined at constant chemical potential of solvent component (subscript μ). Since there is no solvent rearrangement when the pressure is shifted from P+Π to P (Fig. 1b), preferential binding parameters can be measured at pressure P (see below part 3.3). The preferential solute binding parameter $(\partial m_s/\partial m_p)_\mu$ is related to the relative variation of the chemical potentials of protein and solute when increasing the solute concentration:

$$(\partial m_s/\partial m_p)_\mu = -\,(\partial\mu_p/\partial m_s)_m \,/\, (\partial\mu_s/\partial m_s)_m = -\,a_{ps}/a_{ss}\,, \qquad (8)$$

where $a_{ij} = (\partial\mu_i/\partial m_j)_m/RT$ and the subscript m signifies constancy of molalities of all components except the derived one, at constant pressure (the subscript P being omitted). The preferential binding parameter $(\partial m_s/\partial m_p)_\mu$ corresponds to the number of moles of solute that would have to be added (or removed) with one mole of the macromolecules in order to maintain constant the chemical potential of solvent components. In a structural approach, considering the solvent perturbed by the protein binding N_w moles of water and N_s moles of solute per mole of protein, the preferential solute binding

parameter is related to the solvent composition expressed in molar ratio m_s/m_w:

$$(\partial m_s/\partial m_p)_\mu = N_s - N_w \,.\, (m_s/m_w) \,. \qquad (9)$$

Solvent binding sites statistically occupied with water and solute in the composition of the bulk lead to a null contribution in $(\partial m_s/\partial m_p)_\mu$. Alternatively, the preferential hydration parameter $(\partial m_w/\partial m_p)_\mu$ can be used since it is related to the preferential solute binding parameter by the solvent composition (see Fig. 2).

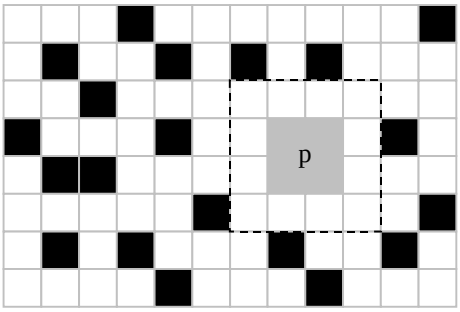

Figure 2

Understanding preferential solvent binding parameters. Small white square: water (w); small black square: small solute (s); large gray square: macromolecule (p). In this example, the ratio of water on solute m_w/m_s is 3 in the bulk solvent and 12 molecules of water are bound to the protein. They have to be added with the (nude) protein to maintain constant the chemical potential of the solvent: $(\partial m_w/\partial m_p)_\mu = 12$. On the other hand, we can consider that 4 molecules of small solute, previously associated with the 12 molecules of water in the solvent, were removed upon the introduction of the macromolecule, so that $(\partial m_s/\partial m_p)_\mu = -4$. The preferential solute binding parameter and preferential hydration parameter are related by the solvent composition by $(\partial m_s/\partial m_p)_\mu \,/\, (\partial m_w/\partial m_p)_\mu = -m_s/m_w$. Note that the hypothesis of strong binding is not needed, nor the definition of the limit of the domain corresponding to perturbed solvent.

2.5. Second virial coefficient for a three component system

The general Gibbs-Duhem equation, with n_i the number of particles i in the considered system, is:

$$SdT - VdP + \Sigma n_i d\mu_i = 0 \,. \qquad (10)$$

In the dialysis equilibrium conditions, i.e. with $dT = 0$, $dP = d\Pi$, $d\mu_w = d\mu_s = 0$, it reduces to:

$$V_m \,(d\Pi/dm_p) = m_p \,(\partial\mu_p/\partial m_p)_\mu \,. \qquad (11)$$

For the 3-component system considered here, μ_p is a function of P, m_p, m_s (m_w being constant by definition) so that $d\mu_p$ can be expressed via $\overline{V}_p\, dP = (\partial\mu_p/\partial P)_m dP$, $a_{pp} dm_p$ and $a_{ps} dm_s$; and $(\partial\mu_p/\partial m_p)_\mu$ via $\overline{V}_p\,(d\Pi/dm_p)$, a_{pp} and $a_{ps}(\partial\mu_p/\partial m_p)_\mu$. From Eq. (1), a_{pp}, related to protein concentration and a non-ideality term, can be expressed as:

$$a_{pp} = (\partial\mu_p/\partial m_p)_m/RT = 1/m_p + a_{pp}^{(e)} \,. \qquad (12)$$

