PROTEINS: Structure, Function, and Genetics 34:1–3 (1999)

EDITOR'S CORNER

Arm-Domain Interactions in Proteins: A Review

As the first multimeric protein whose structure was determined by X-ray crystallography, hemoglobin became a paradigm for how complex proteins are built and function. Hemoglobin's compact, unitary structure with extensive and complicated interfaces between the subunits suggested that proteins would be globular, and that mechanisms for responding to the presence of ligands would be most intricate. Therefore, less than a decade ago many of us were surprised to learn that proteins frequently consist of functionally independent domains. Now, as we probe deeply into the regulation of protein function and activity will we be surprised once again? Perhaps so, for somewhere between needing the involved intersubunit interfaces of hemoglobin and needing only the entropic effects of tethering domains to be near one another lies an intermediate situation. Apparently proteins or domains of proteins can communicate by means of arms or tails, stretches of five to twenty or more amino acid residues, often at the N- or C-terminus, that under certain conditions extend from one domain and make specific contacts with another domain. In contrast to domains, which fold and assume defined structures on their own, arm regions are unstructured until the arm makes specific contact with a domain.

The general motif of an arm interacting with a globular domain has been known for many years. Such an interaction was reported in 1970 in the structure of lactate dehydrogenase.[1] Since then, many other examples of such static, structural arm-domain interactions have been observed, including enzymes,[2–6] the linking of capsid subunits in viruses,[7] and even a heat shock protein.[8] For the most part, these arm-domain interactions appear to augment domain-domain interactions between structured interfaces. They also appear to be static in the sense that the arms remain in place during the functioning of the proteins. More interesting, however, are the cases in which arm-domain interactions are not permanent, are made with multiple partners, constitute the entire interaction between two proteins or domains, or are regulated.

An early example of a transient arm-domain interaction, although extending the definition of domain to include a sequence of DNA, is the N-terminal arm of the lambda phage repressor. When the protein binds to DNA, the first three residues of the eight residue N-terminal arm bind in the major groove on the back face of the DNA.[9,10] A much more complex DNA binding entity, the nucleosome also behaves similarly. The arms of the eight histone proteins in a nucleosome can also contact DNA[11] and also interact with adjacent nucleosomes and other chromosomal proteins.[12,13] Acetylation of the N-terminal arms of the core histones reduces their stability with DNA and allows greater access of transcription factors and facilitates transcription.[14]

Arm-domain interactions are particularly useful in the formation of hetero-oligomeric families. One such class are the homeodomain proteins that activate transcription of developmental genes in higher organisms. In these proteins, the residues flanking the conserved homeodomain DNA binding region are variable and help the proteins interact with appropriate coactivator proteins or DNA or both. For example, the Hox and Engrailed homeodomain proteins interact with the Pbx and Extradenticle homeodomain proteins. In HOXD-4, the sequence YPWMK that lies N-terminal to the homeodomain is required for the interaction.[15] The extradenticle EXD protein also uses the YPWM motif to bind to another protein. Here, the conserved residues also may be involved with reading the DNA sequence, and thus may simultaneously contact DNA and protein.[16] Arg3 and Arg5 of the N-terminal arm of the Engrailed homeodomain protein contact the minor groove of DNA.[17] The N- terminal arms of a number of other homeodomain proteins are known to contact the minor groove as well.[15,18–21] On the other side of the homeodomain region, residues of the MATα2 homeoprotein that lie C-terminal to the homeodomain are unstructured in the absence of MATa1. In the presence of MATa1, 16 residues of MATα2 bind to a1 and hold the two proteins together.[22]

