Peptide Clusters

DOI: 10.1002/anie.200802679

Charge-Based Interactions between Peptides Observed as the Dominant Force for Association in Aqueous Solution**

Sylvia E. McLain, Alan K. Soper, Isabella Daidone, Jeremy C. Smith, and Anthony Watts*

The process by which proteins fold in solution into their biologically functional forms is still not well understood despite intense research. The association of hydrophobic amino acid side chains in proteins—the hydrophobic effect—is frequently invoked to be the fundamental driving force behind protein folding in vivo.[1–4] However, there is little direct experimental evidence that supports this assertion, and protein assembly purely from hydrophobic association gives an incomplete picture of the folding process. While many fully folded protein cores contain associated hydrophobic residues, ion pairs or salt bridges are important in stabilizing protein structures, and a proportion of proteins have ion pairs buried in their core.[5] Moreover, the presence of a hydrophobic core does not necessarily implicate hydrophobic forces as the primary driving force of folding.

To gain further understanding of the relative roles of hydrophobic and hydrophilic interactions in the process of protein formation, we determined the structure in aqueous solution of three dipeptide fragments containing both hydrophobic and hydrophilic portions exposed to the surrounding water solvent by using a combination of neutron diffraction and computer simulation techniques. The series of peptides investigated consisted of glycyl-L-alanine, glycyl-L-proline, and L-alanyl-L-proline (Figure 1). The hydrophobicity of these dipeptides increases across the series; glycine has the smallest hydrophobic group (-H), alanine a single methyl group (-CH_3), and proline has the largest hydrophobic group with its pyrrolidine ring (-$CH(N)(CH_2)_3$).[6] Proline was chosen for this investigation as it is both hydrophobic and soluble enough to make the neutron diffraction experiments feasible. Note that the peptide bond in glycyl-L-alanine is a secondary amide, whereas the other two dipeptides are tertiary amides (Figure 1).

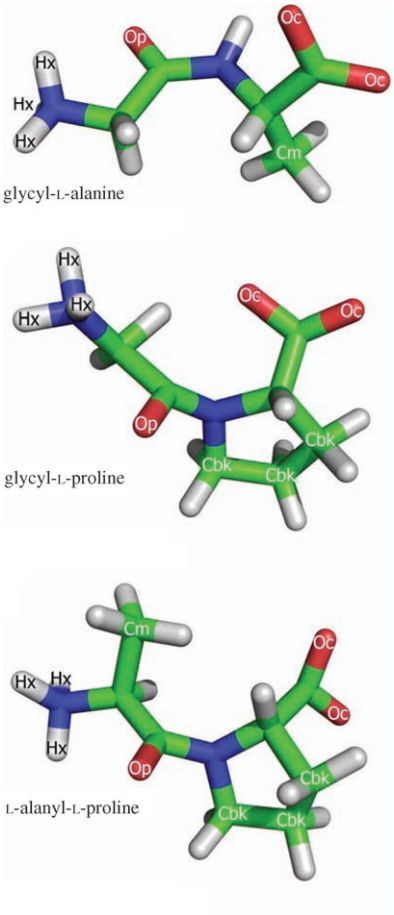

Figure 1. Molecular structure of the dipeptides studied (taken from the EPSR modeling box).

Neutron diffraction enhanced by hydrogen isotope substitution (NDHIS) when combined with computer simulation provides atomic-length-scale information about the arrangement of molecules in solution.[7–11] Through the application of NDHIS coupled with modeling by empirical potential structure refinement (EPSR; see the Experimental Section), it is possible to extract three-dimensional structures of the solution which are consistent with the diffraction experi-

[*] Dr. S. E. McLain
Neutron Scattering Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831 (USA)
Fax: (+1) 865-574-6080
E-mail: mclainse@ornl.gov

A. K. Soper
ISIS Facility, Rutherford Appleton Laboratory
Chilton Didcot, Oxon, OX11 0QX (UK)

I. Daidone, J. C. Smith
Interdisciplinary Center for Scientific Computing
University of Heidelberg, 69120 Heidelberg (Germany)

J. C. Smith
Center for Molecular Biophysics, Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831 (USA)

A. Watts
Biochemistry Department, University of Oxford
Oxford, Oxon, OX1 3QU (UK)

[**] We thank Leighton Coates (Oak Ridge National Laboratory) and Thomas Splettstoesser (University of Heidelberg) for help with graphical representations and Christopher Stanley (Oak Ridge National Laboratory) for useful discussions. JCS was funded by Laboratory Research and Development funds provided by the US Department of Energy.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200802679.

ment.[12,13] Molecular dynamics (MD) simulations[14,15] (see the Experimental Section) were used to assess independently the structure of the three peptide fragments in solution.

