A Pairwise Preferential Interaction Model for Understanding Peptide Aggregation

Myungshim Kang · Paul Edward Smith

Abstract A pairwise preferential interaction model (PPIM), based on Kirkwood–Buff integrals, is developed to quantify and characterize the interactions between some of the functional groups commonly observed in peptides. The existing experimental data are analyzed to determine the preferential interaction (PI) parameters for different amino acid and small peptide systems in aqueous solutions. The PIs between the different functional groups present in the peptides are then isolated and quantified by assuming simple pairwise additivity. The PPIM approach provides consistent estimates for the pair interactions between the same functional groups obtained from different solute molecules. Furthermore, these interactions appear to be chemically intuitive. It is argued that this type of approach can provide valuable information concerning specific functional group correlations which could give rise to peptide aggregation.

1 Introduction

Common diseases such as Alzheimer's, Parkinson's, and Huntington's disease, as well as other amyloid and fibril related disorders, are directly linked to the misfolding and subsequent aggregation of specific peptides or proteins [1–3]. However, our current understanding of the factors that lead to aggregation is poor. An important question is whether one can predict if a specific peptide will self-aggregate based on the functional groups present in the molecule? In addition, can one even quantify the

M. Kang · P. E. Smith (✉)
Department of Chemistry, Kansas State University, 213 CBC Building,
Manhattan, KS 66506-0401, USA
e-mail: pesmith@ksu.edu

degree of aggregation at the molecular level? These are very simple questions that a variety of researchers would like to be able to answer. Unfortunately, our understanding of molecular interactions in complex solvents such as water remains rather limited. While several statistical mechanical theories of solutions have been presented and remain formally accurate [4–6], there has been slow progress in determining the required integrals which form the basis of many of these theories, and extending these theories to understand solution mixtures. Hence, it is not currently possible to accurately calculate the self-association behavior of a particular peptide in solution. It is clear that a simple general methodology which provides (even limited) success in this direction would be welcomed.

Studies of small peptide aggregation represent an active area in which a general approach for predicting the tendencies of different peptides to self-aggregate would also be very useful [7]. The aggregation of small peptides should be a function of the properties of the different amino acid side chains, as modified by the presence of the solvent. Experimental studies of peptide aggregation have led to the characterization of a series of short peptide sequences based on longer sequences or fragments of proteins which are known to aggregate [8–10]. It is clear from many of these studies that peptide aggregation is not limited to peptides rich in hydrophobic residues. Hydrophobic, polar, and highly charged peptides can all display tendencies towards aggregation [8,9]. Hence, studies to design and modify peptide structures in an attempt to investigate the aggregation propensity and mechanism generally involve a trial-and-error approach [8]. Clearly, an improved understanding of aggregation would help to focus efforts towards specific peptide sequences.

In principle, aggregation or self-association of a solute in a solution mixture results from differences in the intermolecular interactions between the solute and solvent molecules. If the solute–solute interactions are larger than solute–solvent interactions, self-association is likely to occur, and vice versa. Hence, the tendency for aggregation could be predictable, if one can determine the difference between the various solute–solute and solute–solvent interactions in a quantitative manner. Proteins are large complex molecules. Therefore, it would be more useful to quantify the interactions between amino acids, or even between functional groups, rather than to deal with the protein as a whole. This assumes that the total interaction between two peptides or proteins can be decomposed into a series of interactions between the component functional groups.

The concepts of preferential solvation (PS) and preferential interactions (PIs) have been developed to help understand solution mixtures [6,11–15]. Kirkwood–Buff (KB) integrals have played an important role in quantifying these PIs [13,16–18], especially from computer simulations [19–22]. In this work, we use the concept of PIs to investigate factors which may lead to peptide aggregation. A pairwise preferential interaction model (PPIM) is developed to quantify and characterize the interactions between some of the common functional groups observed in peptides and proteins. First, the existing experimental data are analyzed to determine PI parameters for different amino acid and small peptide systems in aqueous solutions. Then, the PIs between function groups on those peptides are isolated and quantified using a simple pairwise preferential interaction approach.