Expanding V_m with $V_m°$ and m_p, and dropping inconsequential terms, leads to the following relation (Scatchard, 1946; Casassa & Eisenberg, 1964; Eisenberg, 1976):

$$\begin{aligned} d\Pi/dm_p &\approx RT/V_m^o \cdot \left(a_{pp} - a_{ps}^2/a_{ss}\right) m_p \\ &= RT/V_m^o + RT/V_m^o \cdot \left(a_{pp}^{(e)} - a_{ps}^2/a_{ss}\right) m_p \,. \end{aligned} \qquad (13)$$

Using a concentration in mass/volume units, the term $RT/V_m°$ leads to $1/M_p$ of the osmotic equations (6) and (7) that provides the protein molar mass, and the second virial coefficient A_2 is expressed by:

$$A_2 = \partial V_m/\partial m_p)_\mu°/M_p^2 + A_{2pp}^{(e)} + A_{2ps} \,. \qquad (14)$$

The first term arising from concentration scale change is small and can be approximated by $\overline{v}_p/M_p$, $\overline{v}_p$ being the partial specific volume. The second one $A_{2pp}^{(e)}$ is:

$$A_{2pp}^{(e)} = (V_m°/2M_p^2)\, a_{pp}^{(e)} \,. \qquad (15)$$

The $a_{pp}^{(e)}$ term is not (at least easily) measurable. It results form the introduction of a nude protein to the solution: so that the solvation process will change the solvent component chemical potentials. It can be estimated considering protein-protein potentials unrelated to the solvent redistribution, such as the excluded-volume, electrostatic repulsion... The third term of Eq. (14), A_{2ps}, can be expressed with the preferential solute binding parameter (Eq. 8):

$$A_{2ps} = -\,(V_m°/2M_p^2)\, a_{ps}^2/a_{ss} = -\,(V_m°/2M_p^2)\, a_{ss}\, (\partial m_s/\partial m_p)_\mu^2 . \qquad (16)$$

A_{2ps} is related to the protein-induced solvent redistribution. As a_{ss} is always positive and $(\partial m_s/\partial m_p)_\mu$ is squared, A_{2ps} is always negative. The equation (16) brings to the fore that a solvation shell different from the bulk solvent could induce an effective attraction between macromolecules.

2.6. Polyelectrolytes and salt dissociation

The above expressions (except Eq. 9) are general. The dissociation of the macromolecule and its co-ions will modify the development of a_{pp}, a_{ps}, A_{2pp} and A_{2ps} in terms of species molalities and add a Donnan contribution (Eisenberg, 1976). The equation (9), which provides a structural interpretation of the preferential binding parameters, will be expressed using $(N_s\text{-}E_s)$ instead of N_s, with the Donnan term E_s related to the dissociation of the co-ions of the macromolecule.

3. Techniques for the measurement of the thermodynamic parameters

3.1. Second virial coefficient from equilibrium properties

The second virial coefficient is obtained by the protein concentration dependencies of Π or $(d\Pi/dc)$ as expressed by the equations (6) or (7). $(d\Pi/dc)$ can be obtained from the forward intensity I(0) of static light scattering, small angle neutron scattering or small X-rays scattering, or from the concentration profiles $dlnc/dr^2$ at equilibrium in ultracentrifugation (r is the distance to the axe of rotation). These techniques are used in the diluted case for the determination of the molar mass M_p of the protein, but the experimental parameters can be expressed as a function of $(d\Pi/dc)^{-1}$ (Eisenberg, 1976; 1981):

$$I(0) \propto (\partial\rho_x/\partial c)_\mu^{\,2}\, c\, (d\Pi/dc)^{-1}, \qquad (17)$$

$$\mathrm{d}\ln c/\mathrm{d}r^2 \propto \omega^2 (\partial\rho/\partial c)_\mu \ (\mathrm{d}\Pi/\mathrm{d}c)^{-1}. \quad (18)$$