The dipeptide molecules were studied separately in their zwitterionic form, with the C and the N termini as CO_2^- and NH_3^+ groups, respectively. Aqueous solutions of the peptides were prepared by using a variety of hydrogen isotope substitutions at a concentration ratio of 1 mol dipeptide to 20 mol water and pH ≈ 6 in each case. The diffraction data and the EPSR-model fits to this data can be found in the Supporting Information. Figure 2 shows the atom–atom

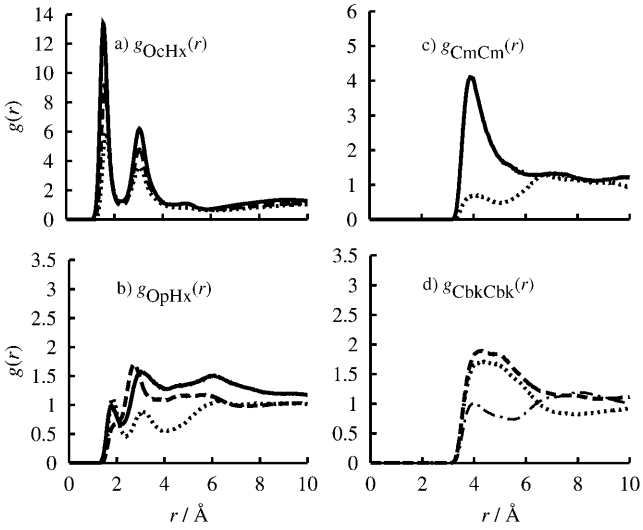

Figure 2. Site–site radial distribution functions ($g(r)$) for peptides in aqueous solutions. The solid lines correspond to glycyl-L-alanine, the dashed lines to glycyl-L-proline, and the dotted lines to L-alanyl-L-proline. The charge–charge interactions for the dipeptides are a) the $g_{OcHx}(r)$ interaction between the C-terminal oxygen atoms (CO_2^- group) and N-terminal hydrogen atoms (NH_3^+ group) on the peptides and b) the $g_{OpHx}(r)$ interaction between the C=O group of the peptide bond and the N-terminal hydrogen atoms (NH_3^+) on the peptides. The other two functions show c) the hydrophobic methyl–methyl interactions ($g_{CmCm}(r)$) for the solutions of glycyl-L-alanine and L-alanyl-L-proline and d) the pyrrolidine ring carbon–carbon interactions ($g_{CbkCbk}(r)$) for the solutions of glycyl-L-proline and L-alanyl-L-proline. The dot-dashed line in (d) corresponds to the methyl–ring interactions ($g_{CbkCm}(r)$) in L-alanyl-L-proline.

correlations ($g(r)$) from the experimental EPSR-model fits. These correlations correspond to both the hydrophobic and "hydrophilic" or charge–charge interactions between the peptides in solution. Specifically, $g_{OcHx}(r)$ and $g_{OpHx}(r)$ are the "hydrophilic" interactions between peptides. They correspond to C-terminal oxygen atom (Oc)–N-terminal hydrogen atom (Hx) contacts and peptide bond oxygen atom (Op)–N-terminal hydrogen atom (Hx) contacts, respectively. The hydrophobic interactions, namely, the methyl–methyl ($g_{CmCm}(r)$), ring–ring ($g_{CbkCbk}(r)$), and methyl–ring ($g_{CbkCm}(r)$) interactions, are also shown in Figure 2.

The maxima of the $g(r)$ curves in Figure 2 show that the most pronounced interactions are found between the glycyl-L-alanine peptides, as is evident in both the "hydrophilic" group contacts ($g_{OcHx}(r)$; Figure 2a) and hydrophobic group contacts ($g_{CmCm}(r)$; Figure 2c). Both of the more hydrophobic peptides, glycyl-L-proline and L-alanyl-L-proline, show less pronounced interactions between the atomic sites. In all of the peptides, the most prominent correlation is between the NH_3^+ and CO_2^- end groups on the peptide fragments, rather than between hydrophobic groups. A substantial number of NH_3^+ and CO_2^- contacts occur in each case, as shown by the coordination numbers for these functions at 2.43 Å, which is the first minimum of $g_{OcHx}(r)$, of 0.75, 0.57, and 0.44 for glycyl-L-alanine, glycyl-L-proline, and L-alanyl-L-proline, respectively.