By virtue of their arm-domain interactions, many of the homeodomain proteins are able to function with a number of different protein partners. The regulatory region of a gene possessing appropriately located binding sites for the two transcriptional activator proteins would be significantly occupied and hence expressed only when both of the two appropriate partner proteins are present in the cell and can form the correct functional heterodimer by protein-protein interactions. That is, homeoproteins are readily capable of combinatorial gene regulation. For example, a set of ten different homeodomain proteins, each capable of binding the same arm sequence of amino acids, and ten different coactivators, each with the same arm sequence, would be theoretically capable of generating 10×10 different patterns of gene regulation with only 20 proteins. Such a scheme allows cells to function with a minimum number and minimum amounts of transcription factors. Similar combinatorics can be achieved without the use of protein-protein interactions, but at the loss of cooperativity in DNA binding, necessitating higher concentrations of the proteins. Arm-domain interactions are thus an efficient way to generate promiscuous oligomerization, for it is

© 1999 WILEY-LISS, INC.

almost inconceivable that protein-protein interfaces like those present in hemoglobin could be used for such a purpose.

Eukaryotic transcription factors other than homeodomain proteins also use arms for interactions, both to activate transcription or to inhibit it. The murine homeodomain protein Msx-1 associates with TATA binding protein through residues in its N-terminal arm.[23] Conversely, the retinoblastoma (Rb) tumor protein, which itself is a repressor of gene activation, blocks progression through the cell cycle by binding to and inactivating activators of the E2F family. Rb is inactivated by oncoproteins that bind to it, including the E7 protein of human papilloma virus, large T of SV40, and adenovirus E1A. These proteins displace E2F activators, thereby freeing the E2F activators to function when they should not. E7, large T, and E1A bind to Rb via a short peptide including the sequence LxCxE[24]. Finally, 12 amino acids at the C-terminus of the progesterone receptor are necessary and sufficient for interaction with a transcriptional corepressor.[25]

Cells contain many receptors whose activation must be transmitted to particular intracellular targets, often through specific protein cascades. Arms permit unique interactions in these pathways. For example, in at least some cases, the C-terminal three or four residues of the alpha subunit of a G protein are crucial for interactions with the appropriate receptor protein. Interchanging C-termini interchanges specificity.[26–28] The C-terminus of the γ subunit may also play a similar role, at least in the rhodopsin-G protein interaction.

Phages T4 and N4 provide two well-studied prokaryotic examples where arms mediate interaction between specific proteins. In phage T4, the products of gene 45 form a ring that is oligomerized around DNA. This sliding ring can thus hold proteins in the immediate vicinity of DNA. The phage sigma factor gp55 and co-activator gp33[30] as well as the DNA polymerase and the gp44/62 protein[31] use a conserved C-terminal sequence to bind to the sliding ring-clamp product of gp45. Phage N4 late gene expression, recombination, and replication require N4 single-strand binding protein, N4SSB. Mutations show the C-terminus of N4SSB is required for interaction with the *E. coli* RNA polymerase as well as for replication and recombination.[32]

Direct regulation involving arms is also known. Proteins in the Src family, including Lck, Hck, Lyn, Blk, Src, Fyn, Yrk, Fgr, and Yes, are regulated by phosphorylation of a tyrosine in the protein's C-terminal tail. When the tails of Src or Hck are phosphorylated, they bind to the SH2-SH3 domains (which bind phosphotyrosine and polyproline arms themselves) and help hold the proteins so that the kinase domains are inactive.[33,34] When the tails are not phosphorylated and do not bind the SH2 and SH3 domains, or when they are displaced by the presence of tight-binding peptides,[34] the kinase is active.

AraC protein, a regulator of the L-arabinose operon in *E. coli*, is another example of an arm being involved in regulation.[35] In the absence of arabinose, the protein's N-terminal arms, which extend from the dimerization domain of the protein, bind to the DNA binding domains. They hold the DNA binding domains in such an orientation that the protein much prefers to loop between two well-separated half-sites on the DNA. When arabinose is present, the protein's N-terminal arms prefer to fold over arabinose and thus are pulled off the DNA binding domains. This frees the DNA binding domains and allows them to bind to adjacently located half-sites rather than looping between distantly located half-sites. The *cis*-binding then activates transcription by stimulating the binding of RNA polymerase to the adjacent promoter.

