Molecular Design of Inorganic-Binding Polypeptides

John Spencer Evans, Ram Samudrala, Tiffany R. Walsh, Ersin Emre Oren, and Candan Tamerler

Abstract

Controlled binding and assembly of peptides onto inorganic substrates is at the core of bionanotechnology and biological-materials engineering. Peptides offer several unique advantages for developing future inorganic materials and systems. First, engineered polypeptides can *molecularly recognize* inorganic surfaces that are distinguishable by shape, crystallography, mineralogy, and chemistry. Second, polypeptides are capable of *self-assembly* on specific material surfaces leading to addressable molecular architectures. Finally, genetically engineered peptides offer multiple strategies for their functional modification. In this article, we summarize the details and mechanisms involved in combinatorial-polypeptide sequence selection and inorganic-material recognition and affinity, and outline experimental and theoretical approaches and concepts that will help advance this emerging field.

Introduction

During the last two decades, combinatorial-peptide selection methods[1–3] have been used to generate sequence libraries that recognize and bind to different inorganic solids.[4–12] Surface-exposed or displayed polypeptides produced by phage[5,7–10,12] and bacteria[4,6,10,11] have become the predominant *in vivo* techniques for material-specific peptide selection, and inorganic-binding peptides are quickly becoming molecular tools for biotechnological and nanotechnological applications. With 20 naturally occurring amino acids available for use, biological organisms can craft an extremely large and diverse set of linear sequences for a wide range of materials, including metals, oxides, semiconductors, and minerals. In addition to the numerous linear combinations, the potential two- and three-dimensional configurations of these sequences add another dimension in that there are a number of polypeptide backbone geometries (i.e., secondary structures) that form from different amino-acid sequences of these peptides. Hence, factors such as polypeptide–material affinity and selectivity ultimately are chosen by the sequence and its molecular architecture as well as the chemical composition of the peptide.[13,14] Thus, a successful design of polypeptide-inorganic materials is dependent upon our understanding of the molecular factors that govern sequence–structure–function selection.

This article will summarize our current knowledge of phage and bacterial-generated polypeptides directed against inorganic solids, using examples obtained from experiment and theory to define the molecular trends emerging from the screened polypeptide libraries generated against artificial and biological inorganic materials such as Pt, Au, hydroxyapatite, graphite, and quartz.

First-Generation Peptides and Post-Selection Engineering

A genetically engineered polypeptide for inorganics (GEPI) is defined as an amino-acid sequence that specifically and selectively binds to an inorganic surface.[10] Bacterial-cell surface (BCS) and phage display (PD) libraries have been adapted to select for a variety of GEPIs.[5–12] Typically, these libraries are generated by artificially inserting randomized nucleotides within genes specifying cell-surface or phage coat proteins. The host library (which typically consists of 10^9–10^{11} different members of either cells or phage) then is exposed to the desired substrate. The displayed surface-coat polypeptides on the hosts have different sequences that come in contact with the inorganic surfaces. Mild chemical-elution conditions remove weak or nonspecific binders, and strongly binding cells or viruses are recovered.[15] This biopanning cycle is repeated a number of times to enrich specificity and high-affinity binders. Eventually, the amino-acid sequences of the inorganic-binding regions are deduced by DNA sequencing and cataloged.

Once the first set of binding peptides is obtained, their affinity and specificity can be further "tuned" via the use of various molecular-tailoring strategies. For example, binding properties of a selected polypeptide can be tweaked by site-specific change(s) of amino acids within the sequence. Molecular constraints and the use of multiple sequence repeats (i.e., multimerization) can also be used to tune the binding properties and consequently the structural features of the initial polypeptide sequence.[6,14] As an example, we found that the cyclic form of a Pt binding sequence exhibits a higher affinity than the linear version using surface plasmon resonance (SPR) spectroscopy (Figure 1).[14] In the case of multimerization, we demonstrated that the affinity and selectivity for given inorganic materials improved as a function of the repeat number of the polypeptide sequence (e.g., 3-versus 1-repeat gold binding peptides).[16] Hence, the initial sequences serve as a starting point for further improvement or modification.