2 Pairwise Preferential Interaction Model (PPIM) Approach

2.1 Thermodynamics of Solutions and KB Theory

The notation used here follows the common definition where the subscripts 1 and 2 refer to the primary solvent (usually water) and the solute, respectively. According to statistical mechanics, the chemical potential of a species can be expressed as [6]

$$\mu = W + RT \ln \left[\Lambda^3 \rho q^{-1} \right] = \mu^* + RT \ln \left[\Lambda^3 \rho \right]. \tag{1}$$

Here, q is the internal partition function of the molecule, N is the number of the species, V is the volume of the system, $\rho = N/V$ is the number density (or molar concentration), and Λ is the thermal de Broglie wavelength of the species. The first term (W) quantifies contributions of the interactions between molecules to the chemical potential. Ben-Naim [6] has developed the use of the expression $\mu^* = W - RT \ln q$, referred to as the pseudo-chemical potential. The pseudo-chemical potential describes the free energy change for transfer of a molecule from a fixed position in a vacuum to a fixed position in the solution.

Kirkwood–Buff theory plays an important role in the development of the current model. KB theory is an exact theory of stable solution mixtures containing molecules of any size. The theory provides a direct link between the solution structure and the corresponding thermodynamics. Expressions for derivatives of the pseudo-chemical potential with respect to composition are easily obtained in terms of KB integrals (G_{ij}) which are defined by [6,23]

$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}^{\mu VT}(r) - 1 \right] r^2 \mathrm{d}r, \tag{2}$$

where $g_{ij}(r)$ is the center of mass radial distribution function (rdf) between species i and j.

2.2 Preferential Interactions

Our aim is to develop a method to analyze the experimental data and a model which can be used to quantify and predict peptide aggregation. Here, we present and apply a new approach to quantitatively express features of the interactions between functional groups in amino acids. It is limited in some aspects (discussed below), but in the absence of other predictive approaches, it seems worthy of development. In many respects it is somewhat analogous to, and on the same level as, the Chou–Fasman type of approach used to predict secondary structure elements in proteins [24]. Using KB theory, it is quite easy to show that for any thermodynamically stable mixture of a solute (2) and solvent (1) one can write [6,19]

$$-\beta \left(\frac{\partial \mu_2^*}{\partial \rho_2}\right)_{T,P} = -\left(\frac{\partial \ln y_2}{\partial \rho_2}\right)_{T,P} = \frac{G_{22} - G_{12}}{1 + \rho_2 (G_{22} - G_{12})} \tag{3}$$

where y_2 is the molar activity coefficient of the solute. The above expression reduces to the numerator in the limit of infinite dilution of the solute (2). The value of $G_{22} - G_{21}$ at infinite dilution is the central quantity of interest in this work. We will define this as the PI between two infinitely dilute solute molecules surrounded by solvent molecules

$$P_{22}^\infty = G_{22}^\infty - G_{21}^\infty \tag{4}$$

The PI defined here is the same as previous definitions used to investigate PS [11], except for the infinitely dilute solute restriction. However, it will be used in a different manner. The PI at infinite dilution of solute quantifies the affinity between two solute molecules in a large excess of solvent (no additional many-body solute interactions present). It results from a balance of the pair solute–solute and solute–solvent interactions. A positive value indicates a favorable solute–solute interaction which tends toward solute association or aggregation. A negative value indicates a favorable solvation which tends towards solute hydration and low solute self-association. A value of zero indicates a balance of the interactions, i.e., an ideal solution. The above expression indicates that if the molar activity coefficient decreases with molarity, then the solute must display a tendency towards self-association. The approach therefore provides a way to quantify the degree of molecular association.

