The ACT Domain: A Small Molecule Binding Domain and Its Role as a Common Regulatory Element*

Gregory A. Grant

From the Department of Molecular Biology and Pharmacology and the Department of Medicine, Washington University School of Medicine, St. Louis, Missouri 63110

The ACT domain is a structural motif in proteins of 70–80 amino acids that is one of a growing number of different intracellular small molecule binding domains that function in the control of metabolism, solute transport, and signal transduction (1–5). The first structure of an ACT domain was determined in 1995 with the crystal structure of Escherichia coli d-3-phosphoglycerate dehydrogenase (6), a tetrameric protein containing one ACT domain per subunit. This structure represents the archetypical ACT domain and is composed of four β strands and two α helices arranged in a βαβαβ fold as shown in Fig. 1. However, it was not recognized as a recurring motif until 1999 when Aravind and Koonin (2) proposed a PSI-Blast (position-specific iterating-Blast) sequence database search using the small subunit (IlvN) of acetolactate synthase. This search identified a diverse group of proteins that were noted to be involved in some way in amino acid and purine metabolism and were regulated by specific amino acids. They named this proposed domain the ACT domain. A PSI-Blast analysis of its sequence by Ettema et al. (3) revealed the ACT and RAM domains did not appear to select conserved consensus sequences and the PSI-Blast searches that may be related evolutionarily, but their similarity is intriguing. ACT domains. It is not known how these very similar domains may be related evolutionarily, but their similarity is intriguing.

A recent search of the “protein family data base” (Pfam) listed 3779 proteins as containing the ACT domain. These domains bind RNA through interaction at the face of their β-sheet structure rather than binding small molecules in loop regions like the ACT domains. It is not known how these very similar domains may be related evolutionarily, but their similarity is intriguing. The RNA binding domains have their own unique pattern of conserved consensus sequences and the PSI-Blast searches that identified the ACT and RAM domains did not appear to select any RNA binding domains.

**ACT Domain Sequence Homology**

A recent search of the “protein family data base” (Pfam) listed 3779 proteins as containing the ACT domain. These represent at least 57 different architectures (Pfam accession number PF01842), and ACT domain proteins are found in Bacteria, Archaea, and Eukaryota, including vertebrates, plants, fungi, and single cellular organisms. However, the ACT domain is not found in all organisms nor is it used consistently within species or in all proteins with like function (1–4). The sequence homology among the ACT domains of proteins is not immediately obvious and went undetected until the dimer interfaces of each are significantly different. The ACT domain dimers of phosphoglycerate dehydrogenase form a side-by-side structure producing an extended 8-stranded β sheet (Fig. 3A). On the other hand, the β sheets of the ACT domain dimers of the Lrp transcription factor assume a more face-to-face configuration (Fig. 3F). These observations formed the basis for the proposed division into ACT and RAM domains. As will be seen later in this review, recently determined structures demonstrate that the ACT domain shows an increasing diversity in tertiary and quaternary architecture as well as ligand binding interactions.

A novel type of ACT domain-containing protein family whose members contain ACT domain repeats has also been identified by sequence analysis in Arabidopsis (8). These proteins were termed ACR proteins. There are at least 8 genes in the 5 chromosomes of Arabidopsis that belong to the “ACR” protein family. Proteins similar to the ACR family have also been identified in rice (Oryza sativa) (9).

The majority of ACT domain-containing proteins appear to interact with amino acids and are involved in some aspect of regulation of amino acid metabolism (2–4) (Fig. 1). These include both metabolic enzymes and transcription regulators. In fact, the expression of some ACT-containing enzymes is under the control of ACT-containing transcription regulators. This has resulted in the ACT domain being referred to as “the regulatory domain in amino acid metabolism” in the SCOP (structural classification of proteins) data base. However, notable exceptions to this generality have been revealed in recent years. These include the NikR transcriptional regulator that binds nickel and functions in the regulation of intracellular nickel levels (10) and the YkoF protein (11) that binds thiamine and is thought to be involved in thiamine transport.