The contrast term $(\partial\rho_x/\partial c)_\mu^2$ related to protein-solvent interactions is the square of the refractive index increment $(\partial n/\partial c)_\mu$, the neutron scattering length density increment $(\partial\rho_N/\partial c)_\mu$, the electron density increment $(\partial\rho_{el}/\partial c)_m$ or the mass density increment $(\partial\rho/\partial c)_\mu$. ω is the angular velocity in the centrifuge. Neglecting non-ideality in data treatment provides an apparent molar mass M_{app} for the protein in solution, whose value depends on the protein concentration according to the following virial expansion:

$$1/M_{app} = 1/M_p + 2A_2c + \ldots \quad (19)$$

3.2. Second virial coefficient from hydrodynamic properties

The spatial and velocity correlations between interacting particles in solution also modify the dynamic properties of the macromolecules in solution: the collective diffusion coefficient D, frictional coefficient f, and sedimentation coefficient s, and their concentration dependencies. For moderate protein concentrations, s, D and f can be described by the following linear approximations:

$$s/s° = f°/f = 1/(1+ k_s c + \ldots) , \quad (20)$$
$$D/D° = (1+ k_D c + \ldots) . \quad (21)$$

The Svedberg equation relating s, D and M_p can be also expressed as a function of $(\mathrm{d}\Pi/\mathrm{d}c)^{-1}$:

$$s/D = (\partial\rho/\partial c)_\mu \ (\mathrm{d}\Pi/\mathrm{d}c)^{-1}. \quad (22)$$

A_2 can be calculated from k_D and k_s by:

$$k_D \approx 2\ A_2M_p - k_s . \quad (23)$$

We have recently investigated the potential of sedimentation velocity analytical ultracentrifugation for the measurement of the second virial coefficients of proteins. Using global modeling for three sets of data obtained at three different protein concentrations, very good estimates for k_s and s°, and also for D° and the buoyant molar mass, and good estimates for k_D and the second virial coefficients were obtained (Solovyova *et al.*, 2001).

3.3. Measurement of the preferential binding parameters

The mass density increment $(\partial\rho/\partial c)_\mu$ or the neutron scattering length density increment $(\partial\rho_N/\partial c)_\mu$ corresponds to the increase in density ρ, or neutron scattering length density ρ_N, due to the addition of 1 g/ml protein. $(\partial\rho/\partial c)_\mu$ can be obtained by density measurement, or analytical ultracentrifugation, since it determines the buoyancy of the macromolecule (Eq. 18). $(\partial\rho_N/\partial c)_\mu$ can be obtained by the measurement of I(0)/c in small angle neutron scattering experiments (Eq. 17). Similar expressions can also be derived for small angle X-ray scattering.

$(\partial\rho/\partial c)_\mu$ and $(\partial\rho_N/\partial c)_\mu$ can be expressed as functions of the partial molal volume, molar masses and neutron scattering length densities of the components, $\overline{V}_i$, M_i and b_i (cm/mol), and of the preferential solute binding parameter $(\partial m_s/\partial m_p)_\mu$:

$$\begin{aligned} M_p(\partial\rho/\partial c)_\mu &= (M_p - \rho^o\overline{V}_p) + (\partial m_s/\partial m_p)_\mu \cdot (M_s - \rho^o\overline{V}_s) \\ M_p(\partial\rho_N/\partial c)_\mu &= (b_p - \rho_N^o\overline{V}_p) + (\partial m_s/\partial m_p)_\mu \cdot (b_s - \rho_N^o\overline{V}_s) \end{aligned} \quad (24)$$

The complementarities of these equations were discussed in (Ebel, 1995; Kernel *et al.*, 1999; Ebel *et al.*, submitted).

4. Results and discussion

In section 2, it was emphasized that protein-solvent interactions modify the chemical potential of all components in solution, protein, water and co-solute. Particularly, it was shown that the solvent interactions could lead to a negative contribution in the second virial coefficient. In Section 3, we have described the way to measure the thermodynamic parameters related to inter-particle interactions. The second virial coefficient is related to the protein distribution in solution, and to effective inter-protein potentials. The preferential solute binding parameter depends on the changes in the solvent composition in consequence of protein-solvent interactions. In the present section, we evaluate the strength of the negative contribution related to the protein-solvent interactions. Are they negligible or have they to be taken into account?