Although the combinatorial biology techniques are relatively straightforward, there are several important considerations that need to be addressed when combining inorganic materials with biological agents. First, the method for separating material-bound hosts from unbound ones may disqualify a particular display technology. For instance, phage particles are limited in size and thus are suitable for work with inorganic powders and enrichment by centrifugation separation techniques.[15] The bacterial flagellin cell-surface-display system would not be amenable to this enrichment process since centrifugal forces would shear off the long flagella or tail from the bacterial cell.[11] Second, the chemical and physical states of the inorganic surface itself may affect the efficiency of polypeptide binding. For example, many materials rapidly develop a surface-oxide layer when exposed to air or solution or may become modified when incubated in

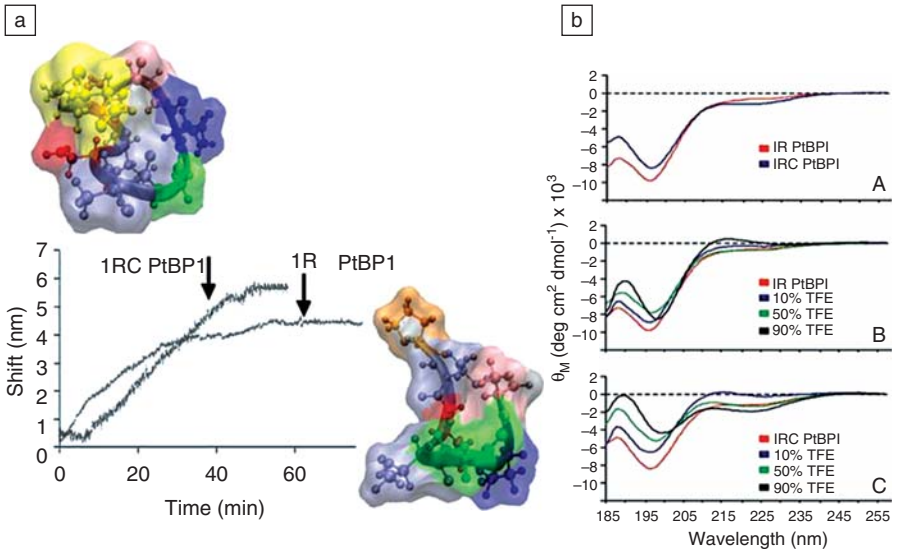

Figure 1. (a) Adsorption isotherms of two forms of the platinum binding peptide: 1 repeat linear (1R-PtBP1) and 1 repeat constraint (1RC-PtBP1). Pseudo-3D views show their molecular architectures. The integral sequence in both peptides is CPTSTGQAC. The amino acids are colored according to their chemical properties (hydrophobic = blue; acidic = yellow; basic = green; polar = red). (b) Circular-dichroism spectra of 30 micromolar 1R-PtBP1 and 1RC-PtBP1 peptides at pH 7.5 and in the presence of varying volume percentages of 2,2,2-trifluoroethanol (TFE) at the same pH. Here, circular dichroism measures the rotation of plane-polarized light by carbonyl chemical groups in the polypeptide backbone, and the rotation correlates with the peptide backbone geometry (i.e., RC, alpha-helix, beta-sheet, etc.). TFE is a structure-stabilizing solvent that helps polypeptides fold into more compact conformations. (c) In the presence of varying volume percentages of TFE at pH 7.5

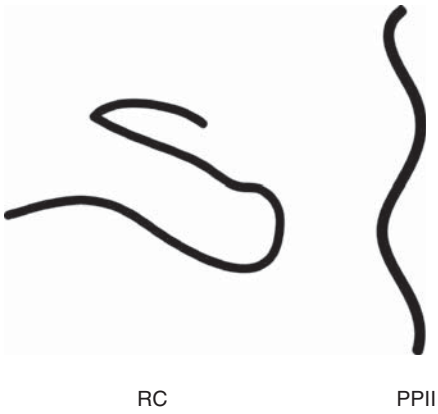

Figure 2. Cartoon representation of random-coil (RC) and polyproline Type II (PPII) polypeptide backbone chains. Chain length is arbitrary.

