Review

# Protein oligomerization: How and why

## Mayssam H. Ali and Barbara Imperiali*

*Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

**Abstract**—A large fraction of cellular proteins are oligomeric. Protein oligomerization may often be an advantageous feature from the perspective of protein evolution and has probably evolved by a variety of mechanisms. The study of protein oligomerization may provide insights into the early protein environment and the evolution of modern proteins. Oligomeric mini-proteins, short peptides with discrete protein-like structures, may serve as valuable models for understanding features of protein oligomerization.
© 2005 Elsevier Ltd. All rights reserved.

## Contents

## 1. Introduction

Oligomeric proteins, comprising two or more associating polypeptide chains, represent a significant fraction of cellular proteins. The broad category of oligomeric proteins can be classified by subunit type, strength of subunit association, and duration and avidity of subunit association. Protein oligomerization may be an advantageous feature from the perspective of protein evolution for a number of reasons, including new opportunities for functional control, such as allosteric regulation and the establishment of higher-order complexity. Many early, primitive, proteins may have been homo-oligomeric or hetero-oligomeric to better support function, and thus the study of the nature of protein oligomerization may elucidate features of protein evolution. Inter-subunit interfaces share common features with those of both hydrophobic protein cores and polar protein surfaces. Protein oligomerization has probably evolved by a variety of mechanisms. Recent developments in the design of oligomeric mini-proteins, short peptides with discrete protein-like structures, may serve as valuable models for understanding the details of protein oligomerization.

## 2. Characteristics of oligomeric proteins

Oligomeric proteins abound in nature. They are composed of multiple subunits (polypeptide chains), which

may be the same (a homo-oligomeric protein) or different (a hetero-oligomeric protein). It has been calculated that the average oligomeric state of cellular proteins is tetrameric,[1] and a recent survey suggests that 35% or more of the proteins in a cell are oligomeric.[2] The proportion of oligomeric protein structures deposited in the Protein Data Bank is significantly lower.[3] However, this may simply reflect experimental constraints favoring the structural determination of small monomeric proteins. Most oligomeric proteins are homo-oligomers.[2] Higher-order oligomers are less prevalent[2,3] and a relatively small fraction of oligomeric structures have odd-numbered stoichiometries. Most oligomeric proteins, and essentially all homo-oligomeric proteins, are symmetrical. This symmetry is most frequently cyclic, dihedral, or cubic.[2]

The association between subunits can vary in strength and duration. Some proteins are found only, or primarily, in the oligomeric state. These proteins generally have dissociation constants in the nanomolar range.[4] Others have a weak tendency to associate, with oligomerization dependent on environmental conditions, such as concentration, temperature, and pH. Such proteins often have higher $K_d$ values in the micromolar or even millimolar range.[4] Still other proteins oligomerize dynamically in response to a stimulus, such as a change in nucleotide binding, nucleotide hydrolysis or phosphorylation state. Such a change can have a dramatic effect on the affinity of the subunits for one another, often by orders of magnitude.[4,5]

Monod et al. have characterized homo-oligomeric proteins by mode of interaction (Fig. 1).[6] A homodimeric protein can only have an isologous inter-subunit interaction between the same surface on each monomer, giving rise to a dimer with 2-fold symmetry. A heterologous interaction, with two complementary sites, is possible for higher oligomers. Such an interaction can give rise to a discrete oligomer if the interaction is circularly symmetrical or can lead to indefinite self-association if it is not, as in the case of actin.[6] These classifications continue to provide a valuable framework for thinking about protein interfaces, but may not adequately encompass oligomeric interactions, such as domain swapping.[2]
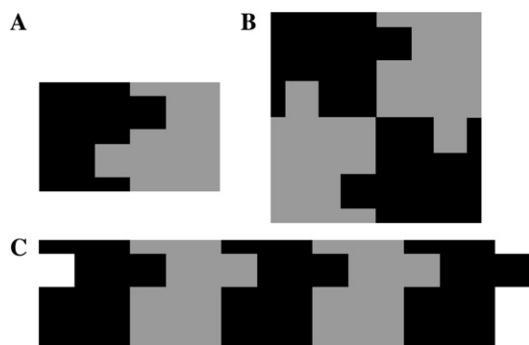
Mini-protein models can provide valuable insights into many attributes of oligomeric proteins. Mini-proteins are short polypeptides that adopt a stable discrete structure in aqueous solution. These systems are typically derived by paring down a natural fold to a more compact size. There is a wealth of oligomeric mini-proteins in the literature, including all-α coiled coils and helical bundles,[7–9] mixed α/β-motifs,[10–12] and all-β motifs.[13–16] Mini-proteins have been extensively used as minimal models for the study of key features of natural proteins, such as the thermodynamic determinants of protein stability[17] and the introduction of catalytic or other functionality.[18–21]