It is noteworthy that the 80–90-amino acid long ribonucleoprotein motif of RNA binding proteins (12, 13) also possesses the same βαβαββ fold as the ACT domain. These domains bind RNA through interaction at the face of their β-sheet structure rather than binding small molecules in loop regions like the ACT domains. It is not known how these very similar domains may be related evolutionarily, but their similarity is intriguing. The RNA binding domains have their own unique pattern of conserved consensus sequences and the PSI-Blast searches that identified the ACT and RAM domains did not appear to select any RNA binding domains.
MINIREVIEW: ACT Domain

more sensitive sequence comparison methods, such as PSI-Blast, were developed. Because of the large number of alignments now available, they can only be presented here in a limited fashion. The reader is referred to the original articles (2–5) as well as the Pfam data base for additional information. Fig. 2 presents the sequence alignments of representative ACT domains including all of those whose structures have been determined. The structures provide the additional advantage of considering the secondary structure elements in the alignment. The amino acid residues are highlighted by residue type following the secondary structure elements as listed at the top of the figure. The color scheme generally follows that of Aravind and Koonin (2) and then according to the RAM domain alignment of Ettema et al. (3). When the pattern of secondary structure is also taken into consideration, it can be seen that sequences of the RAM domain group are not inconsistent with the original ACT domain alignment (2). To illustrate this, Fig. 2 presents four putative RAM domains highlighted first according to the ACT domain alignment of Aravind and Koonin (2) and then according to the RAM domain alignment of Ettema et al. (3). A common characteristic of the initial alignment of Aravind and Koonin (2) was the presence of a nearly invariant glycyl residue at the turn between the first β strand and the first α helix that coincided with the binding site for L-serine in phosphoglycerate dehydrogenase. This led them to propose a common ligand binding mode for all ACT domains.

The RAM domain proteins are missing the glycyl residue in this location. This coincides with the presence of another highly conserved glycyl residue in the loop between the β2 and β3 strands. Although the RAM domain proteins share many features of the ACT domain alignment, the β2-β3 loop area does appear to be rather distinctive for the RAM grouping (Fig. 2).

ACT Domain Structures

To date, the structures of at least 10 ACT domain-containing proteins have been determined (Fig. 3), and five of these have been solved with bound ligand. Even with this limited number of structures, they display a remarkable diversity in how the domains associate.

The structures range from stand-alone ACT domains, as seen for the YbeD protein from E. coli (14) (Fig. 3K), to large multimeric proteins with ACT domains either isolated or associating into groups of 2, 3, or 4 (6, 7, 10, 11, 14–19, 21, 22). The first structures determined, and the most numerous, are the monomeric proteins with a single ACT domain. Of these, the PheOH, H3O, Lrp, and IlvR proteins (23) have been determined in complex with bound ligand.

FIGURE 1. The ACT domain fold. Left, a partial list of proteins containing the ACT domain and their regulatory ligands; right, a single ACT domain from E. coli d-3-phosphoglycerate dehydrogenase showing the (βββαβ) motif. In the diagram, the triangles point up or down to indicate the direction of the strand in the β sheet. Numbers progress from N terminus to C terminus.