the biological media used during the panning process. Thus, it is important to characterize inorganic surfaces prior to and after panning procedures to determine if any alterations have taken place. Third, the compatibility of inorganic materials with biological buffers may need to be addressed. One may need to monitor the effect of wash or elution buffers on inorganic-surface integrity and optimize parameters to guard against surface modification, etching, or other forms of surface deterioration. Fourth, inorganic compounds come in a variety of forms, from polydisperse powders to single crystals. With diverse interfacial features available on different surfaces of the same solid, peptides may reconform to recognize different surface features. Thus, a different binding sequence could emerge depending on the nature of the surface topology. Finally, our expectations regarding sequence library convergence need to be re-examined when we use inorganic materials for selection. In traditional biological applications of peptide libraries, at the end of three to five biopanning cycles, the selected sequences typically converge toward a consensus consisting of identical sequences. However, this rule does not apply in the case of inorganic-binding sequences where similarities, rather than a strict consensus, are generally observed. This presumably reflects the heterogeneity of the inorganic substrate at the atomic, crystallographic, and morphological levels, as well as other, perhaps chemical, factors.[15]

Molecular Structures: Experimental Perspective

Polypeptide structure influences function, and therein lies the challenge for GEPI research. Although it is relatively easy to generate sequence libraries against a given material, it is not so straightforward to wade through this expansive library, determine the individual polypeptide structures of this ensemble, and examine how these structures relate to function. Recent experimental studies that have been carried out with the GEPIs selected for a wide variety of materials such as Au,[4,16,17] Pt,[15,18] carbon nanohorns,[19] and hydroxyapatite[20] give us a glimpse of what may be general structural "rules" that exist within peptide sequences. The first trend is that both M13 phage pIII (7 or 12 amino acids (AA))[8–10] and bacterial-cell receptor-generated[11] polypeptides (14 AA, 42 AA) exhibit unfolded conformations that fall within two classifications (Figure 2). The first classification is predominantly random-coil (RC) structures in equilibrium with other secondary structures such as alpha-helix, beta-strand, and beta-turn.[10,11] The second classification is polyproline Type II (PPII), an extended helical secondary structure common to sequences containing Pro, Ala, Gln, and other PPII-forming amino acids.[14,20] This secondary structure is believed to exist in equilibrium with RC conformation but not with alpha helix or beta strand.

What is the significance of unfolded structures that exist either as RC or PPII in material-selected polypeptide sequences? To answer this question, there are two hypotheses to consider. The first consideration is that both structures allow side-chain accessibility to the solvent and interfacial environments.[14,21] This means that potential peptide–material interface interactions would be expected to be maximal for an unfolded polypeptide as opposed to a folded peptide that, due to internal contacts and folded topology, can offer only limited surface(s) for interaction. This phenomenon is also observed in biomineral-associated polypeptides.[22–24] The second consideration is that labile, unfolded conformations are potentially better at adapting to irregular surface topologies at an inorganic interface than a polypeptide with an internal structure that is already stabilized and fixed.[21–24] Thus, focusing on unstructured sequences appears to be the approach that nature uses for selection with inorganic materials.

In addition to these structural considerations, there is also the fact that M13 phage pIII sequences are expressed in two geometric configurations: either as the 7 AA or 12 AA linear form or as a 7 AA cyclic (loop-constrained) form.[14,20]

Side-by-side comparison of the cyclic and linear versions of the same 7AA Pt binding sequence showed that the cyclic version exhibited higher affinity for Pt and also required an additional step in the binding process.[14] Further analysis revealed that, under certain conditions, the cyclic version can fold but the linear version cannot.[14] This explains why a second step in the binding process was essential for the cyclic form but not the linear form: the cyclic form needed an additional step for folding before final adsorption could take place. Hence, the closure of a polypeptide sequence into a loop configuration changes the structure and behavior of the integral sequence. These features may be useful in designing higher-affinity polypeptides for material interaction.