4.1. Calculation of A_{2ps} from protein solvation

We consider a model of a three-component system composed of water, solute and protein of 60 kDa. We define on the protein a limited number of strong (saturated) binding sites for water (N_w) and solute (N_s). A number of other weak solvent binding sites may exist, which have the same affinity for all the solvent components and provide a null contribution in the preferential binding parameter. Of course, the model of strong binding sites is simple, but we note that all models of protein-solvent interactions leading to the same preferential binding parameter – which is the experimentally accessible parameter – will have the same consequence.

We consider realistic values for N_w of 100, 1000 and 3000 mol/mol and selected N_s values of 0, 3, 10 and 60 mol/mol. Indeed, the accessible surface area of bovine serum albumin (BSA) of similar molar mass, is given as 29,000 Å^2, which corresponds to a continuous monolayer of about 3200 water molecules, assuming that a water molecule occupy an area of 9 Å^2. For BSA, in the presence of sugars and organic solvents, values for N_w between 500 and 2800 with N_s = 0 can be inferred from experimental measurements (Courtenay *et al.*, 2000, Ebel *et al.*, 2000). For a very acidic protein of 130 kDa, malate dehydrogenase from *Haloarcula marismortui*, in the presence of salts, we determined values for N_w between 2000 and 4000, and for N_s, up to 85 (Kernel *et al.*, 1999, Ebel *et al.*, submitted). We consider solvent compositions characterized by (m_s/m_w) values of 0.001, 0.01 and 0.1 mol/mol, corresponding to solute molalities of 0.055, 0.55, and 5.5 mol/kg of water (corresponding molarities in mol/l are in the same order of magnitude).

N_w and N_s are used to calculate the preferential binding parameter $(\partial m_s/\partial m_p)_\mu$ using Eq. (9). This value is used in Eq. (16) for the calculation of A_{2ps}, the contribution to the second virial coefficient A_2 that is related to protein-induced solvent redistribution. A global second virial coefficient A_2 is calculated considering for $A_{2pp}^{(e)}$ only the excluded volume interaction contribution, which is positive, independent of the solvent composition, and can be approximated in the case of spheres by: $A_{2pp}^{(e)} = 4\,\overline{v}_p/M_p$, with $\overline{v}_p$ = 0.74 ml/g the partial specific volume of the protein (Tanford, 1961). The results are given on Table 1.

On the top part of Table 1, the protein is solvated only by water (N_s = 0). The preferential solute binding parameter $(\partial m_s/\partial m_p)_\mu$ is always negative. Its absolute value increases at constant solvent composition (same ratio m_s/m_w) with the amount of bound water (N_w) or at constant amount of bound water with the solute concentration (m_s/m_w). Non-zero values of $(\partial m_s/\partial m_p)_\mu$ correspond always to negative values of A_{2ps}. Because A_{2ps} is related also to a_{ss},

which is in the ideal case related to $1/m_s$, the effect on A_{2ps} of a given value of $(\partial m_s/\partial m_p)_\mu$ is larger at small solute concentration (compare lines 2 and 4, or lines 3 and 5 of Table 1). The contribution of the negative values of A_{2ps} could lead to negative values for the net second virial coefficient A_2. For example, we obtain (line 6) a A_2 value of -2 10^{-4} mol ml g^{-2} for 1000 water molecules strongly bound to the 60 kDa protein, at high solute concentration ($m_s/m_w = 0.1$).

Table 1

Influence of the solvation on weak interactions. One mole of a protein of molar mass M_p = 60 kDa binds strongly N_w moles of water and N_s moles of solute, in a binary solvent of composition m_s/m_w. m_i are molalities (mol/kg of water). The preferential solute binding parameter $(\partial m_s/\partial m_p)_\mu$ and its contribution A_{2ps} to the second virial coefficient A_2 are calculated from Eq. (9) and (16) with the following approximations: $V_m° = 1000$ ml, $a_{ss} = 1/m_s$. We considered for $A_{2pp}^{(e)}$ only the excluded-volume part (see the text). A_2 is the sum $A_{2ps} + A_{2pp}^{(e)}$.