FIGURE 2. Sequence alignment of selected ACT domains. The sequences are aligned with regard to residue type and secondary structure. The upper group is a sampling of ACT domains using the alignment scheme shown by Aravind and Koonin (2). The middle group is a sampling of ACT domains that were classified as RAM domains by Ettema et al. (3) using the same alignment scheme as in the upper group. The lower group has the same proteins as in the middle group but uses the alignment scheme presented by Ettema et al. (3) for RAM domains. The residues numbers of the ACT domains within the protein sequence are given, and the asterisk denotes that the structure has been determined (see Fig. 3). The residues in bold designate those within the secondary structure elements as listed at the top of the figure. The color scheme generally follows that of Aravind and Koonin (2) and Ettema et al. (3) and refers to the following residue types: green, hydrophobic (ILVCGMFYWTP); red, conserved glycine or glycine next to ligand binding site. The protein abbreviations are: PGDH, d-3-phosphoglycerate dehydrogenase; H3O, the regulatory subunit of acetohydroxyacid synthase; PheOH, phenylalanine hydroxylase; IlvR, aspartate kinase; Lrp, glycine cleavage system transcription regulator; ATPPR, ATP-phosphoribosyltransferase; Lrp, NikR transcription regulator; YbeD, YbeD binding protein; ActR, first ACT repeat of ActR protein; ThrD, threonine deaminase; Lrp, Lrp transcription regulators; PHS023, PHS023 protein; LeuA, a -isopropylmalate synthase. The species abbreviations are: E. coli; rat, Rattus norvegicus; Ath, Arabadopsis thaliana; Vch, Vibrio cholerae; Mbt, Mycobacterium tuberculosis; Bsu, Bacillus subtilis; Pho, P. furiosus; Pho, P. harikoshii; Sso, Sulfolobalus solfataricus; Pmu, Pasturella multocida.

<table>
<thead>
<tr>
<th>Protein/ ap</th>
<th>residues</th>
<th>beta 1</th>
<th>alpha 1</th>
<th>beta 2</th>
<th>beta 3</th>
<th>alpha 2</th>
<th>alpha 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGDH Ec</td>
<td>338–407*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>IlvR-1 Ec</td>
<td>2–75*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>PheOH rat</td>
<td>35–110*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>AspKl Ath</td>
<td>346–419*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>Glycle 1 Vch</td>
<td>5–73*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>Glycle 2 Vch</td>
<td>92–175*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>ATPPR mtb</td>
<td>211–277*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>NikR Ec</td>
<td>51–120*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>YbeD Ec</td>
<td>15–86*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>YkoF-1 Bau</td>
<td>11–84*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>YkoF-2 Bau</td>
<td>117–189*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>ACR 1 Ath</td>
<td>36–103</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>ThrD Ec</td>
<td>338–407*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>LrpA Pfu</td>
<td>63–137*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>PHS023 Pho</td>
<td>1–75*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>LysA-2 Sau</td>
<td>381–457</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>Lrp Pau</td>
<td>70–147</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>LrpA Pfu</td>
<td>63–137*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>PHS023 Pho</td>
<td>1–75*</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>LysA-2 Sau</td>
<td>381–457</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
<tr>
<td>Lrp Pau</td>
<td>70–147</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td>92815</td>
<td></td>
</tr>
</tbody>
</table>
domain because it is not formed from a sequential fold. It also does not bind a ligand. It seems to play a purely structural role, being part of the dimer interface.

*E. coli* threonine deaminase is particularly interesting because it contains an ACT domain followed closely, within the same chain, by a $\beta\alpha\beta$ fold that together forms a seven-stranded $\beta$ sheet (19). Mutagenesis studies (20) indicate that the effector ligand (isoleucine) binding site is located at the interface between these two domains in a position analogous to $\gamma$-serine in phosphoglycerate dehydrogenase.

The rat phenylalanine hydroxylase (21) (Fig. 3G) structure is a dimer where the two ACT domains do not associate with each other. However, the protein in this case is lacking the C-terminal 24 residues that are responsible for tetramer formation. Thus, a tetrameric structure could eventually show a dimeric interaction of ACT domains similar to some of the other structures.

Proteins whose ACT domains associate with their $\beta$ strands in a more or less face-to-face arrangement include the Lrp-like transcription regulator (7), the NikR transcription regulator (10, 15), the glycine cleavage system repressor protein (unpublished data, Protein Data Bank code 1U8S), ATP phosphoribosyltransferase (16, 17), and the YkoF thiamine binding protein (11) (Fig. 3, J, D, I, C, and F, respectively). However, among these there is notable variation in the relative orientation of the domains. The $\beta$ strands of the Lrp regulator face each other with a diagonal arrangement whereas those of the NikR regulator display more or less face-to-face arrangements. ATP phosphoribosyltransferase displays a face-to-face arrangement of three ACT domains rather than two.