Molecular Structure and Interactions with Materials: Theoretical Considerations

In tandem with experimental approaches, simulation can be a useful complementary tool in the characterization of the interface of the polypeptide-inorganic solid. Simulation is crucial to unraveling the complex interplay of sequence, structure, and function that determines the binding affinity and specificity at these complex interfaces. Here, we invoke the unifying concept of the energy landscape where stable low-energy states of the system are characterized by minima on the landscape.[25] The framework of energy landscapes has been successfully used to resolve a range of physical/chemical problems from protein folding to understanding glassy materials.[26] For the polypeptide-inorganic interface, this landscape will be a function of the peptide conformation, overall position, and orientation of the peptide on the surface. There are two aspects to this problem: first, generating the energy landscape, and second, exploring representative regions of this landscape.

For the first aspect, we must describe the chemical and physical interactions among components in the system. The quality and appropriateness of the description of these interactions is a vital ingredient in these calculations. The parameters of some intermolecular potentials are available for describing biomolecules (for example, CHARMM[27] or AMBER[28]) with a similar array for describing inorganic surfaces (such as for titania[29] and silica[30]), but very few potentials have been specifically designed to simultaneously describe both entities. Therefore, care and caution must be exercised to ensure that the balance of intermolecular interactions is properly described.[31] The construction of "tailored" potentials is expected to be a key development in this field.

For the second aspect, we need a method for exploring the energy landscape and finding the low-energy minima that will correspond with polypeptide-inorganic-binding configurations (Figure 3). There is a growing body of evidence suggesting that the polypeptide-inorganic interface typically supports many different strong-binding configurations on the energy landscape.[32] Typically, Monte Carlo[33,34] or molecular-dynamics[32,35] approaches are employed to estimate thermodynamic quantities such as free-energy differences needed for interpreting binding affinity. While no one approach can guarantee an exhaustive survey of the landscape, further developments in sampling are of a prime necessity. By seeing how these various minima are distributed on the landscape, we can use statistical mechanics to interpret and predict the behavior of polypeptide–solid interfaces.

Recently some successful modeling studies have been emerging. To date, atomistic studies have emphasized structural and energetic data in an attempt to explain peptide-binding affinity.[14,16,18,22,32] One issue yet to be rigorously addressed is the change in free energy upon mutation of a polypeptide. Mutation of "key" binding residue(s) in a sequence would not only remove a significant molecule/surface binding interaction but could also significantly alter the secondary structure of the peptide, giving rise to nonlocal effects that may play a crucial role in solid recognition. In partnership with SPR (energetic) and nuclear magnetic resonance (NMR) (structural) experiments, calculations of the change in binding free energy upon mutation will help to unravel these simultaneous contributions from residue–surface contact and secondary structure changes.

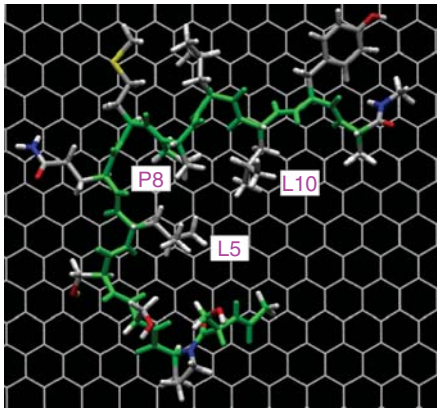

Figure 3. Snapshot from a molecular-dynamics simulation of a peptide bound on a graphite (0001) surface. This configuration shows two features that govern the secondary structure of the peptide. Intra-peptide interactions are visible by the close proximity of the leucine (L5), proline (P8), and leucine (L10) residues at the 5th, 8th, and 10th positions, respectively, along the peptide chain. Residue-surface contact is visible at the 11th position where the tyrosine ring is flat on the surface in near alignment with an underlying graphite ring.