We cannot now easily determine the extent of Nature's use of arms. Likely, functional arms are under represented in structure databases because unstructured arms are not seen in X-ray crystallography, and, aside from stable oligomers, the structures of relatively few protein complexes have been determined thus far. Whether arms that do not permanently bind to their target domains need to have developed special mechanisms to avoid digestion by intracellular proteases is also not yet clear. Perhaps arms below a critical size are resistant, or perhaps arms lack certain protease sensitive subsequences. Aside from their structural use in the formation of stable oligomers, arm-domain interactions are particularly well suited for situations in which one protein must interact with a number of different partners. Two general areas in which such interactions are important are signal transmission pathways in which the same proteins may be utilized in several pathways, and in the regulation of transcription in which the same transcription factor may be involved in the regulation of several genes.

In summary, arms on proteins provide versatile handles for interaction with other proteins. A protein with a binding site for a short peptide can bind to any other protein possessing an unstructured and accessible "arm" containing the peptide sequence. It is far easier both for evolution or genetic engineers to add such a sequence to a protein than it is to tailor a specific and extensive two-dimensional interface whose structure is retained even in the absence of the partner protein. It may well turn out that given any protein, a peptide can be found, either by direct genetic selection or ultimately by structure analysis, that binds to it. If a peptide can be found that can bind only in the presence of ligand to the protein, then we can imagine engineering desirable ligand-regulated protein-protein associations or even engineering regulation into chimeric proteins by means of a regulating arm. Such a ligand-mediated interaction should be far easier to generate and manipulate than those like the extraordinarily complex ligand-mediated inter-subunit interactions that are seen in hemoglobin.

REFERENCES

1. Adams M, Ford G, Koekoek R et al. Structure of lactate dehydrogenases at 2.8 Å resolution. Nature 1970;277:1098–1103.
2. Imada K, Tanaka N, Katsube Y, Matsuura Y, Oshima T. Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of thermus thermophilus at 2.2 Å resolution, J Mol Biol 1991;222:725–738.