Two of the proteins, the YkoF and the glycine cleavage regulatory subunit of acetohydroxyacid synthase from *E. coli* (18) (Fig. 3H), possess both the face-to-face and the side-by-side domain arrangements. Both proteins consist of two tandem ACT domains in each of two chains. They form the side-by-side arrangement between ACT domains within the same polypeptide and a face-to-face arrangement across the subunit interface.

**FIGURE 3.** Structures of ACT domains and ACT domain proteins. A, *E. coli* $\alpha$-3-phosphoglycerate dehydrogenase tetramer (Protein Data Bank code 1PSD). B, *A. thaliana* aspartate kinase dimer (Protein Data Bank code 2CDQ). C, *M. tuberculosis* ATP phosphoribosyltransferase hexamer (Protein Data Bank code 1NH8). D, *E. coli* NikR transcription regulator tetramer (Protein Data Bank codes 1Q5V and 1Q5Y). E, *E. coli* threonine deaminase tetramer (Protein Data Bank code 1TDJ). F, *B. subtilis* YkoF thiamine binding protein dimer (Protein Data Bank code 1U8S). G, *rat* phenylalanine hydroxylase dimer (Protein Data Bank code 1PHZ). H, *E. coli* IlvH regulatory domain dimer of acetohydroxyacid synthase (Protein Data Bank code 2F1F). I, the view on the left shows the side-by-side extended $\beta$ sheet of the upper domains and the offset in the face-to-face arrangement of the lower domains. The view on the right is rotated 90°. J, *V. cholerae* D-3-phosphoglycerate dehydrogenase (21). K, the YkoF thiamine binding protein dimer (Protein Data Bank code 1U8S). J, P. furiosis Lrp-like transcription regulator dimer (Protein Data Bank code 1115). K, *E. coli* YbeD protein stand-alone ACT domain monomer (Protein Data Bank code 1RWU). Within the proteins the subunits are colored gray, green, and aqua with the ACT domains of each subunit colored red, dark blue, and dark green, respectively. For the YkoF protein (F), the IlvH protein (H), and the glycine cleavage system protein (I), the tandem ACT domains in each subunit are colored red and dark blue. The association of the ACT domains is depicted in more detail for proteins A–E. The arrows illustrate their approximate location in the structure. The $\beta$ strands are in aqua and the helices in red. The ligands are depicted in CPK format.
MINIREVIEW: ACT Domain

(Fig. 3, F and I). However, as can be seen in Fig. 3, they do not adopt the same type of overall structure. The IlvH polypeptide also consists of two $\beta$A$\beta$A$\beta$ folds in tandem. The first of these, residues 2–75 from each chain, associate into a side-by-side extended sheet structure whereas the second, residues 81–150, forms a more face-to-face association. However, in this case the spacing is wider than in the face-to-face association in the Lrp protein (Fig. 3f), and the domains are offset so that their $\beta$ sheets are not directly across from each other. Furthermore, the sequence alignment of the second domain in the IlvH protein (residues 81–150) does not match either the ACT or RAM alignments as well as the first. Although the glycine residue between $\beta$1 and $\alpha$1 is present in the IlvH second domain, it is not well conserved throughout other species (18), and there is no conserved glycine found between $\beta$2 and $\beta$3. These second sets of domains may have lost their ligand binding function in favor of forming higher order quaternary structure. The crystal lattice of the IlvH protein shows the dimers forming octamers through association of this second domain with magnesium ions.