Finally, we need to address two additional issues. First, we need to develop tools to explain and predict *specificity* from a modeling and simulation viewpoint. The energy landscape for a given peptide on different material interfaces could provide a unifying framework to address this question. However, surveying numerous energy landscapes remains a nontrivial task. Second, recent reports indicate that the presence of water, structuring at hydrophilic inorganic surfaces, may play a significant role in mediating peptide-inorganic recognition process(es).[36] However, inclusion of explicit solvent in simulations naturally leads to increased computational cost. It remains unclear how implicit solvation approaches can accommodate solvent-structuring effects. Coarse-graining of the solvent would be one tractable way to address the solvation problem.[36]

In Silico Design of Inorganic-Binding Peptides

While developments toward understanding the nature of peptide–solid interaction are progressing, which will eventually provide us with the fundamental tools for the design of robust peptides with predictable functions (i.e., binding), we have initiated a novel knowledge-based approach to enhancing binding and designing new inorganic-binding peptides.[37] In nature, functionally similar proteins usually exhibit some degree of sequence similarity.[38] By extension, GEPI inorganic-binding peptides that recognize the same inorganic material should also exhibit some sequence similarities as well. If this is true, then it is possible to design new inorganic-binding peptides derived from the characteristics "similar" to our existing inorganic material-specific peptide sequences.[37] We have combined sequence alignment techniques[39,40] and produced unique, material-specific scoring matrices.[37] Using these, we then developed a bioinformatics method that allows the design of new peptides exhibiting enhanced affinities

and specificities to inorganic materials. We will now describe this approach.

The test case for developing this GEPI-based bioinformatics approach involved a library of quartz-binding peptides[37] that were grouped into three sequence clusters (i.e., strong, moderate, and weak) according to their material-binding affinities (Figure 4).[37] Next, focusing on the known strong-affinity quartz-binding sequences, we created a quartz-specific sequence scoring matrix (QUARTZ 1)[37] that would account for the specific sequence patterns responsible for quartz binding as opposed to sequence patterns found in nature.

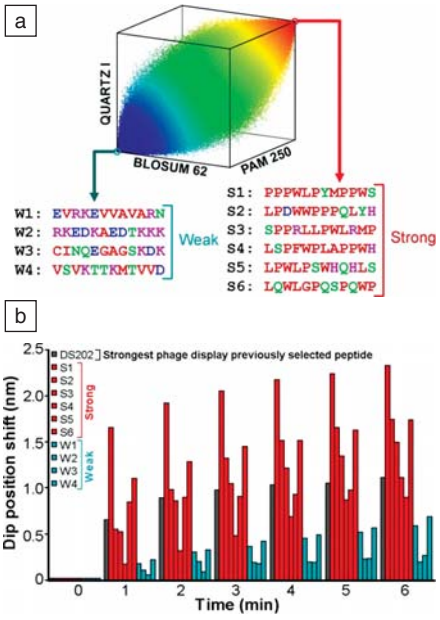

Figure 4. (a) Correlation between the similarity scores of random peptide sequences to strong quartz binders calculated using three different scoring matrices (QUARTZ I, BLOSUM 62, and PAM 250). The rainbow colors indicate the increasing similarity scores (from blue to red) along the diagonal of the cube. The listed sequences of designed six strong (S) and four weak (W) peptides were chosen among the top and the bottom scored ones consistent with all three different scoring matrices used. The amino acids are colored according to their chemical properties (hydrophobic = red; acidic = blue; basic = purple; polar = green). (b) Experimental validation of computationally designed peptides using surface plasmon resonance spectral analysis that measures the amount of bound peptide versus time (performed at 4 μM peptide concentration) for six strong (red) and four weak (teal) peptides along with DS202 (black), the strongest phage display selected peptide (adapted from Reference 37).