Mini-proteins are ideal for the study of oligomeric proteins for several reasons. (1) Mini-proteins are simple systems, having far fewer variables than do natural proteins, an attribute that is experimentally advantageous as it enables the directed study of a single feature of interest. (2) Mini-proteins, because of their small size, are highly amenable to computational study. By contrast, natural proteins, particularly larger oligomeric proteins, are more complex and computationally demanding to model. (3) Additionally, it is straightforward to design and chemically synthesize a series of mutants with which to rigorously study a feature of interest. Moreover, nonstandard amino acids that probe subtle features not discernible through the standard twenty amino acids can be easily introduced.

## 3. Characteristics of oligomeric interfaces

The specific characteristics of an oligomeric protein interface depend on the nature of that interface and on the duration of the interaction defining it. However, some general features may be described. Interfacial residues tend to protrude from the surface of the protein,[3] and the interaction surface tends to be circular in shape.[22,23] Protein–protein interaction interfaces are relatively planar,[24] as are many hetero-oligomer interfaces.[3,25] By contrast, many obligate homodimers and heterodimers have intertwined monomer units, and thus less planar interfaces.[3,25] As might be predicted, large oligomeric interfaces are often associated with strong interactions. However, a small oligomeric interface can be manifest in both weak and strong associations.[4] The buried surface area in obligate homodimeric proteins is usually greater than 1400 Å$^2$.[26,27] In nonobligate complexes, the interface buried surface area is usually less than 2500 Å$^2$, whereas for weak and transient associations the buried surface area of the interface is less than 1000 Å$^2$.[4]

In general, the residues at an oligomeric interface may be slightly more conserved than other surface residues, although the extent and significance of this preference has been questioned.[4,28–30] It has been found that certain conserved residues, or 'hot spots', generally at the center of an interface, are responsible for most of the binding energy of an oligomeric interaction. These hot spots are often composed of polar residues that engage in van der Waals contacts and hydrogen bonding.[31–33]



**Figure 1.** (A) Isologous dimer; (B) heterologous tetramer; (C) heterologous polymer. Figure adapted from Ref. 6.

Hydrophobic interactions play an important role in defining homo-oligomeric interfaces. About two-thirds of the residues in inter-subunit interfaces are nonpolar[27,34] a much larger fraction than in noninteracting surface residues. However, inter-subunit interfaces are less nonpolar, and have a greater proportion of hydrophilic and polar residues, than a typical protein hydrophobic core.[35] Approximately one-fifth of the residues at oligomeric interfaces are polar, a greater proportion than is found in buried hydrophobic cores.[34] The interaction surfaces of transient, or 'on/off,' hetero-oligomers and weakly associating homo-oligomers are less nonpolar than those of tightly bound oligomers.[3,4,25] In this case, the greater surface polarity may help to solubilize the unassociated state of the individual subunits. Hydrogen bonds and salt bridges are important for the stabilization of oligomeric interfaces, as suggested by the prevalence of polar hot spot residues. Early studies suggested that there is about one hydrogen bond per 200 $\text{Å}^2$ of subunit interface.[27] Charged hydrogen bonds can often be found at subunit interfaces, but such interactions are less common than in the core.[34]

Oligomeric interfaces often have significant electrostatic and geometrical shape complementarity that gives rise to the specificity of the interaction.[23,36,37] Nonetheless, most interaction surfaces are not entirely complementary. Hubbard and Argos have reported that oligomeric proteins often have one or more cavities at the inter-subunit interface; most of these cavities are filled with water.[38]

Mini-proteins have been used to examine the general features of oligomeric interfaces. Earlier studies have revealed that simple hydrophobic patterning is insufficient to encode a specific and well-packed inter-subunit interface.[9] The role of buried polar residues has been extensively researched using coiled coil mini-protein models, revealing that buried polar residues can impart geometric specificity, as well as a discrete oligomeric state.[39–42] The role of salt bridges at inter-subunit interfaces, and the extent to which they may contribute to stability and specificity of subunit interactions, has also been investigated using coiled-coil and helical-bundle mini-protein models.[43–46] The findings from these studies have applicability to more complex protein interfaces.