2.3 Analysis of the Experimental Data

It is common practice to perform an analysis of the experimental data (solute or water activity, partial molar volumes, and isothermal compressibilities) to obtain the composition-dependent KB integrals and use them to quantify PS in solutions [11,14,25,26]. In principle, we wish to do the same here—but there is a slight problem. We require the KB integrals at infinite dilution of the solute (2). Traditionally, obtaining reliable values of G_{ij} at low concentrations of i or j has been difficult [25,27]. For instance, the partial pressure of a solute above a solute–solvent solution is difficult to determine experimentally at low solute concentrations. Consequently, the KB integrals tend to be sensitive to the exact fitting procedure used to determine the solute activity. Part of the problem is that the two most common fitting equations for the solute activity, the Wilson and Redlich–Kister equations, are largely empirical in nature and therefore may not fully capture the physics of the process. Recently, a rigorous statistical mechanical approach based on a semi-grand ensemble (open to solute, closed to solvent) was presented which can accurately model the molal activity coefficients of nonvolatile solutes over large concentration ranges using just one or two parameters [28].

In the original approach by Rosgen et al. [28,29], either the molal or molar activity coefficient was fitted using ratios of polynomials in the solute activity (a). Here, the approach is extended to simultaneously fit both the molal and molar activity

coefficient data. This requires the corresponding density data. The derivations are quite straightforward, but too long to present here. The final result for the solute molality (m_2) and solute molarity (c_2) after including terms up to a^2 is given by

$$m_2 = \frac{a + 2a^2 A_1^{-1}}{1 + a A_2 + a^2 A_2 A_1^{-1}} \quad \text{and} \quad c_2 = \frac{a + 2a^2 A_1^{-1}}{d_0^{-1} + a V_1 + a^2 V_2 A_1^{-1}} \tag{5}$$

where the molality scale solute activity ($a = \gamma_2 m_2$) has been used in both cases, and d_0 is the density of the pure solvent. At this level, the fitting equations involve four unknowns (A_1, A_2, V_1, and V_2). The major advantage of this approach is that one can fit both activity coefficients over a large range of compositions with just four parameters, one of which (A_1) appears in both expressions. As the molar and molal concentrations are related through the solution density, this is equivalent to fitting one of the activity coefficients and the density simultaneously. Further analysis of the limiting values of the activity derivatives and comparison with Eq. 5 indicates that the PI value required for this work is then given by

$$P_{22}^{\infty} = 2 (d_0 A_1)^{-1} - V_1 \tag{6}$$

The above fitting procedure involves an expression for the concentration in terms of the activity instead of the normal situation of activity (coefficient) in terms of concentration. While this is somewhat unusual, it is a perfectly valid approach to fit the experimental data. If required, the above equations can be used to express the molal activity coefficient in terms of the molality via solutions to a quadratic equation in γ_2 [29]. This is not necessary here. We will see that the above approach produces excellent results. This is important, as it provides significant confidence in the resulting KB integrals, including those evaluated at low solute concentrations.

3 Results

Using KB theory and analyzing the experimental data, one can obtain the PI between two solute molecules at infinite dilution thus providing fundamental information on the degree of molecular association. In this section, the existing experimental data on activity coefficients and densities [15,30–33] are analyzed in order to isolate a series of P_{22}'s. Our initial investigations have focused on NMA (N-methylacetamide) and a series of small amino acids and peptides. In the zwitterionic form, the amino acids are nonvolatile, while NMA is a solid at 298 K and a good model for the peptide functional group.