From this matrix, we randomly generated new quartz-binding sequences. Then we chose a final set of six strong (high-scored) and four weak (low-scored) predicted quartz binders and tested their binding affinities to quartz (Figure 4). Consistent with our predictions, the experimental affinities of both the high- and low-affinity quartz-binding sequences were consistent with our expected binding behavior (Figure 4). Hence, we can use the existing knowledge of inorganic-binding sequences to generate a new set of inorganic material-specific polypeptide sequences.

The attractiveness of the bioinformatics approach is that it is general and could be used to design novel peptides with any functional property (in addition to binding) as a utility in a wide range of applications in materials science, nanotechnology, biology, and medicine. For example, using different material-specific matrices, one can design peptides that have multiple inorganic-material-recognition and binding functionalities. Assuming that initial sequence data exist, this procedure could be applied to any given inorganic material. Alternatively, one can utilize the bioinformatics protocol to explore and better understand the very model system that inspired the development of GEPIs (i.e., biomineralization proteins). In this case, GEPI sequences that are specific for hydroxyapatite could be used as a database for identifying mineral-binding regions within the sequences of poorly understood biomineralization proteins such as those involved in tooth and bone formation. Such information could then be used to understand protein function in these medically important hard tissues and to eventually develop hard-tissue regeneration or engineering applications.

Conclusions

Materials science has entered a new paradigm in the design and synthesis of novel inorganics (i.e., peptide-based materials and systems).[5–16] At the same time, materials science and engineering is also in a complex and poorly understood realm involving biological macromolecules and artificial materials. Although we can make complexes of these two disparate entities and utilize them successfully, we do not yet fully understand how or why this occurs. This article underscores the importance of grasping some important details and mechanisms involved in polypeptide–inorganic interactions, points to recent achievements, and encourages further research into this rapidly growing area of materials science that interfaces with biology and the physical sciences.

Acknowledgments

This work was supported by grants from the National Science Foundation through the Genetically Engineered Materials Science & Engineering Center, a Materials Research Science and Engineering Center (RS, EEO, CT, JSE) (DMR-0520567), the U.S. Army Research Office through the Defense University Research Initiative on NanoTechnology program (JSE, EEO, CT), the Turkish State Planning Organization via Advanced Technologies in Engineering (CT), and the Engineering and Physical Sciences Research Council (GR/S80127/01) (TRW). Portions of this work represent contribution number 44 from the Laboratory of Chemical Physics, New York University (JSE).