## 4. Folding of oligomeric proteins

The primary sites for protein synthesis and folding are the cytosol and the endoplasmic reticulum. Cytosolic proteins are synthesized, fold, and oligomerize (where relevant), in the cytosol. Membrane and secretory proteins are synthesized in the endoplasmic reticulum (ER), and oligomerization typically occurs within the ER, although, in some cases, oligomerization takes place in the intermediate compartment and Golgi apparatus.[47] The stage at which oligomerization is protein-dependent and may be co-translational or post-translational, or may be subsequent to folding of the individual subunits.[48–51] Protein folding in cells is a complex phenomenon and often occurs with the assistance of molecular chaperones.[52] (In the case of oligomeric proteins that fold with the assistance of chaperones, oligomerization may occur after release of the folded subunits from the chaperone.[52]) In vitro studies of protein folding are generally only feasible with relatively simple proteins. These studies are generally performed under reversible and dilute conditions, which may not accurately represent the cellular milieu. Nonetheless, a great deal has been learned about protein folding from basic in vitro studies.[53,54] The majority of these studies have focused on monomeric proteins because of their greater simplicity. Computational studies have also added to our knowledge of the protein folding process.[53–56] However, it can be challenging to study even simple multichain systems computationally.[57]

Protein folding studies, both experimental and computational, are an area where simple model systems can be of great use. The folding pathways of a number of monomeric mini-proteins have been studied, including β-sheet mini-proteins,[58] the mixed α/β monomer BBA5,[59,60] and the Trp cage.[61,62] Furthermore, interesting results have been found in the studies of folding mechanisms of homo-oligomeric[63–66] and hetero-oligomeric[67] leucine zipper mini-proteins. An all-β homotetramer having four-stranded anti-parallel beta sheet subunits has been used for the study of folding thermodynamics.[16] We have recently reported the structure of the first oligomeric mixed α/β mini-protein BBAT2,[10] and anticipate that it will serve as a particularly valuable and protein-like model for the study of folding of oligomeric proteins. This tetrameric scaffold was further elaborated to result in the heterotetrameric α/β mini-proteins BBAhetT1 and BBAhetT2.[11] These structurally characterized heterotetrameric mini-proteins will serve as valuable models for the study of protein folding in a heterospecific system.

## 5. Determinants of oligomeric state

There is some subtlety in the specification of a unique oligomeric state. Often small changes in protein composition or environment can tip the balance from one state to the next. Some proteins coexist in more than one oligomeric state.[68–70] Many receptors undergo dimerization upon ligand binding.[71,72] Structural proteins, such as actin, can polymerize.[73] Other proteins can polymerize after undergoing a conformational change, giving rise to amyloid fibrils.[53,54] Moreover, closely related proteins may vary in their oligomeric forms. Furthermore, it is known that variation of inter-domain linker lengths can result in variations in oligomeric state.[74–76] Some examples are the legume lectins, which can dimerize by various modes, as well as tetramerize,[77] the cystine-knot growth factors,[78] and lumazine synthase.[79]

The determinants of a unique oligomeric state have been explored through the design of oligomeric and monomeric mini-proteins. Substitutions of core polar residues in coiled-coil peptides have been found to result in the loss of a specific oligomeric state.[80] Shortening an inter-domain linker has resulted in a domain-swapped

oligomeric mini-protein.[81] DeGrado and co-workers converted a domain-swapped dimeric three-helix bundle into a hexameric helical bundle by mutating three amino acids per monomer.[82] Negative design features preventing self-association have also been implemented through the iterative design of all-β mini-proteins.[83] Finally, a number of groups have used short peptides to study amyloid fibril formation.[84,85]

## 6. How does oligomerization arise?

There are several functional advantages that may be conferred by oligomerization and that may have been selected for through evolution.[4,5,70,86] (1) More complex scaffolds may better support function, for example, by the introduction of a new active site at the interface between subunits. It has been estimated that roughly one-sixth of oligomeric enzymes has an active site located at the inter-subunit interface. (2) Oligomeric proteins can be allosterically regulated, introducing an additional level of control. (3) There is a greater likelihood of an error-free transcript in a shorter protein sequence. A large protein composed of multiple, short, subunits, is more likely to be synthesized without errors than a single-chain protein of comparable size. (4) Where the monomer and oligomer differ in activity, additional regulatory flexibility may be achieved by regulating the conditions of oligomerization. (5) Oligomeric proteins may be subjected to amplified evolutionary pressures, as deleterious mutations may be more pronounced and thus removed sooner from the gene pool. Conversely, the advantages of beneficial mutations may also be made evident sooner. (6) Larger proteins are more resistant to degradation and denaturation. Indeed, an increase in oligomerization state is one of the protein stabilization strategies observed in thermophilic organisms.[87] However, it is worth noting that, in certain cases, oligomerization may be incidental to protein activity, and neither selected for, nor against, by evolutionary pressures[4,5,86] Moreover, there are cases where oligomerization might be disadvantageous, such as that of a protein with a large, hence slowly diffusing, substrate, that may be subjected to evolutionary pressures to be smaller and hence more rapidly diffusing so as to accelerate rates of collision with the cognate substrate.[4] Furthermore, it is not necessarily the case that protein evolution proceeds invariably in the direction of oligomerization.