	N_w	N_s	m_s/m_w	$(\partial m_s/\partial m_p)_\mu$	A_{2ps}	$A_{2pp}^{(e)}$	A_2
	mol/mol				10^{-4} mol ml g^{-2}		
1	100	0	0.001	-0.1	-0.00025	0.5	0.5
2	100	0	0.01	-1	-0.0025	0.5	0.5
3	100	0	0.1	-10	-0.025	0.5	0.5
4	1,000	0	0.001	-1	-0.025	0.5	0.5
5	1,000	0	0.01	-10	-0.25	0.5	0.2
6	1,000	0	0.1	-100	-2.5	0.5	-2.0
7	3,000	0	0.001	-3	-0.225	0.5	0.3
8	3,000	0	0.01	-30	-2.25	0.5	-1.8
9	3,000	0	0.1	-300	-22.5	0.5	-22.0
10	1,000	3	0.001	2	-0.1	0.5	0.4
11	1,000	3	0.01	-7	-0.1	0.5	0.4
12	1,000	3	0.1	-97	-2.4	0.5	-1.9
13	1,000	10	0.001	9	-2.0	0.5	-1.5
14	1,000	10	0.01	0	0	0.5	0.5
15	1,000	10	0.1	-90	-2.0	0.5	-1.5
16	1,000	60	0.001	59	-87.0	0.5	-86.5
17	1,000	60	0.01	50	-6.3	0.5	-5.8
18	1,000	60	0.1	-40	-0.4	0.5	0.1

On the second part of Table 1, we consider the protein as solvated by both water and solute. The values of $(\partial m_s/\partial m_p)_\mu$ can be negative, null or positive. A positive value corresponds to a solvation shell that contains more solute than the bulk solvent. For constant values of N_w and N_s, a positive value of $(\partial m_s/\partial m_p)_\mu$ is more likely to be obtained at low solute concentration (lines 10, 13, 16, 17). Increasing the solute concentration in the solvent leads to a solvation shell containing less solute than the bulk solvent, and thus to negative values of $(\partial m_s/\partial m_p)_\mu$ (lines 10 to 12, 13 to 15, and 16 to 18). For a given value of N_w and in a given solvent composition, increasing values of N_s provides increasing values of $(\partial m_s/\partial m_p)_\mu$ (compare lines 10, 13 and 16; 11, 14 and 17; or 12, 15 and 18). Values of A_{2ps} are always negative or null. They are null when the solvation composition exactly matches the solvent one (line 14). This can be intuitively understood because it means that the protein does not perturb the solvent. As mentioned above, the effect of protein-solvent interactions on A_2 is strongly modulated by the solute concentration. For a given value of $(\partial m_s/\partial m_p)_\mu$, it is larger at small solute concentration in the solvent. The presence of a limited number of strong solute binding sites has a large effect on A_{2ps} at low solute concentration (lines 13, 16, 17). As an example, we calculate a positive $(\partial m_s/\partial m_p)_\mu$ value of 9 for N_w = 1000 and N_s = 10 mol/mol for a solvent composition m_s/m_w of 0.001 mol/mol (line 13). The corresponding A_{2ps} negative value is about -2 10^{-4} mol ml g^{-2}.

These examples underline the fact that whatever the sign of $(\partial m_s/\partial m_p)_\mu$, the protein-induced solvent redistributions always introduce an effective protein-protein attraction as evaluated from the second virial coefficient ($A_{2ps} < 0$). When the protein is only solvated by water, this effective protein-protein attraction increases with the solute concentration. When the protein is solvated by both water and solute, this effect is expected not only at high, but also at low solute concentration.

4.2. Solvent interactions and protein solubility

It is admitted that more negative values of the second virial coefficient are related to decreased values of the solubility. The theoretical relationships between A_2 and protein solubility were investigated considering the differences between the free energy of the protein in solution and in the crystal state (Haas *et al.*, 1999; Ruppert *et al.*, 2001). Restricting only to the properties of the solution, basics thermodynamics shows that the chemical potentials of all components are related to osmotic pressure, and thus to A_2. For the water and solute, combining Eq. (4) and (6) gives:

$$\begin{aligned}\mu_{w\,or\,s\,B(P)} - \mu_{w\,or\,s\,A(P)} &= -\overline{V}_{w\,or\,s}\cdot\Pi \\ &= -\overline{V}_{w\,or\,s}\,RT\left(c/M_p + A_2c^2\cdots\right).\end{aligned} \qquad (25)$$