References

1. G.P. Smith, *Science* **228**, 1315 (1985).
2. K.D. Wittrup, *Curr. Opin. Biotechnol*. **12**, 395 (2001).
3. P. Amstutz, P. Forrer, C. Zahnd, A. Plückthun, *Curr. Opin. Biotechnol*. **12,** 400 (2001).
4. S. Brown, *Nat. Biotechnol*. **15**, 269 (1997).
5. M. Schembri, K. Kjaergaard, P. Klemm, *FEMS Microbiol. Lett*. **170**, 363 (1999).
6. S. Brown, M. Sarikaya, E. Johnson, *J. Mol. Biol*. **299**, 725 (2000).
7. S.R. Whaley, D.S. English, E.L. Hu, P.F. Barbara, A.M. Belcher, *Nature* **405**, 665 (2000).
8. R.R. Naik, S.J. Stringer, G. Agarwal, S.E. Jones, M.O. Stone, *Nat. Mater*. **1**, 169 (2002).
9. C.M. Li, G.D. Botsaris, D.L. Kaplan, *Cryst. Growth Des*. **2**, 387 (2002).
10. M. Sarikaya, C. Tamerler, A.K.Y. Jen, K. Schulten, F. Baneyx, *Nat. Mater*. **2**, 577 (2003).
11. C.K. Thai, H.X. Dai, M.S.R. Sastry, M. Sarikaya, D.T. Schwartz, F. Baneyx, *Biotechnol. Bioeng*. **87**, 129 (2004).
12. M. Umetsu, M. Mizuta, K. Tsumoto, K. Ohara, S. Takami, S. Watanabe, H. Kumagai, I. Adschiri, T. *Adv. Mater*. **17**, 2571 (2005).
13. K.I. Sano, H. Sasaki, K. Shiba, *Langmuir* **21**, 3090 (2005).
14. O.U.S. Seker, B. Wilson, S. Dincer, I.W. Kim, E.E. Oren, J.S. Evans, C. Tamerler, M. Sarikaya, *Langmuir* **23**, 7895 (2007).
15. M. Sarikaya, C. Tamerler, D.T. Schwartz, F. Baneyx, *Ann. Rev. Mat. Res*. **34**, 373 (2004).
16. C. Tamerler, E.E. Oren, M. Duman, E. Venkatasubramanian, M. Sarikaya *Langmuir* **22** (18), 7712 (2006).
17. L. Kulp III, M. Sarikaya, J.S. Evans, *J. Mater. Chem*. **14**, 2325 (2004).
18. E.E. Oren, C. Tamerler, M. Sarikaya, *Nano Lett*. **5**, 415 (2005).
19. J.L. Kulp III, K. Shiba, J.S. Evans, *Langmuir* **21**, 11907 (2005).
20. M. Gungormus, H. Fong, I.W. Kim, J.S. Evans, C. Tamerler, M. Sarikaya, *Biomacromolecules* **9**, 966 (2008).
21. J.S. Evans, *Curr. Opin. Colloid Interface Sci*. **8**, 48 (2003).
22. S. Collino, J.S. Evans, *Biomacromolecules* **8**, 1686 (2007).
23. I.W. Kim, S. Collino, D.E. Morse, J.S. Evans, *Cryst. Growth Des*. **6**, 1078 (2006).

24. S. Collino, I.W. Kim, J.S. Evans, *Cryst. Growth Des*. **6**, 839 (2006).
25. D.J. Wales, M.A. Miller, T.R. Walsh, *Nature* **394**, 758 (1998).
26. D.J. Wales, "Energy Landscapes," Cambridge University Press (Cambridge) (2003).
27. N. Foloppe and A.D. Mackerell, *J. Comput. Chem*., **21**, 86 (2000).
28. J. Wang, P. Cieplak and P.A. Kollman, *J. Comput. Chem*., **21**, 1049 (2000).
29. A.V. Bandura and J.D. Kubicki, *J. Phys. Chem. B*., **107**, 11072 (2003).
30. T.S. Mahadevan and S.H. Garofalini, *J. Phys. Chem. C*, **112**, 1507 (2008).
31. C.L. Freeman, J.H. Harding, D.J. Cooke, J.A. Elliott, J.S. Lardge, D.M. Duffy, *J. Phys. Chem. C* **111**, 11943 (2007).
32. S.D. Tomasio, T.R. Walsh, *Mol. Phys*. **105**, 221 (2007).
33. M. Bachmann and W. Janke, *Phys. Rev. Lett*., **95**, 058102 (2006).
34. M.J. Biggs and M. Mijajlovic, *J. Phys. Chem. C*, **111**, 15839 (2007).
35. V. Carravetta and S. Monti, *J. Phys. Chem. B*, **110**, 6160 (2006).
36. P. Schrevandijk, L.M. Ghiringhelli, L. Delle Site, N.F.A. van der Vegt, *J. Phys. Chem. C* **111**, 2631 (2007).
37. E.E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U.O.S. Seker, M. Sarikaya, R. Samudrala, *Bioinformatics* **23**, 2816 (2007).
38. T.K. Attwood, *Science* **27**, 471 (2000).
39. S.B. Needleman, C.D. Wunsch, *J. Mol. Biol*. **48**, 443 (1970).
40. T.F. Smith, M.S. Waterman, *J. Mol. Biol*. **147**, 195 (1981). □