Figure 1 provides an example of the quality of fits one can obtain using the above fitting equations. The molal activity and densities are very well reproduced. In comparison, the original activity data alone for Gly (glycine) was fitted to an expression involving 4 terms and 4 unknowns [30]. The corresponding parameters and resulting PIs are presented in Table 1. It can be seen that the PI values are positive for all of the systems presented here. This indicates a tendency towards aggregation of the solutes
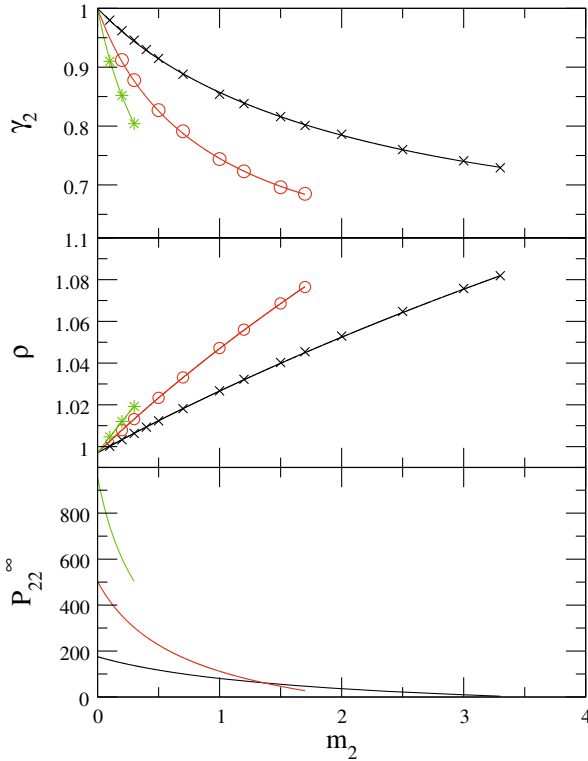
Fig. 1 Experimental data at 298 K and 0.1 MPa for the molal activity coefficients (γ_2), solution density (ρ in $g \cdot cm^{-3}$) and the resulting preferential interactions (P_{22}^{∞} in $cm^3 \cdot mol^{-1}$) for a series of Gly_n peptides as a function of peptide molality (m_2). The *symbols* indicate raw experimental data, while the *solid lines* are the corresponding fits after using Eq. 5 and the parameters presented in Table 1. Gly ($\times$), Gly_2 (O), and Gly_3 ($*$). All PIs are positive indicating a tendency for self-association, which decreases as the solute concentration increases

at low solute concentrations. The aggregation of the Gly_n peptides increases with n for reasons which will become clear shortly.

3.1 Decomposition of the Preferential Interactions

While the previous results are interesting in themselves, we wish to take this a step further and develop a model which can be used to rationalize the available data and eventually make predictions. To do this, we need to be able to decompose the PI between two solute molecules into a combination of effects from the different functional groups present in the molecule. Our approach is to investigate the simplest model possible and then develop the model as it is applied to more systems, where the possibility of (hopefully) small corrections may be required. The basic assumption of the model proposed here is that the PI between a pair of solute molecules can be decomposed into a series of pairwise preferential interactions between the different

Table 1 Fitting parameters obtained for Eq. 5

Solute (2)	Abbrev.	Molal activity range	A_1 (mol·kg^{-1})	A_2 (kg·mol^{-1})	V_1 (L·mol^{-1})	V_2 (L·mol^{-1})	P_{22}^{∞} (cm^3·mol^{-1})
NMA	N	0.5–2.1	56.82	−0.101	−0.020	−0.134	55
Gly	G	0.1–3.3	6.83	0.075	0.118	0.174	176
Gly$_2$	G2	0.2–1.7	2.70	0.164	0.241	0.330	502
Gly$_3$	G3	0.1–0.3	1.68	0.129	0.240	0.372	955
Ala	A	0.1–1.4	18.43	0.0	0.073	0.0	36
Ala$_2$	A2	0.2–1.0	2.89	0.555	0.651	0.0	43
Ala-Gly	AG	0.2–1.0	2.09	0.463	0.814	0.0	147
Gly-Ala	GA	0.2–1.0	2.57	0.366	0.673	0.0	108

