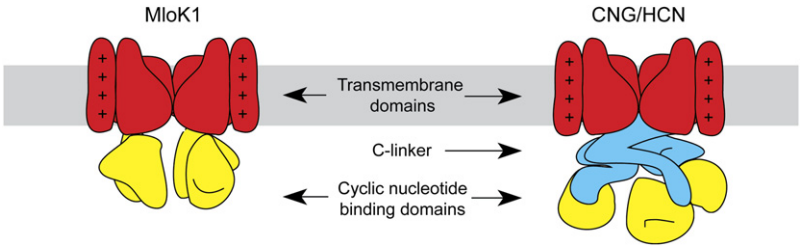

Figure 1. Domain Architectures of Prokaryotic (Left) and Eukaryotic (Right) Cyclic Nucleotide-Regulated Ion Channels

fit the activation of eukaryotic cyclic nucleotide-regulated channels (Ulens and Siegelbaum, 2003). Instead, the difference likely stems from the fact that the X-ray structure was solved from an isolated fragment of the channel, a fragment that likely formed a nonphysiological arrangement in the crystal.

Determining the biological unit—the physiological arrangement of subunits—in a crystal is not a trivial matter. There is no a priori way of knowing which intermolecular contacts are physiological and which are a product of crystal formation. This is particularly problematic for protein fragments. In MloK1, it seems that, without a C-linker, the transmembrane domains are necessary to ensure the correct physiological arrangement of the subunits.

Over and above the results, the work by Chiu et al. (2007) demonstrates the power of EM to solve intractable structural problems. EM studies can help elucidate the structures of full-length proteins and protein complexes that have been difficult to solve with X-ray crystallography. EM also can capture proteins in different conformational states, which opens the window for studies on dynamic structural rearrangements—arguably the future of biochemistry. Improvements in EM will lead only to more beautiful structures. And joined with X-ray crystallography, we should expect new and tantalizing insights into the structures of notoriously difficult proteins. A combination of approaches always leads to a more robust understanding of the microscopic world.

REFERENCES

Biskup, C., Kusch, J., Schulz, E., Nache, V., Schwede, F., Lehmann, F., Hagen, V., and Benndorf, K. (2007). Nature *446*, 440–443.

Chiu, P.-L., Pagel, M.D., Evans, J., Chou, H.-T., Zeng, X., Gipson, B., Stahlberg, H., and Nimigean, C.M. (2007). Structure *15*, this issue, 1053–1064.

Clayton, G.M., Silverman, W.R., Heginbotham, L., and Morais-Cabral, J.H. (2004). Cell *119*, 615–662.

Craven, K.B., and Zagotta, W.N. (2006). Annu. Rev. Physiol. *68*, 375–401.

Cukkemane, A., Gruter, B., Novak, K., Gensch, T., Bonigk, W., Gerharz, T., Kaupp, U.B., and Seifert, R. (2007). EMBO Rep. *8*, 749–755.

Long, S.B., Campbell, E.B., and Mackinnon, R. (2005). Science *309*, 897–903.

Nimigean, C.M., and Pagel, M.D. (2007). J. Mol. Biol. *371*, 1325–1337.

Nimigean, C.M., Shane, T., and Miller, C. (2004). J. Gen. Physiol. *124*, 203–210.

Ulens, C., and Siegelbaum, S.A. (2003). Neuron *40*, 959–970.

Zagotta, W.N., Olivier, N.B., Black, K.D., Young, E.C., Olson, R., and Gouaux, E. (2003). Nature *425*, 200–205.

How to Get All "A"s in Polyadenylation

Corey R. Mandel[1] and Liang Tong[1,*]
[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA
*Correspondence: ltong@columbia.edu
DOI 10.1016/j.str.2007.08.002

In an elegant study in this issue of *Structure*, Balbo and Bohm (2007) report the crystal structure of yeast poly(A) polymerase in a ternary complex with its substrate MgATP and the elongating poly(A) tail, providing molecular insights into the mechanism of polyadenylation.

Most eukaryotic mRNA precursors (pre-mRNAs) must undergo extensive processing before they can be exported from the nucleus to the cytoplasm and translated into proteins. At the 3′ end, the pre-mRNA is cleaved at a specific location and a polyadenylate tail (poly(A) tail) of about 200–300 nucleotides is added. A large complex of more than 15 proteins is required for this 3′-end processing. It has been known for over thirty years that poly(A) polymerase (PAP, Pap1p in yeast) catalyzes the addition of the poly(A) tail (Edmonds, 2002). PAP belongs to the DNA polymerase β superfamily of enzymes, but does not require a template.