For an ideal solution ($A_2 = 0$), the presence of the macromolecule decreases the value of the chemical potential of the solvent components (water and solute). A positive value of A_2 strengthens the effect of dilution at higher macromolecule concentration. A negative one has the opposite effect. The evolution of the chemical potentials of water and solute as a function of protein concentration, considering the virial expansion up to the second terms, is plotted on Fig. 3 for three values of A_2. For A_2 value of 0.5 10^{-4} mol ml g^{-2}, corresponding to the excluded volume effect only, they decrease continuously when increasing the protein concentration. For negative A_2 values, they increase at high protein concentration. Above a protein concentration of about 80 and 30 mg/ml, for A_2 values of -2 and -5.8 10^{-4} mol ml g^{-2}, respectively, the chemical potential of the solvent would be higher in the protein solution than in the bulk, which is an improbable situation. The system has thus to evolve. One possibility is a phase separation (towards two liquid phases, or with protein precipitation as ordered crystals or aggregates). An other one is a modification of the macromolecule conformation, leading to a change in the preferential binding parameter, and thus decreasing the chemical potential of the solvent. The latter will not be discussed here.

Considering Fig. 3, moderately negative second virial coefficients in the range -2 to -10 10^{-4} mol ml g^{-2} correspond to reduce the protein solubility in the mg/ml or tenths of mg/ml range. They correspond potentially to good crystallization conditions. From Table 1, these second virial coefficient values could be induced by the solvent redistribution close to the protein. They would correspond to a moderate number of bound water at high solute concentration (lines 6 and 12), a high number of bound water at moderate solute concentration (line 8). When there are binding sites for the solute on the macromolecule, these conditions could be also found at low or moderate solute concentration (lines 13, 17) and at high solute concentration (line 15). A solute is called salting-in or salting-out when its presence in solution increases or decreases, respectively, the protein solubility. It is clear from Table 1 that the salting-in and salting-out situation could be obtained for the same solute at different concentrations, depending on the protein-solute interactions.

5. Conclusion

The effect of protein preferential hydration on protein-protein interactions is generally admitted (Parsegian, 2000). The salting-out

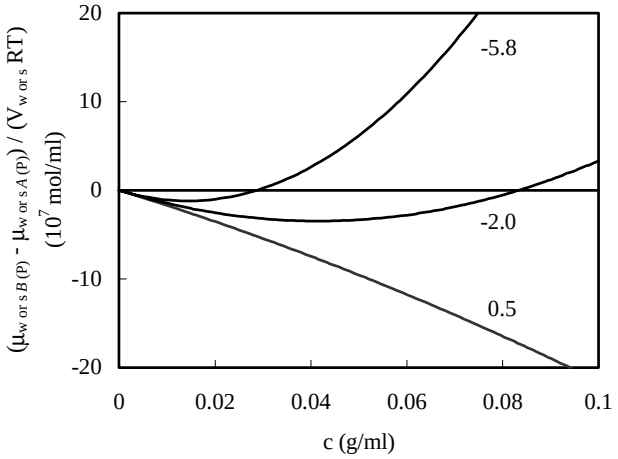

Figure 3

Second virial coefficient and protein solution stability. The difference in the chemical potentials of the solvent components between the protein solution and the bulk solvent is calculated with Eq. (25) for three A_2 values, indicated in 10^{-4} mol ml g^{-2}.

efficiency of various salts was related to the measurement of increasing values of preferential hydration for protein. The solutes that increase the solution surface tension, inducing a non-specific hydration of the interfaces, are described to decrease protein solubility (Timasheff, 1993). The depletion of large particles as polyethylene glycol also explains their precipitating effect (Arakawa & Timasheff 1985; Finet & Tardieu 2001). Less documented are the effects of the interactions between solute and macromolecule. We have shown here from classical thermodynamic relationships that water accumulation or water depletion at the surface of protein – corresponding to small solute depletion or accumulation – have similar consequences on the values of the second viral coefficient: a solvation shell composition different from the bulk solvent decreases A_2 values, and thus lowers protein solubility. Our conclusions neglect however the possible effect of the solvent composition in the second component of the second virial coefficient, the $a_{pp}^{(e)}$ term. We have recently measured in a variety of salt solvents the experimental preferential binding parameters and second virial coefficients of malate dehydrogenase from *Haloarcula Marismortui* (Costenaro *et al.*, submitted). This very acidic protein is adapted to high salt – the cytoplasm of *H. marismortui* is nearly saturated in salt – and show unusual solvent binding properties (Ebel *et al.*, submitted). The experimental results on the relationships between the protein-solvent and weak protein-protein interactions using this model protein comfort the theoretical relationships presented here, with decreased values of the second virial coefficient at either low or high salt concentration in the solvent.