Errors in the final PIs are estimated as <10 %. Several parameters adopted values close to zero during the fitting procedure. In this case, the fit was repeated with these values set at zero. All the experimental peptide data refer to racemic mixtures taken from the literature [15,30–33]

chemical groups present in the solutes. This type of approach has been used before for enthalpies of mixing [34], second cross virial coefficients [35], Gibbs free energy of hydration [36], and phase equilibria and excess properties of complex mixtures [37]. But to our knowledge it has not been used for the decomposition of PIs. The groups involve the usual chemical functionalities such as hydroxyl, amide, carbonyl, etc., as well as hydrocarbon-based "interactions" such as those between methyl groups (Me–Me). Hence, the PPIM expresses the total PI between two solute molecules (i and j) at infinite dilution as

$$P_{ij}^{\infty} = \sum_a \sum_b p_{ab}^{\infty} \qquad (7)$$

where the sum is over all functional groups (a) on molecule i and all functional groups (b) on molecule j. The p_{ab}'s then have to be determined. This represents one of the main objectives of the current work. As there are far fewer functional groups than potential solutes, the PPIM model can then be used to predict the degree of molecular association between a variety of different solutes in solution. In the next section, we present evidence that this simple model appears to be reasonable.

To further illustrate the current model, let us consider a solute molecule which contains two functional groups a and b. Both groups on each molecule interact with each other and therefore contribute to the total PI at infinite dilution. In the PPIM model, it is assumed that the solute-to-solute interaction at infinite dilution can be written as a combination of the group interactions such that $G_{22} - G_{21} = p_{aa} + p_{bb} + 2p_{ab}$. The values of p_{aa} and p_{bb} can be obtained from the results for a solute where only one of the groups (a or b) is present. The value of p_{ab} corresponds to the PI between two different groups and can, in general, be obtained via the decomposition process involving the data for many solutes (see later). We note that there is no explicit dependence of the pair interaction on the distance between the different groups on the solute molecules in the PPIM approach. This is because the KB integrals quantify changes

in the molecular distributions over all distances as indicated in Eq. 2. However, it is to be expected that the pair interactions will have a limited range, the extent of which remains to be determined.

3.2 Evaluation of Group Contributions

The PIs presented in Table 1 can be used to evaluate the required group contributions. This is illustrated in Table 2, where we have indicated the number and type of the

Table 2 Decomposition of the PIs observed between molecules into group contributions

Solute	Number of group pairs					
	PP	QQ	PQ	MQ	MP	MM
N	1					
G		1				
G2	1	1	2			
G3	4	1	4			
A		1		2		1
A2	1	1	2	4	4	4
AG	1	1	2	2	2	1
GA	1	1	2	2	2	1

	Group p_{ab} (cm$^3 \cdot$ mol^{-1})						Source
	PP	QQ	PQ	MQ	MP	MM	
	55						N
	63						1/2 (G3 − 2G2 + G)
		176					G
			135				1/2 (G2 − G − N)
			144				1/2 (G3 − G2 − 3N)
					−112		1/2 (AG − A − PP − 2PQ)
						130	1/2 (A2 − AG − A + QQ − 2MP)
				−135			1/2 (A − QQ − MM)
				−137			1/2 (AG − QQ − PP − 2PQ − 2MP − MM)
				−131			1/2 (GA − A − PP − 2PQ)
						169	1/2 (A2 − GA − A + QQ − 2MP)
				−155			1/2 (A − QQ − MM)
				−157			1/2 (GA − QQ − PP − 2PQ − 2MP − MM)
Average	59	176	140	−146	−122	149	
σ	6	−	6	20	13	28	

Solute abbreviations are given in Table 1. Group abbreviations are peptide (P), charge (Q), and methyl (M). The QQ group interaction includes all combinations between two zwitterions (i.e., two +/−, one +/+, one −/− interaction). The PQ interaction is a combination of the P+ and P− interactions (and similarly for MQ). Estimates for the standard deviation (σ) were based on the variation in values obtained from different sources. Additional errors will arise between different experimental activity datasets [25]