Earlier structural studies of yeast and mammalian PAP free enzyme and complex with MgATP and dATP show that the enzyme contains three domains: N-terminal domain, middle domain, and C-terminal domain

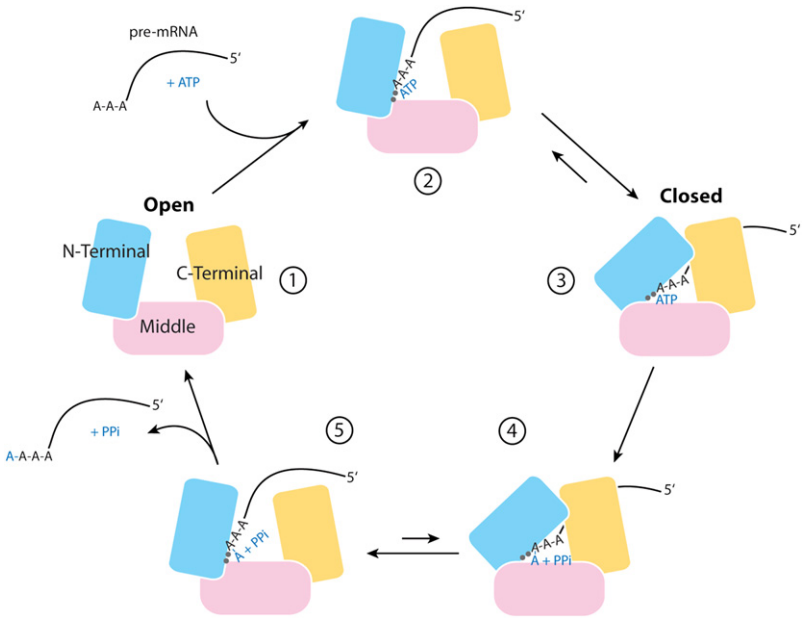

Figure 1. A Model for the Catalytic Cycle of Poly(A) Polymerase
The free enzyme exists in an open form. Upon binding both substrates (in random order), the enzyme undergoes a conformational change to the closed form, where catalysis occurs. After catalysis, the enzyme returns to the open form to release the products. The two metal ions (Mg^{2+} or Mn^{2+}) in the active site are shown as gray spheres.

(Figure 1) (Balbo et al., 2007; Bard et al., 2000; Martin et al., 2000, 2004). The active site, with two divalent cations, is located at the bottom of a large cleft between the N- and C-terminal domains, near the interface between the N-terminal and middle domains. Structural and biochemical data hint that PAP is an induced-fit enzyme (Balbo et al., 2005, 2007). However, the molecular mechanism of polyadenylation is still poorly understood, as the structure of PAP in a ternary complex with MgATP and the elongating poly(A) tail is not available.

In this issue of *Structure*, Balbo and Bohm report that they have succeeded in determining the structure at high resolution (1.8 Å) of such a ternary complex (Balbo and Bohm, 2007). To prevent the enzyme from turning over the substrates, the authors mutated the crucial Asp154 residue to alanine in the active site of yeast Pap1p, and this clever trick enabled them to trap the enzyme in the ternary complex. A five-residue oligoadenylate RNA molecule was used to mimic the elongating poly(A) tail of the pre-mRNA, and the crystallographic analysis showed that only its last three residues (A_{-3}-A_{-2}-A_{-1}) are highly ordered. The structure therefore suggests that PAP can meaningfully interact with only the incoming MgATP molecule and the last three residues in the elongating poly(A) tail, consistent with earlier biochemical data (Zhelkovsky et al., 1998).

The structure of the ternary complex reveals Pap1p in its most closed form observed so far, with a large movement of the N-terminal domain relative to the other two domains (Figure 1). There are now some direct contacts between the N- and C-terminal domains in this closed form, burying about 360 $Å^2$ of the surface area, and the active site cleft has been reduced to an active site tunnel. MgATP is bound at one end of the tunnel, and the poly(A) tail enters the tunnel from the opposite face of the enzyme. This tunnel is larger than the substrates, as it also contains a large number of ordered solvent molecules. The last three residues of the poly(A) tail assume an extended conformation. Their bases are completely buried by the enzyme but are located far away from each other. Many of the polar interactions between Pap1p and the poly(A) are mediated by waters, and the three adenine bases are not specifically recognized by the enzyme. This is consistent with the biochemical requirement that PAP has to initiate poly(A) synthesis at freshly cleaved pre-mRNA 3′ ends, which generally do not have poly(A) sequences. On the other hand, the 2′ hydroxyl group at the −1 and −2 positions of the poly(A) is hydrogen-bonded directly to the enzyme, explaining PAP's preference for RNA substrates.

The 3′ hydroxyl group of the last residue (A_{-1}) in the poly(A) is positioned only 3.2 Å from the α-phosphorus atom of MgATP, in the correct position for initiating the adenylation reaction. Therefore, mutation of the Asp154 residue may not have seriously affected the substrate binding modes.