The study of the evolution of protein oligomerization is an active field of scientific discourse. There are several mechanisms whereby oligomerization could have arisen. It is likely that there is not one general mechanism, but several mechanisms, and that oligomerization in different proteins has evolved by different routes. Genetically, oligomeric species could be created from monomeric proteins through some combination of a few genetic events.[5,88,89] Substitutions, insertions, and deletions comprise the basic set of mutations that, over time, will result in sequence drift. Such mutations could affect the electrostatic or geometric properties of a surface patch, or the length or geometry of an inter-domain linker, resulting in a change in the oligomeric state of a former-

ly monomeric protein. Recombination could result in the fusion of an oligomerization domain to a previously monomeric protein. Some of these mutations would be beneficial and selected for, some detrimental and selected against, and some neutral.

Mutations can lead to the introduction, expansion, or exposure of a hydrophobic patch on the surface of a protein, creating a new interface for oligomerization (Fig. 2). As all residues do not contribute equally to the oligomeric interface, it is reasonable to suppose that mutations at a few key residues necessary for mediation of an interaction could have such an effect. It has been proposed that relatively few mutations would be necessary for such a step, particularly in the case of dimeric proteins that fold via a three-step pathway.[35,90] Moreover, the close resemblance of subunit interfaces to surface residues may imply that they originated from surface residues.[35]

Oligomeric proteins that fold via a two-step pathway, that is, wherein the individual monomers are unfolded in the absence of oligomerization, may have evolved specifically toward the dimeric state, without the intermediacy of an independently folded monomeric state.[90] In particular, Xu et al.[90] have proposed a direct evolutionary pathway for dimeric proteins exhibiting two-state kinetics and either having a large and flat interface or having two short intertwined chains. Some examples of proteins in this class might include dimeric coiled coils, such as GCN4, which are unfolded in the absence of oligomerization interactions (although some studies suggest that the folding of GCN4 may be more complex than a simple two-state transition[91]).

Domain swapping is an elegant theory that proposes one mechanism whereby multi-domain proteins could evolve to a different oligomeric state. Certain proteins are known to coexist in both monomeric and oligomeric forms, wherein the interdomain interactions can occur both within a single monomer and between monomer units (Fig. 3).[70] Approximately, 40 such domain-swapped proteins have been documented at this time. These include the bacteriophage λ cro repressor, barnase, and diphtheria toxin.[70] In this case, a single chance
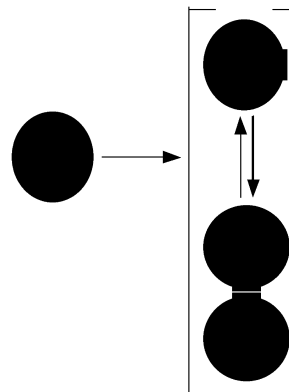


**Figure 2.** The evolution of a hydrophobic patch on the surface of a monomeric protein results in a protein capable of oligomerization.
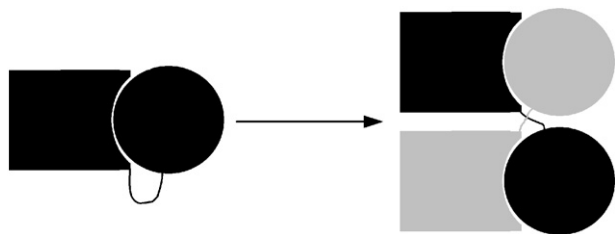
**Figure 3.** Domain swapping. A monomeric protein with two interacting domains may coexist with oligomeric state in which the domains have the same interaction between monomer units.

mutation that favors a higher oligomeric state over the monomeric form could be selected for, resulting in a sustained oligomer.[70,92]