group interactions for each pair of solute molecules. By manipulation of the data, one can extract the individual p_{ab} values which are also presented in Table 2. There are several features of the resulting data which are encouraging and important. First, comparison of the p_{ab} value for the peptide group obtained from a separate analysis of the Gly$_n$ peptides and NMA indicates essentially the same result of 59 cm$^3 \cdot$ mol^{-1}. As these correspond to two totally different types of solute systems, it is satisfying that the same result is obtained in both cases. A consistent result for the peptide-to-amino acid termini (PQ) is also obtained from two different decompositions. Inclusion of the Ala-Gly (and Gly-Ala) data provided estimates for the PI between a methyl group and the two charged termini (MQ) as well as the MM and MP interactions. Again, consistent results were obtained for the MP interaction using the same data but from two different routes. Some differences were observed between the Ala-Gly and Gly-Ala results. However, these differences were relatively small (<10 %) and hence, to a first approximation, the sequence dependence effect appears to be minor for short peptides and will therefore be neglected in the present studies.

4 Discussion

The PPIM has several advantages. First, it is very simple. Second, the PIs include both short-range and long-range contributions to changes in the solution distributions. Hence, both the effects of direct molecular interactions (such as hydrogen bonds) and the consequences of the packing of solvent molecules around the solutes are included in the model. The latter is traditionally very difficult to determine. Third, the use of the PIs ensures that one focuses on effects that lead to changes in the pseudo-chemical potential, i.e., that result solely from intermolecular interactions [38]. Fourth, by performing a simultaneous fit of the activity and density data using a rigorous statistical mechanical theory, one ensures accurate experimental data are obtained at low solute concentrations for the determination of the PIs.

One of the disadvantages of the model is that, in its present form, it is restricted to infinitely dilute solute molecules. Obviously, at finite solute concentrations, the solute–solute interactions are modified by the presence of other solute molecules. This will change the values of $G_{22} - G_{21}$. In addition, the solutes cannot be ionic in nature. For ions, the distribution at low solute concentrations is dominated by the electroneutrality constraint [26,39,40] and is therefore independent of the character of the ionic species involved. Furthermore, the current model does not distinguish between different peptide sequences with the same composition of amino acids, such as Gly-Ala and Ala-Gly, or between chiral molecules. As one of the potential applications of the PPIM is to understand peptide aggregation, this could present a problem. However, the experimentally observed activity coefficients of Gly-Ala and Ala-Gly are very similar [32], and hence the corresponding p_{ab}'s do not differ significantly. An additional complication arises as the peptides increase in size. There is no distance dependence associated with the p_{ab}'s and hence the results are independent of the peptide register displayed during aggregate formation. The initial work presented here merely attempts to elucidate possible pair interactions which display a high propensity for aggregation. How this leads to specific peptide structures remains a far more challenging problem.

Finally, the present analysis is restricted to that of nonvolatile solutes in water, primarily due to the fitting procedure. This does not mean that the model is not applicable to other solvents or to molecules which are volatile. Once the p_{ab} for a particular group has been determined, it can be used for any solute in that particular solvent. It is the determination of the p_{ab}'s from representative molecules which requires the use of a fitting equation valid only for nonvolatile solutes. In summary, while there are some restrictions, we feel the model has significant promise, and due to the many potential applications deserves to be developed further. It can certainly provide the first-order effect of group pair interactions on the association process.

We also note that the signs estimated by the model are consistent with those intuitively expected from simple physical chemistry arguments. For instance, it is expected that the interaction between a methyl group and both the charged group (MQ) and the peptide group (MP) would be unfavorable from a desolvation perspective. This is indicated by the negative values of p_{ab}. In contrast, the positive value of p_{ab} for the MM interaction indicates significant self-association, which is in agreement with a simple picture of the hydrophobic effect. The weaker self-association of the peptide group is not unrealistic, although it is difficult to predict the exact balance between self-association and solvation. This, of course, is precisely what makes the present model so attractive. The positive value for PP indicates a tendency for the peptide group to self-associate. The implications of this for protein folding or peptide aggregation are unclear but deserve further study.