References

Arakawa, T. & Timasheff, S. N. (1985). *Biochemistry*, **24**, 6756-6762.
Belloni, L. (2000). *J. Phys.Condens. Matter* **12**, R549-R587.
Bonneté, F., Finet, S. & Tardieu, A. (1999). *J. Crystal Growth*, **196**, 403-414.
Casassa, F. E. & Eisenberg, H. (1964). *Adv. Protein Chem.* **19**, 287-395.
Costenaro, L., Zaccai, G. & Ebel, C. (2002). Submitted.
Courtenay, E. S., Capp, M. W, Anderson, C. F. & Record M. T. Jr (2000). *Biochem.* **39**, 4455-4471.
Ducruix, A., Guilloteau, J. P., Ries-Kautt, M. & Tardieu A. (1996). *J. Crystal Growth*, **168**, 28-39.
Ebel, C. (1995). *Progr.Colloid Polym. Sci.* **99**, 17-23.
Ebel, C., Eisenberg, H. & Guirlando, R. (2000). *Biophys. J.* **78**, 385-393.
Ebel, C., Costenaro, L., Pascu, M., Faou, P., Kernel, B., Proust - De Martin, F. & Zaccai, G. (2002). Submitted.
Eisenberg, H. (1976). *Biological macromolecules and polyelectrolytes in solution*. Oxford: Clarendon Press.
Eisenberg, H. (1981). *Q.. Rev. Biophys.* **14**, 141-172.
Finet, S. & Tardieu, A. (2001). *J. Crystal Growth*, **232**, 40-49.
Georgalis, Y. & Saenger, W. (1999). *Sci. Prog.* **82**, 271-294.
George, A. & Wilson, W. W. (1994). *Acta Cryst.* D**50**, 361-365.
George, A., Chiang, Y., Guo, B., Arabshahi, A., Cai, Z. & Wilson, W. W. (1997). *Methods Enzymol.* **276**, 100-110.
Guo, B., Kao, S., McDonald, H., Asanov, A., Combs, L. L. & Wilson, W. W. (1999). *J. Crystal Growth*, **196**, 424-433.
Hansen, J. P. & McDonald, I. R. (1986). *Theory of simple liquids*, 2nd edition. London: Academic Press.
Haas, C., Drenth, J. & Wilson, W. W. (1999). *J. Phys. Chem. B*, **103**, 2808-2811.
Kernel, B., Zaccai, G. & Ebel, C. (1999). *Prog. Colloid Polym. Sci.* **113**, 168-175.
Malfois, M., Bonneté, F., Belloni, L. & Tardieu, A. (1996). *J. Chem. Phys.* **105**, 3290-3300.
Parsegian, V. A., Rand, R. P. & Rau, D. C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3987-3992.
Pusey, P. N. & Tough, J. A. (1985). *Dynamic light scattering, Applications of photon correlation spectroscopy*, pp. 85-179. New York: Plenum Press.
Riès-Kautt, M. M. & Ducruix, A. F. (1991). *J. Cryst. Growth*, **110**, 20-25.
Riès-Kautt, M. M. & Ducruix, A. F. (1997). *Methods Enzymol.* **276**, 23-59.
Ruppert, S., Sandler, S.I. & Lenhoff, A.M. (2001). *Biotechnol. Prog.* **17**, 182-187.
Scatchard, G. (1946). *J. Am. Chem. Soc.* **68**, 2315-2319.
Solovyova, A., Schuck, P., Costenaro, L. & Ebel, C. (2001). *Biophys. J.* **81**, 1868-1880.
Tanford, C. (1961). *Physical Chemistry of Macromolecules*, pp. 180-274. New York: John Wiley & Sons Inc.
Timasheff, S. N. (1993). *Annu. Rev. Biophys. Biomol. Struct.* **22**, 67-97.
Von Hippel, P. H. & Schleich, T. (1969). *Structure and Stability of Biological Macromolecules*, edited by S. N. Timasheff & G. D. Fasman, Vol. 2, pp. 417-574. New York: Marcel Dekker.

 Received 10 April 2002 • Accepted 7 August 2002