The question arises as to whether the dominant pairwise hydrophilic charged interactions also drive clustering of the peptide molecules, or rather, if any clustering is driven by hydrophobic interactions. To determine whether any large-scale association occurs, we analyzed both the experimentally derived EPSR model and MD simulations by using the same clustering criteria (Figure 3; see the Experimental Section).

As reflected by the $g(r)$ curves in Figure 2, the most prominent clustering interaction for all of the peptides in the series is through contacts between the NH_3^+ and CO_2^- groups. In the case of the glycyl-L-alanine peptide, the most probable NH_3^+–CO_2^- cluster, with a probability of approximately 90% in the experimentally derived EPSR model, is a fully percolating cluster containing 50 peptides. The MD simulation of the same peptide also shows this trend towards larger

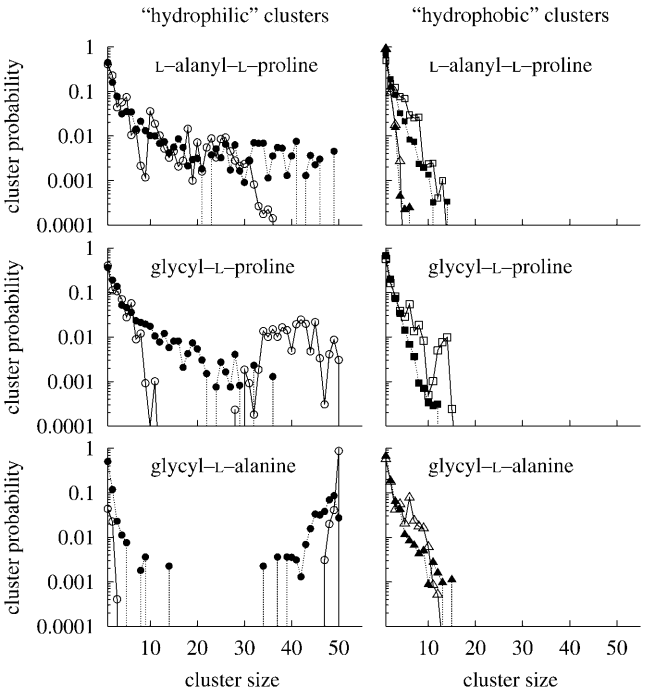

Figure 3. Cluster analysis for peptides in solution. The cluster probability has been normalized with respect to each type of cluster and is depicted on a logarithmic scale. The left-hand graphs show the NH_3^+–CO_2^- clusters (results derived from ○ the EPSR model, ● MD simulations). The right-hand graphs depict the hydrophobic interactions between peptides (methyl–methyl interactions: △ results derived from the EPSR model, ▲ results derived from MD simulations; proline pyrrolidine ring–ring interactions: □ results derived from the EPSR model, ■ results derived from MD simulations).

clusters of 40–50 NH_3^+–CO_2^--linked peptides, which have a collective probability of approximately 40%. Although there are unassociated single glycyl-L-alanine molecules (cluster size 1) and dimers present, the occurrence of intermediate-sized NH_3^+–CO_2^- clusters (about 5–35 peptides) has a very low probability in the MD simulation; in the experimentally derived EPSR model these intermediate-sized clusters are completely absent from the solutions of glycyl-L-alanine. In contrast, the most hydrophobic peptide, L-alanyl-L-proline, predominantly forms small-to-intermediate-sized NH_3^+–CO_2^- clusters (2–25 peptides) with a steadily decreasing probability of larger-sized clusters of this type in the EPSR modeling box. A similar trend was observed in the L-alanyl-L-proline MD simulations, although with a higher probability of NH_3^+–CO_2^- peptide association into larger clusters (25–50 peptides).

The probability distribution of the NH_3^+–CO_2^- glycyl-L-proline clusters is "intermediate" between the distributions of the NH_3^+–CO_2^- clusters of the other two peptides in the EPSR model and shows a similar likelihood of formation of either small clusters (less than 10 peptides) or larger clusters (30–50 peptides). The glycyl-L-proline NH_3^+–CO_2^- clusters derived from the MD simulation show a high probability of intermediate clusters of 2–25 peptides. Although the results derived from EPSR and MD models for hydrophilically linked NH_3^+–CO_2^- clusters between L-alanyl-L-proline peptides, and to a larger extent between glycyl-L-proline peptides, differ somewhat, it is clear from both approaches that the least hydrophobic peptide, glycyl-L-alanine, shows the greatest NH_3^+–CO_2^- association, whereas the two more hydrophobic peptides show a much lower probability of forming larger NH_3^+–CO_2^- clusters.