3. Erskine P, Senior N, Awan S et al. X-ray structure of 5-aminolaevulinate dehydratase, a hybrid aldolase. Nat Struct Biol 1997;4: 1025–1031.
4. Huang W, Jia J, Cummings J, Nelson M, Schneider G, Lindqvist Y. Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. Structure 1997;5:691–699.
5. Kingston R, Scopes R, Baker N. The structure of glucose-fructose oxidoreductase from zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-dissociable NADP. Structure 1996;4:1413–1428.
6. Katti S, Katz B, Wyckoff H. Crystal structure of muconolactone isomerase at 3.3Å resolution. J Mol Biol 1989;205:557–571.
7. Lee S, Owen K, Choi H et al. Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. Structure 1996;4:531–541.
8. Kim K, Kim R, Kim S. Crystal structure of a small heat-shock protein. Nature 1998;394:595–599.
9. Pabo C, Lewis M. The operator-binding domain of lambda repressor: Structure and DNA recognition. Nature 1982;298:443–447.
10. Pabo C, Krovatin W, Jeffrey A, Sauer R. The N-terminal arms of lambda repressor wrap around the operator DNA. Nature 1982; 298:441–443.
11. Luger K, Mäder A, Richmond R, Sargent D, Richmond T. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 1997;389:251–260.
12. Lee K, Mayes J. The N-terminal tail of histone H2A binds to two distinct sites within the nucleosome core. Proc Natl Acad Sci USA 1997;94:8959–8964.
13. Georgel P, Tsukiyama T, Wu C. Role of histone tails in nucleosome remodeling by Drosophila NURF. EMBO J 1997;16:4717–4726.
14. Ura K, Kurumizaka H, Dimitrov S, Almouzni G, Wolff A. EMBO J 1997;16:2096–2107.
15. Ekker S, Jackson D, von Kessler D, Sun B, Young K, Beachy P. The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J 1994;13:3551–3560.
16. Shanmugam K, Featherstone M, Saragovi H. Residues flanking the HOX YPWM motif contribute to cooperative interactions with PBX. J Biol Chem 1997;272:19081–19087.
17. Ades D, Sauer R. Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. Biochemistry 1995; 34:1460–14608.
18. Furukubo-Tokunaga K, Flister S, Gehring W. Functional specificity of the antennapedia homeodomain. Proc Natl Acad Sci USA 1993;90:6360–6304.
19. Phelan M, Featherstone M. Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX-PBX heterodimers. J Biol Chem 1997;272: 8635–8643.
20. Gruschus J, Tsao D, Wang L, Nirenberg M, Ferretti J. Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. Biochemistry 1997;36:5372–5380.
21. Billeter M, Qian Y, Otting G, Muller M, Gehring W, Wüthrich K. Determination of the nuclear magnetic resonance solution structure of an antennapedia homeodomain-DNA complex. J Mol Biol 1993;234:1084–1093.
22. Li T, Stark M, Johnson A, Wolberger C. Crystal structure of the MATa1/MATalpha2 homeodomain heterodimer bound to DNA. Science 1995;270:262–269.
23. Zhang H, Catron K, Abat-Shen C. A role for the Mxs-1 homeodomain in transcriptional regulation: Residues in the N-terminal arm mediate TATA binding protein interaction and transcriptional repression. Proc Natl Acad Sci USA 1996;93:1764–1769.
24. Lee J, Russo A, Pavletich N. Structure of the retinoblastoma tumor-suppressor pocket domain bound to a peptide from HPV E7. Nature 1998;391:859–865.
25. Xu J, Nawaz Z, Tsai S, Tsai M, O'Malley B. The extreme C terminus of progesterone receptor contains a transcriptional repressor domain that functions through a putatative corepressor. Proc Natl Acad Sci USA 1996;93:12195–12199.
26. Conklin R, Farfel Z, Lustig K, Julius D, Bourne H. Substitution of three amino acids switches receptor specificity of Gq alpha to that of Gi alpha. Nature 1993;363:274–276.
27. Dratz E, Furstenau J, Lambert C.et al. NMR structure of a receptor-bound G-protein peptide. Nature 1993;363:276–281.
28. Hamm H, Deretic D, Arendt A, Hargrave P, Koenig B, Hofmann K. Site of G protein binding to rhodopsin mapped with synthetic peptides from the alpha subunit. Science 1998;241:832–835.
29. Kisselev O, Pronin A, Ermolaeva M, Gautam N. Receptor-G protein coupling is established by a potential conformational switch in the beta-gamma complex. Proc Natl Acad Sci USA 1995;92:9102–9106.
30. Sanders G, Kassavetis G, Geiduschek E. Dual targets of a transcriptional activator that tracks on DNA. EMBO J 1997;16: 3124–3132.
31. Berdis A, Soumillion P, Benkovic S. The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation. Proc Natl Acad Sci USA 1996;93:12822–12877.
32. Choi M, Miller A, Cho N, Rothman-Denes L. Identification, cloning, and characterization of the bacteriophage N4 gene encoding the single-stranded DNA-binding protein. J Biol Chem 1995; 270:22541–22547.
33. Xu W, Harrison S, Eck M. Three-dimensional structure of the tyrosine kinase c-Src. Nature 1997;385:595–602.
34. Sicheril F, Moarefi I, Juriyan J. Crystal structure of the Src family tyrosin kinase Hck, Nature 1997;385:602–609.
35. Saviola B, Seabold R, Schleif R. Arm-domain interactions in AraC. J Mol Biol 1998;278:539–548.

Robert Schleif
Biology Department
Johns Hopkins University
Baltimore, Maryland