Moreover, it may not be necessary to proceed through the intermediacy of a true domain-swapping protein with coexisting oligomeric states. Should a mutation affect the linker region such that the protein can no longer fold on itself, it may have the option of burying the hydrophobic surfaces in a domain-swapping interaction with another monomer. This typically occurs when a linker between the domains is shortened, precluding an intramonomer association.[74,75] Varying a specific linker length can result in an increase or decrease in oligomeric state.[76] Several oligomeric proteins and mini-proteins have been engineered from the corresponding monomers by shortening a linker to form a domain-swapped oligomer.[74,75,81,93]

Finally, oligomerization can arise via fusion of a gene encoding a dimerization or oligomerization domain, such as a coiled-coil domain, onto a previously monomeric protein (Fig. 4).[92] The newly oligomeric protein will then associate through this oligomerization domain and potentially form new contacts in regions brought into proximity by the association. Mutations favoring such an interaction, if beneficial, may be selected for thus producing a more robust oligomer.

It has been proposed that the basic helix-loop-helix (bHLH) family of oligomeric transcription factors may have arisen via modular evolution as described above.[94] Over time, the primordial bHLH oligomerization domain gets linked to various other proteins through gene duplication events and insertion into other genes, resulting in the large and diverse array of modern proteins bearing bHLH domains.[94] It is likely that other oligomerization modules may also have been propagated in this fashion.
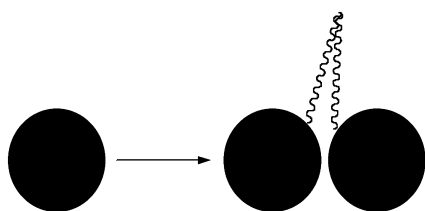


**Figure 4.** Fusion of an oligomerization domain to a monomeric protein results in an oligomeric fusion protein.

## 7. Relevance of oligomerization to evolution of modern proteins

Many proteins that are monomeric in the modern or evolved state are believed to have arisen from smaller, associating, fragments, that have, via gene duplication and gene fusion, become a single encoded protein.[95] Some examples of monomeric proteins that are believed to have arisen from homo-oligomeric precursors are the periplasmic binding proteins,[96,97] the 8-fold β/α barrel,[98] and the ribosome anti-association factor eIF6.[99] It is not essential that the merged genes be homologous, as in the above examples. The fusion of two different genes would result in an asymmetric protein. Each of the original associating fragments may comprise a domain of the fusion protein. Chothia found that approximately 67% of prokaryotic proteins, and about 80% of eukaryotic proteins, are composed of at least two domains.[89] It is likely that many of these multi-domain proteins arose from the genetic fusion of smaller, associating single-domain proteins.

The study of oligomerization may also have relevance to the evolution of single protein domains. Russell and co-workers have proposed that many protein domains arose from antecedent domain segments (ADSs).[100] ADSs are peptide sequences, encoded by short primordial genes, that spontaneously homo-oligomerize in aqueous solution to adopt a fold similar to a modern domain. These noncovalent, single domain, mini-proteins would have served structural or functional roles in the early protein world. Over time, evolutionary pressures would have favored the development of a single fused gene composed of multiple ADS repeats because of entropic and thermodynamic factors. Support for this hypothesis comes from the study of several domains exhibiting internal symmetry, with structural and sequence repeats.[100] Among these are the beta trefoil, the beta propeller, and the beta spiral.[101] Furthermore, the ability of proteolytic fragments of several independently folded domains and small proteins to reassemble into noncovalent assemblies resembling the native could be explained by an ADS-like evolutionary pathway.[100,102–104] An ADS could also have formed a hetero-oligomeric protein with a different ADS. Such a hetero-oligomer could then, following gene duplication and fusion, form a more complex single domain. Russell and co-workers propose that the monomeric protein thioredoxin (Fig. 5) could have been one such example, formed from a combination of two ββα ADSs and one βα ADS.[100]

Studies of designed mini-proteins have revealed that it is difficult to introduce functionality into a single peptide, such as an ADS, even if the structures are well-folded.[19,105] Apparently, a larger scaffold is preferable from the vantage of substrate binding and catalysis, allowing for more residues to be devoted to function rather than structure. By contrast, there has been considerable success in introducing functionality into oligomeric mini-proteins.[7,18,20,21,105] We could thus expect significant evolutionary pressure on an ADS or microgene product to homo- or hetero-oligomerize. As noted above, designed mini-proteins, and in particular designed oligo-
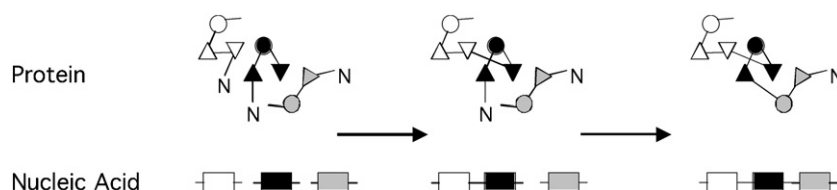
**Figure 5.** The proposed evolution of thioredoxin from three ADSs. Figure adapted from Ref. 100. Triangles represent β-sheets and circles represent α-helices.