Given that PAP does not specifically recognize the poly(A) tail, the fact that it predominantly synthesizes poly(A) sequences must reside in the recognition of the incoming MgATP substrate. Here the studies of Balbo and Bohm provide a somewhat unexpected answer. While the adenine base of MgATP is π-stacked with that of the last adenylate in the poly(A), it is not specifically recognized by the enzyme as there are no direct hydrogen bonds to its N1 or N6 atoms. Instead, the specificity appears to be determined by several indirect factors:

(1) The π-stacking interaction favors purine over pyrimidine nucleotides. This explains the biochemical observations that MgCTP is a poorer substrate than MgATP, and that replacement of the last residue of the poly(A) by a C reduced the catalytic efficiency 10-fold (Balbo et al., 2005). Freshly cleaved pre-mRNAs generally have an A at the 3′ end.

(2) Discrimination between MgATP and MgGTP may be due in part to disruption of the water-mediated hydrogen-bonding network, and the fact that the C2 atom of adenine is placed close to the enzyme and it may not be able to accommodate the 2-amino group of guanine. However, kinetic data show that

the discrimination against MgGTP is due primarily to its 400-fold lower V_{max} value, while its K_m is similar to that for MgATP.

(3) The induced-fit behavior of PAP may play a more important role in defining the nucleotide specificity. Detailed analyses of kinetic data on wild-type Pap1p and several mutants distributed around the active site region, together with the structural data, led to an overall model for polyadenylation (Figure 1). The enzyme exists in an open form in the absence of substrates. Upon binding of MgATP and poly(A) (in random order), the enzyme undergoes a large conformational change to the closed form, which allows the adenylation reaction to occur. After catalysis, the enzyme returns to the open form and releases the products. In comparison, the open form is more stable in the ternary complex with MgGTP and poly(A), thereby making MgGTP a much poorer substrate and contributing to the nucleotide specificity of the enzyme. Mutations can directly affect substrate binding and/or indirectly affect the equilibrium between the open and closed forms of the enzyme.

The studies by Balbo and Bohm (2007) provide a clearer view of the structural basis of polyadenylation and also point to directions for further studies on this important enzyme. A catalytic residue is mutated to obtain the current structure, and only one Mg^{2+} ion is present in the active site. Therefore, the detailed chemical mechanism of the catalysis is not known. Also, it will be important to completely understand why MgGTP cannot induce the closure of the active site. Finally, PAP functions as part of a large protein machinery in pre-mRNA 3′-end processing, and the molecular mechanism for how other protein factors in this complex regulate PAP activity remains to be elucidated.

REFERENCES

Balbo, P.B., and Bohm, A. (2007). Structure *15*, this issue, 1117–1131.

Balbo, P.B., Meinke, G., and Bohm, A. (2005). Biochemistry *44*, 7777–7786.

Balbo, P.B., Toth, J., and Bohm, A. (2007). J. Mol. Biol. *366*, 1401–1415.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L., and Bohm, A. (2000). Science *289*, 1346–1349.

Edmonds, M. (2002). Prog. Nucl. Acid Res. Mol. Biol. *71*, 285–389.

Martin, G., Keller, W., and Doublie, S. (2000). EMBO J. *19*, 4193–4203.

Martin, G., Moglich, A., Keller, W., and Doublie, S. (2004). J. Mol. Biol. *341*, 911–925.

Zhelkovsky, A.M., Helmling, S., and Moore, C.L. (1998). Mol. Cell. Biol. *18*, 5942–5951.

Another Window into Disordered Protein Function

A. Keith Dunker[1,*]
[1]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine Program in Bioinformatics, School of Informatics, Indiana University, Indianapolis, IN 46202, USA
*Correspondence: kedunker@iupui.edu
DOI 10.1016/j.str.2007.08.001

Multiple crystal structures of the same proteins often have specific regions that switch between structure and disorder. In this issue of *Structure*, Zhang et al. (2007) show that these "dual personality fragments" are distinct from both structured and disordered protein and are functionally important.

Current biochemistry textbooks discuss the typical functions of globular proteins in terms of "lock and key" and "induced fit" models. The lock and key model depends on a structured protein with a rigid binding site, while the original induced fit model was described in terms of a structured protein with a flexible binding site that undergoes conformational change upon interaction with its ligand. Induced fit was later extended to include binding site changes resulting from domain shifts. For both the lock and key and the induced fit models, the formation of protein 3D structure is a prerequisite to function and can be described as the sequence → structure → function paradigm.

While still not discussed in biochemistry textbooks, another model for globular protein function has been in the focus of active theoretical and experimental research for a long time. In this alternative model, the protein starts as an interconverting ensemble under physiological conditions, a state that has been called "natively unfolded" (Weinreb et al., 1996), "intrinsically unstructured" (Wright and Dyson, 1999), and "disordered" coupled with various adjectives. Different functions, such as molecular recognition, occur as the disordered protein undergoes coupled binding and folding.

A number of prior studies focused on the prediction of disordered regions in a protein (Ferron et al., 2006). Within a given protein sequence, local