In Figure 3, only the results of the NH_3^+–CO_2^- clustering analysis are shown in the graphs on the left-hand side. The other potential site for charge-based hydrophilic clustering is through a peptide bond (C=O) oxygen atom in an interaction with the N-terminal NH_3^+ group. However, all peptides in both the EPSR models and MD simulations show a very low probability of clustering of this type, and these clusters have therefore been omitted.

For all three peptides, the hydrophobic clusters (formed through methyl–methyl or pyrrolidine ring–ring interactions) are much smaller than the NH_3^+–CO_2^--linked clusters. Moreover, the largest hydrophobic peptide cluster for any of the measured peptides only contains about 12–15 peptides, and the probability of larger-scale hydrophobic clusters decreases rapidly with increasing cluster size. Glycyl-L-alanine shows a higher degree of methyl–methyl clustering, with a maximum cluster size of around 15 peptides, than L-alanyl-L-proline, which exhibits a maximum cluster size of around 3–5 peptides, consistent with the $g_{CmCm}(r)$ curves (Figure 2d). Both glycyl-L-proline and L-alanyl-L-proline show approximately the same distribution of ring–ring-based clusters, whereby larger-sized clusters of this type (above about 10 molecules) have very low probabilities. Strikingly, even though hydrophobicity increases across the series of peptides, there is no increase in hydrophobically driven large-scale clustering with increasing hydrophobic surface area (Figure 3). Additional simulations performed on the same systems after removing all charges so that no hydrophilic charge-based clusters could occur (see the Supporting Information) indicated that the hydrophobic clusters in Figure 3 are in fact not significantly more prevalent than the hydrophobic clusters that occur in a randomly packed solution.

Experimentally derived EPSR models and MD simulations demonstrate that, of the peptides investigated in this study, glycyl-L-alanine shows the most association between peptides in solution, and the two peptides with the larger hydrophobic surface area, glycyl-L-proline and L-alanyl-L-proline, show the least association (Figures 2 and 3). Moreover, the clustering profiles and the site-specific $g(r)$ values indicate that the most prevalent interaction between any of the peptides in solution is charge-based or hydrophilic. A typical snapshot of the MD simulation box for glycyl-L-alanine shows the aggregation of these peptides in solution (Figure 4).

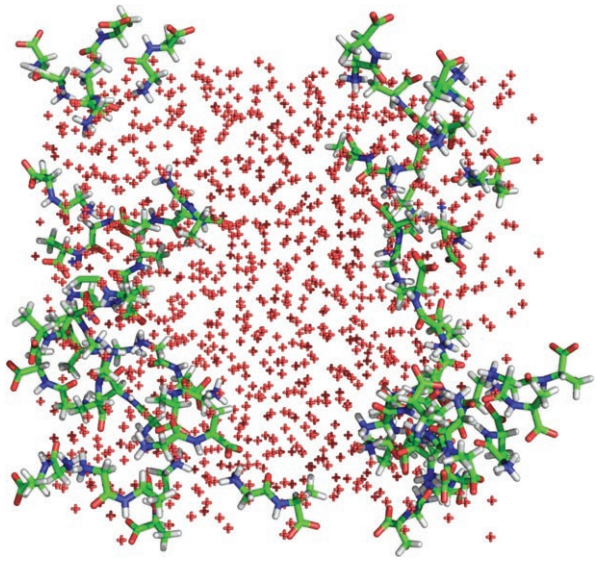

Figure 4. Representative snapshot of the peptide/water box from MD simulations on glycyl-L-alanine in solution. The water molecules are depicted by the red crosses with the hydrogen atoms eliminated for clarity.

The present findings indicate that the charged sites on the peptides dominate the formation of their structures in solution. Moreover, the interactions between peptides decrease as the hydrophobicity of the peptides increase. A possible explanation for this phenomenon might have been that the hydrophobic and hydrophilic interactions compete, so that with increasing hydrophobicity there is a "breakdown" of the charged interactions. However, such an effect does not appear to apply, given that the least hydrophobic peptide of those studied, glycyl-L-alanine, showed the greatest number of hydrophobic interactions. This trend is opposite to that which would be predicted by a model of peptide association in which hydrophobicity is the dominant structural driving force.

These results have significant implications for theoretical models concerning the protein-folding process. The present experimental and computer-simulation results furnish quan-

titative structural information on an atomic length scale with respect to the relative importance of charged interactions between the hydrophilic groups and nonpolar hydrophobic interactions in aqueous solution. These results indicate that hydrophilic interactions are the dominant driving force in peptide association and possibly, by extension, in the protein-folding process.