meric mini-proteins, bear a close resemblance to proposed primordial proteins. It may be possible to deduce principles about the early protein environment, and about the evolutionary selection processes of microgene products, through the exercise of introducing function and elaborating structure in mini-proteins.

## 8. Conclusions

Oligomeric proteins are prevalent in nature, comprising roughly one third of cellular proteins.[2] A number of functional advantages favor the evolution of oligomers from monomeric precursors. The study of protein oligomerization can provide insights into the early protein environment and the evolution of modern proteins. The specific composition of an inter-subunit interface is dependent on the nature and affinity of the interactions comprising that interface, but both hydrophobic and polar interactions play key roles in most interfaces. There are several mechanisms whereby protein oligomerization could have arisen, including the evolution of a new site of interaction, the direct evolution of a protein for the oligomeric state, a mutation in the hinge region leading to a domain-swapped oligomer, and the fusion of a preexisting oligomerization domain. Oligomeric mini-proteins constitute simple and tractable model systems for the study of oligomeric proteins.

## Acknowledgments

## References and notes

1. Goodsell, D. S. *Trends Biochem. Sci.* **1991**, *16*, 203–206.
2. Goodsell, D. S.; Olson, A. J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153.
3. Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13–20.
4. Nooren, I. M. A.; Thornton, J. M. *J. Mol. Biol.* **2003**, *325*, 991–1018.
5. Nooren, I. M. A.; Thornton, J. M. *EMBO J.* **2003**, *22*, 3486–3492.
6. Monod, J.; Olsen, K. W.; Sabesan, M. N.; Buehner, M.; Ford, G. C.; Rossman, M. G. *J. Mol. Biol.* **1965**, *12*, 88–118.
7. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779–819.
8. Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F. *Acc. Chem. Res.* **2000**, *33*, 745–754.
9. Kohn, W. D.; Hodges, R. S. *Trends Biotechnol.* **1998**, *16*, 379–389.
10. Ali, M. H.; Peisach, E.; Allen, K. N.; Imperiali, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12183–12188.
11. Ali, M. H.; Taylor, C. M.; Grigoryan, G.; Allen, K. N.; Imperiali, B.; Keating, A. E. *Structure* **2005**, 13.
12. Das, C.; Shankaramma, S. C.; Balaram, P. *Chem. Eur. J.* **2001**, *7*, 840–847.
13. Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8747–8751.
14. Ilyina, E.; Roontga, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245–5250.
15. Venkatraman, J.; Gowda, G. A. N.; Balaram, P. *J. Am. Chem. Soc.* **2002**, *124*, 4987–4994.
16. Jang, H.; Hall, C. K.; Zhou, Y. *Protein Sci.* **2004**, *13*, 40–53.
17. Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419–426.
18. Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425–431.
19. Shogren-Knaak, M. A.; Imperiali, B. *Bioorg. Med. Chem.* **1999**, *7*, 1993–2002.
20. Broo, K. S.; Nilsson, H.; Nilsson, J.; Floodberg, A.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 4063–4068.
21. Allert, M.; Kjellstrand, M.; Broo, K.; Nilsson, Å.; Baltzer, L. *J. Chem. Soc., Perkin Trans.* **1998**, *2*, 2271–2274.
22. Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705–708.
23. Jones, S.; Thornton, J. M. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 31–65.
24. Argos, P. *Protein Eng.* **1988**, *2*, 101–113.
25. Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121–132.
26. Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C. *Nature* **1987**, *328*, 834–836.
27. Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155–164.
28. Caffrey, D. R.; Somaroo, S.; Hughes, J. D.; Mintseris, J.; Huang, E. S. *Protein Sci.* **2004**, *13*, 190–202.
29. Valdar, W. S. J.; Thornton, J. M. *Proteins: Struct. Funct. Genet.* **2001**, *42*, 108–124.
30. Grishin, N. V.; Phillips, M. A. *Protein Sci.* **1994**, *3*, 2455–2458.
31. Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383–386.
32. Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1–9.
33. Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 331–342.
34. Miller, S. *Protein Eng.* **1989**, *3*, 77–83.
35. D'Alessio, G. *Prog. Biophys. Mol. Biol.* **1999**, *72*, 271–298.
36. McCoy, A. J.; Epa, V. C.; Colman, P. M. *J. Mol. Biol.* **1997**, *268*, 570–584.

37. Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946–950.
38. Hubbard, S. J.; Argos, P. *Protein Sci.* **1994**, *3*, 2194–2206.
39. Akey, D. L.; Malashkevich, V. N.; Kim, P. S. *Biochemistry* **2001**, *40*, 6352–6360.
40. Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642–8648.
41. Eckert, D. M.; Malashkevich, V. N.; Kim, P. S. *J. Mol. Biol.* **1998**, *284*, 859–865.
42. Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603–12610.
43. McClain, D. L.; Binfet, J. P.; Oakley, M. G. *J. Mol. Biol.* **2001**, *313*, 371–383.
44. Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1994**, *237*, 500–512.
45. Burkhard, P.; Ivaninski, S.; Lustig, A. *J. Mol. Biol.* **2002**, *2002*, 901–910.
46. Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431–2437.
47. Reddy, P. S.; Corley, R. B. *BioEssays* **1998**, *20*, 546–554.
48. Prabakaran, D.; Kim, P. S.; Dixit, V. M.; Arvan, P. *Eur. J. Cell Biol.* **1996**, *70*, 134–141.
49. Bulleid, N. J.; Dalley, J. A.; Lees, J. F. *EMBO J.* **1997**, *16*, 6694–6701.
50. Chen, W.; Helenius, J.; Braakman, I.; Helenius, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6229–6233.
51. Copeland, C. S.; Zimmer, K.-P.; Wagner, K. R.; Healey, G. A.; Mellman, I.; Helenius, A. *Cell* **1988**, *53*, 197–209.
52. Hartl, F. U.; Hayer-Hartl, M. *Science* **2002**, *295*, 1852–1858.
53. Dobson, C. M. *Nature* **2003**, *426*, 884–890.
54. Dobson, C. M. *Semin. Cell Dev. Biol.* **2004**, *15*, 3–16.
55. Daggett, V.; Fersht, A. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497–502.
56. Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.
57. Saunders, J. A.; Scheraga, H. A. *Biopolymers* **2003**, *68*, 300–317.
58. Searle, M. S.; Ciani, B. *Curr. Opin. Struct. Biol.* **2004**, *14*, 458–464.
59. Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.
60. Zagrovic, B.; Pande, V. S. *Nat. Struct. Biol.* **2003**, *10*, 955–961.
61. Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
62. Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549.
63. Zitzewitz, J. A.; Bilsel, O.; Luo, J.; Jones, B. E.; Matthews, C. R. *Biochemistry* **1995**, *34*, 12812–12819.
64. Dürr, E.; Jelesarov, I.; Bosshard, H. R. *Biochemistry* **1999**, *38*, 870–880.
65. Ibarra-Molero, B.; Makhatadze, G. I.; Matthews, C. R. *Biochemistry* **2001**, *40*, 719–731.
66. Zhu, H.; Celinski, S. A.; Scholtz, J. M.; Hu, J. C. *Protein Sci.* **2001**, *10*, 24–33.
67. Wendt, H.; Leder, L.; Härmä, H.; Jelesarov, I.; Baici, A.; Bosshard, H. R. *Biochemistry* **1997**, *36*, 204–213.
68. Rochet, J.-C.; Brownie, E. R.; Oikawa, K.; Hicks, L. D.; Fraser, M. E.; James, M. N. G.; Kay, C. M.; Bridger, W. A.; Wolodko, W. T. *Biochemistry* **2000**, *39*, 11291–11302.
69. Ali, S. A.; Iwabuchi, N.; Matsui, T.; Hirota, K.; Kidokoro, S.-i.; Arai, M.; Kuwajima, K.; Schuck, P.; Arisaka, F. *Biophys. J.* **2003**, *85*, 2606–2618.
70. Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, *11*, 1285–1299.
71. Schlessinger, J. *Cell* **2002**, *110*, 669–672.
72. Plotnikov, A. N.; Schlessinger, J.; Hubbard, S. R.; Mohammadi, M. *Cell* **1999**, *98*, 641–650.
73. Rafelski, S. M.; Theriot, J. A. *Annu. Rev. Biochem.* **2004**, *73*, 209–239.
74. Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Biochemistry* **1996**, *35*, 735–742.
75. Pei, X. Y.; Holliger, P.; Murzin, A. G.; Williams, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9637–9642.
76. MacBeath, G.; Kast, P.; Hilvert, D. *Protein Sci.* **1998**, *7*, 1757–1767.