The relatively large PI between the peptide and charged groups appears to be just as favorable as the MM self-association. In unrelated simulation studies, we have observed significant similar interactions between Ser hydroxyl groups and the charged C terminus of small peptides [41]. The data in Table 2 suggest that the reason for the increased self-association or aggregation of the Gly_n peptides with increasing n lies in the fact that all the group interactions are positive and thus promote self-association, with the dominant effect occurring for the PQ interaction. The self-association of Ala is lower than that of Gly as the additional methyl group introduces MQ interactions which are unfavorable, presumably due to the effect of the methyl groups on the solvation of the charged groups. Finally, the positive value for p_{ab} between methyl groups is in agreement with the estimates provided previously by analysis of hydrocarbon aggregation in water [27].

Comparison of the present results with known experimental data concerning peptide aggregation is also informative. Several studies have indicated that peptide aggregation is favored for sequences which contain sheet forming residues, often hydrophobic in character, and that oppositely charged residues also play a significant role [8,42,43]. The influence of charged groups has also been observed in computer simulations [44]. However, the role of charged residues appears to be quite involved as several studies have proposed that charge groups often flank aggregation prone sequences in order to prevent aggregation [45,46]. In the latter case the charges are usually both positive. The results shown in Table 2 suggest that charge groups can promote association between peptides by favorable charge–charge and charge–peptide interactions. This clearly supports a significant role for charged groups in promoting peptide aggregation. We note that our definition of the pairwise preferential interaction for QQ is related to interactions between oppositely charged groups. Clearly, sequences flanked by

similar (positively) charged residues will tend to repel and hence decrease the tendency for aggregation [46].

5 Conclusion

In summary, the interactions between different groups present on a solute have been quantified using the simple PPIM. In practice, the PPIM provides consistent estimates for the same pair interactions obtained from different solute molecules, and interactions which are chemically intuitive. The model suggests a role for PQ and MM interactions in peptide aggregation which should be experimentally testable. A relatively limited set of systems and hence a limited number of pairwise interactions have been elucidated here. Furthermore, the model is limited to infinitely dilute solute interactions, although this approximation should be most relevant for biological systems which typically involve very low solute concentrations. Clearly, further study is needed to generalize the model. To some extent the limited set of solutes studied here reflects an absence of activity data for many of the amino acids or other model compounds in the literature. Nevertheless, we feel that the above analysis is promising and provides valuable insights into the role of different functional groups in the aggregation of small peptides.

References

1. C.A. Ross, M.A. Poirier, Nat. Med. **10**, S10–S17 (2004)
2. C. Hetz, C. Soto, Cell. Mol. Life Sci. **60**, 133 (2003)
3. F.E. Cohen, J.W. Kelly, Nature **426**, 905 (2003)
4. D.A. McQuarrie, *Statistical Mechanics* (Harper & Row, New York, 1976)
5. T.L. Hill, *Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, 1960)
6. A. Ben-Naim, *Statistical Thermodynamics for Chemists and Biochemists* (Plenum Press, New York, 1992)
7. D. Thirumalai, D.K. Klimov, R.I. Dima, Curr. Opin. Struct. Biol. **13**, 146 (2003)
8. L. Tjernberg, W. Hosia, N. Bark, J. Thyberg, J. Johansson, J. Biol. Chem. **277**, 43243 (2002)
9. M.F. Perutz, B.J. Pope, D. Owen, E.E. Wanker, E. Scherzinger, Proc. Natl. Acad. Sci. USA **99**, 5596 (2002)
10. C. Wurth, N.K. Guimard, M.H. Hecht, J. Mol. Biol. **319**, 1279 (2002)
11. A. Ben-Naim, Cell Biophys. **12**, 255 (1988)
12. Y. Marcus, J. Chem. Soc. Fara. Trans. **86**, 2215 (1990)
13. E. Matteoli, G.A. Mansoori, *Fluctuation Theory of Mixtures* (Taylor & Francis, New York, 1990)
14. E. Matteoli, L. Lepori, J. Chem. Soc. Farad. Trans. **91**, 431 (1995)
15. J. Zielkiewicz, Phys. Chem. Chem. Phys. **2**, 2925 (2000)
16. P.E. Smith, Biophys. J. **91**, 849 (2006)
17. I.L. Shulgin, E. Ruckenstein, J. Phys. Chem. B **111**, 3990 (2007)
18. S. Shimizu, C.L. Boon, J. Chem. Phys. **121**, 9147 (2004)
19. R. Chitra, P.E. Smith, J. Phys. Chem. B **105**, 11513 (2001)
20. M. Aburi, P.E. Smith, J. Phys. Chem. B **108**, 7382 (2004)
21. B.M. Baynes, B.L. Trout, J. Phys. Chem. B **107**, 14058 (2003)