Experimental Section

Neutron diffraction measurements were performed on the peptide solutions (see the Supporting Information) with the SANDALS diffractometer at ISIS in the UK. The neutron diffraction data were used as constraints for the EPSR models of the peptide solutions. The modeling boxes were fixed at the appropriate experimental density and temperature and contained 50 peptides and 1000 water molecules in each case. The seed potentials for the EPSR models were modified slightly from the potentials used in the MD simulation (for details, see the Supporting Information).

The MD simulations of the three peptides were performed with the GROMACS software package by using the AA-OPLS force field for the peptide and the SPC/E model for water in an NVT ensemble at 300 K with isokinetic temperature coupling.[16] The contents of the MD simulation boxes and EPSR modeling boxes were identical; for both, periodic boundary conditions were used. Electrostatic interactions were treated by using the particle mesh Ewald method (real-space cutoff: 0.9 nm).[17] The bond lengths were fixed,[18] and a time step of 2 fs was used for numerical integration. Simulations were performed for 1 μs with coordinates stored every 1 ps.

The ensemble-averaged site–site radial distribution functions ($g(r)$) and the average clustering profile were calculated from the EPSR and MD molecular assemblies. The $g(r)$ curves can be integrated to obtain the coordination number (n) of atom β around atom α over the distance range r_1–r_2 according to Equation (1):

$$n_{\alpha}^{\beta}(r) = 4\pi c_{\beta}\rho \int_{r_1}^{r_2} g(r) r^2 \, \mathrm{d}r \qquad (1)$$

Cluster analysis is performed by considering that two molecules belong to the same cluster if specified atoms are within a given distance range, which is usually defined by the position of the first minimum between the atoms observed in the appropriate $g(r)$ function. The size of a cluster is determined by counting all of the molecules that are connected to at least one other molecule in the same cluster within the specified distance constraint. Further details are given in the Supporting Information.

Received: June 6, 2008
Revised: August 11, 2008
Published online: October 20, 2008

Keywords: charge–charge interactions · hydrophobic effects · molecular dynamics · neutron diffraction · peptide association

[1] D. Chandler, *Nature* **2005**, *437*, 640 – 647.
[2] D. Chandler, *Nature* **2007**, *445*, 831 – 832.
[3] M. Mezger, H. Reichert, S. Schoder, J. Okasinski, H. Schroder, H. Dosch, D. Palms, J. Ralston, V. Honkimaki, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18401 – 18404.
[4] K. A. Dill, *Biochemistry* **1990**, *29*, 7133 – 7155.
[5] D. J. Barlow, J. M. Thornton, *J. Mol. Biol.* **1983**, *168*, 867 – 885.
[6] P. A. Karplus, *Protein Sci.* **1997**, *6*, 1302 – 1307.
[7] E. C. Hulme, A. K. Soper, S. E. McLain, J. L. Finney, *Biophys. J.* **2006**, *91*, 2371 – 2380.
[8] S. E. McLain, A. K. Soper, A. E. Terry, A. Watts, *J. Phys. Chem. B* **2007**, *111*, 4568 – 4580.
[9] S. E. McLain, A. K. Soper, A. Watts, *J. Phys. Chem. B* **2006**, *110*, 21251 – 21258.
[10] P. E. Mason, G. W. Neilson, J. E. Enderby, M. L. Saboungi, C. E. Dempsey, A. D. MacKerell, J. W. Brady, *J. Am. Chem. Soc.* **2004**, *126*, 11462 – 11470.
[11] Y. Kameda, M. Sasaki, Y. Amo, T. Usuki, *Bull. Chem. Soc. Jpn.* **2007**, *80*, 1746 – 1749.
[12] A. K. Soper, *Mol. Phys.* **2001**, *99*, 1503 – 1516.
[13] A. K. Soper, *Phys. Rev. B* **2005**, *72*, 104204.
[14] I. Daidone, M. Ulmschneider, A. Di Nola, A. Amadei, J. C. Smith, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15230 – 15235.
[15] I. Daidone, M. D'Abramo, A. Di Nola, A. Amadei, *J. Am. Chem. Soc.* **2005**, *127*, 14825 – 14832.
[16] D. Brown, J. H. R. Clarke, *Mol. Phys.* **1984**, *51*, 1243 – 1252.
[17] D. M. York, T. A. Darden, L. G. Pedersen, *J. Chem. Phys.* **1993**, *99*, 8345 – 8348.
[18] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463 – 1472.