77. Prabu, M. M.; Suguna, K.; Vijayan, M. *Proteins: Struct. Funct. Genet.* **1999**, *35*, 58–69.
78. Sun, P. D.; Davies, D. R. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 269–291.
79. Zylberman, V.; Craig, P. O.; Klinke, S.; Braden, B. C.; Cauerhff, A.; Goldbaum, F. A. *J. Biol. Chem.* **2004**, *279*, 8093–8101.
80. Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401–1407.
81. Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002–1003.
82. Ghirlanda, G.; Lear, J. D.; Ogihara, N. L.; Eisenberg, D.; DeGrado, W. F. *J. Mol. Biol.* **2002**, *319*, 243–253.
83. Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2754–2759.
84. Nesloney, C. L.; Kelly, J. W. *Bioorg. Med. Chem.* **1996**, *4*, 739–766.
85. Bosques, C. J.; Imperiali, B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7593–7598.
86. Goodsell, D. S.; Olson, A. J. *Trends Biochem. Sci.* **1993**, *18*, 65–68.
87. Walden, H.; Bell, G. S.; Russell, R. J. M.; Siebers, B.; Hensel, R.; Taylor, G. L. *J. Mol. Biol.* **2001**, *306*, 745–757.
88. Long, M.; Betrán, E.; Thornton, K.; Wang, W. *Nat. Rev. Genet.* **2003**, 4.
89. Chothia, C.; Gough, J.; Vogel, C.; Teichmann, S. A. *Science* **2003**, *300*, 1701–1703.
90. Xu, D.; Tsai, C.-J.; Nussinov, R. *Protein Sci.* **1998**, *1998*, 533–544.
91. Dragan, A. I.; Privalov, P. L. *J. Mol. Biol.* **2002**, *321*, 891–908.
92. *Nat. Struct. Biol.* **1994**, *1*, 411–412.
93. Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10687–10691.
94. Morgenstern, B.; Atchley, W. R. *Mol. Biol. Evol.* **1999**, *16*, 1654–1663.
95. Taylor, W. R.; May, A. C. W.; Brown, N. P.; Aszódi, A. *Rep. Prog. Phys.* **2001**, *64*, 517–590.
96. Dassa, E.; Saurin, W.; Hofnung, M.; Kuan, G.; Saier, M. H., Jr. *Res. Microbiol.* **1995**, *146*, 271–278.
97. Saurin, W.; Dassa, E. *Protein Sci.* **1994**, *3*, 325–344.
98. Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. *Science* **2000**, *289*, 1546–1550.
99. Groft, C. M.; Beckmann, R.; Sali, A.; Burley, S. K. *Nat. Struct. Biol.* **2000**, *7*, 1156–1164.
100. Lupas, A. N.; Ponting, C. P.; Russell, R. B. *J. Struct. Biol.* **2001**, *134*, 191–203.
101. Ponting, C. P.; Russell, R. R. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 45–71.
102. Galaktos, N. G.; Walsh, C. T. *Biochemistry* **1987**, *26*, 8475–8480.
103. Girons, I. S.; Gilles, A.-M.; Margarita, D.; Michelson, S.; Monnot, M.; Fermandjian, S.; Danchin, A.; Bârzu, O. *J. Biol. Chem.* **1987**, *262*, 622–629.
104. Pelletier, J. N.; Campbell-Valois, F.-X.; Michnick, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12141–12146.
105. McDonnell, K. A. *Towards Incorporation of Catalytic Function into Small Folded Peptide Scaffolds*, Massachusetts Institute of Technology 2001.

**Mayssam H. Ali** received her undergraduate degree in chemistry from Harvard College in 1998, and a Ph.D. in organic chemistry from the Massachusetts Institute of Technology, under the supervision of Professor Imperiali, in 2004.

**Barbara Imperiali** is the Class of 1922 Professor of Chemistry and Professor of Biology at the Massachusetts Institute of Technology (MIT). She received a B.Sc. in medicinal chemistry from University College London in 1979, followed by a Ph.D. in synthetic organic chemistry from the Massachusetts Institute of Technology in 1983. Prof. Imperiali was a professor of chemistry at the California Institute of Technology until 1997, when she joined the faculty at MIT.