22. M. Kang, P.E. Smith, Fluid Phase Equilib. **256**, 14 (2007)
23. J.G. Kirkwood, F.P. Buff, J. Chem. Phys. **19**, 774 (1951)
24. P.Y. Chou, G.D. Fasman, Trends Biochem. Sci. **2**, 128 (1977)
25. E. Matteoli, L. Lepori, J. Chem. Phys. **80**, 2856 (1984)
26. R. Chitra, P.E. Smith, J. Phys. Chem. B **106**, 1491 (2002)
27. H.Q. Liu, E. Ruckenstein, J. Phys. Chem. B **102**, 1005 (1998)
28. J. Rosgen, B.M. Pettitt, J. Perkyns, D.W. Bolen, J. Phys. Chem. B **108**, 2048 (2004)
29. J. Rosgen, B.M. Pettitt, D.W. Bolen, Biochemistry. **43**, 14472 (2004)
30. E.R.B. Smith, P.K. Smith, J. Biol. Chem. **117**, 209 (1937)
31. H. Kuramochi, H. Noritomi, D. Hoshino, K. Nagahama, J. Chem. Eng. Data **42**, 470 (1997)
32. E.R.B. Smith, P.K. Smith, J. Biol. Chem. **135**, 273 (1940)
33. J. Zielkiewicz, Int. DATA Ser. Sel. Data Mixtures, Ser. A **28**, 299 (2000)
34. J.J. Savage, R.H. Wood, J. Solut. Chem. **5**, 733 (1976)
35. A.V. Plyasunov, E.L. Shock, R.H. Wood, J. Chem. Eng. Data **48**, 1463 (2003)
36. J. Sedlbauer, P. Jakubu, Ind. Eng. Chem. Res. **47**, 5048 (2008)
37. J. Gmehling, Pure Appl. Chem. **75**, 875 (2003)
38. P.E. Smith, R.M. Mazo, J. Phys. Chem. B **112**, 7875 (2008)
39. K.E. Newman, Chem. Soc. Rev. **23**, 31 (1994)
40. P.G. Kusalik, G.N. Patey, J. Chem. Phys. **86**, 5110 (1987)
41. H.X. Lei, P.E. Smith, Biophys. J. **85**, 3513 (2003)
42. M.L. de la Paz, K. Goldie, J. Zurdo, E. Lacroix, C.M. Dobson, A. Hoenger, L. Serrano, Proc. Natl. Acad. Sci. USA **99**, 16052 (2002)
43. F. Chiti, M. Stefani, N. Taddei, G. Ramponi, C.M. Dobson, Nature **424**, 805 (2003)
44. H. H. Tsai, D. Zanuy, N. Haspel, K. Gunasekaran, B.Y. Ma, C.J. Tsai, R. Nussinov, Biophys. J. **87**, 146 (2004)
45. F. Rousseau, L. Serrano, J.W.H. Schymkowitz, J. Mol. Biol. **355**, 1037 (2006)
46. F. Rousseau, J. Schymkowitz, L. Serrano, Curr. Opin. Struct. Biol. **16**, 118 (2006)