Ligand Binding Sites

The five structures that have been solved with ligands indicate that the ligands tend to bind at the interfaces between ACT domains. There appears to be a correlation of the ligand binding sites with specific glycine residues located in loops between the helices and sheets. $\gamma$-Serine binds to the phosphoglycerate dehydrogenase ACT domains primarily at the loop between $\beta$1 and $\alpha$1. This location is also where $\gamma$-lysine binds in aspartate kinase and has been determined by mutagenesis to be the probable location of $\gamma$-valine binding to IlvH (18). The Arabidopsis aspartate kinase ACT domain also binds S-adenosylmethionine at a second site in the $\beta$3-$\alpha$2 loop that seems to act synergistically with the lysine. This is the only known case of an ACT domain binding more than one kind of ligand. These three proteins are also the only ones that exhibit the extended 8-stranded $\beta$ sheet structure and possess a conserved glycyl residue in the $\beta$1-$\alpha$1 loop region. Threonine deaminase, which has an ACT-like side-by-side dimer, also possesses a glycine residue in this location. The sequences identified as representing RAM domains (3) contain a nearly invariant glycyl residue in the loop between the $\beta$2 and $\beta$3 strands. This is the region that mutagenesis data suggest is involved in ligand binding in the Lrp-like regulator.

ATP-phosphoribosyltransferase contains neither glycyl residue and is the only available structure where the ACT domains form a triad arrangement. In this protein, the substrate, $\alpha$-histidine, appears to associate mainly with the loop between the $\alpha$1 helix and $\beta$2 strand. Interestingly, there is a glycyl residue found in this region that is positioned next to the bound histidine.

The YkoF structure is somewhat different in that it contains two successive ACT domains on the same polypeptide chain. Two of these chains associate to form a dimer that binds four ligand molecules. The main determinants for binding each ligand appear to be contributed largely by only one of the chains because there is little apparent hydrogen bonding of the ligand to the adjacent chain as is seen in most other ACT domain proteins. However, the adjacent chain does have a glycyl residue (Gly–88) in the loop between the ACT domains that is next to where two of the ligands bind. The second two ligand binding sites are opposite the C terminus of the adjacent chains. There is no glycyl residue present here, but there is also less structural obstruction at the terminus of the chain.

The NikR transcription regulator appears to be an exception. The nickel ion binds at the loop between $\beta$2 and $\beta$3, similar to the putative ligand binding location in the Lrp-like transcription regulator (RAM domain), but instead of having a glycyl residue at this location, the nickel chelates to the side chain of a histidine residue found in the analogous position.

Thus, there are at least four different ligand binding motifs that can potentially be discerned from the five available ligand-bound structures and the PSI-Blast sequence alignments. All but one involve the accommodation of enough space for the ligand to occupy the pocket by the absence of a side chain (Gly). The other is the only instance so far of a metal ion binding to an ACT domain by direct chelation with amino acid side chains.

ACT Domain Mechanism

The mechanism by which ACT domains translate the ligand binding event to the active site of enzymes or to the protein-DNA binding site in transcriptional regulators is unclear. Some metabolic enzymes containing ACT domains are inhibited by the ligands that bind whereas others are activated. Some show K-type allosteric effects whereas others display V-type effects. The uneven distribution of ACT domain proteins among and within species suggests that the ACT domain modules may have been recruited independently by each protein. If this is the case, do the ACT domains also function differently from protein to protein or is there a common mechanism?

Crystal structures of E. coli d-3-phosphoglycerate dehydrogenase (6, 23) illustrate that a major effect of serine binding is the retardation of a rotation of the substrate binding domain relative to the nucleotide binding domain. The binding of phenylalanine to phenylalanine hydroxylase, along with phosphorylation of Ser-16, appears to stabilize the structure so that the binding pocket remains open to accept the substrate (21). Histidine binding to ATP-phosphoribosyltransferase synergistically enhances the inhibition of the enzyme by AMP and ADP, and is associated with the conversion of a dimeric form to a hexamer (16, 17). Stabilization of the hexamer by histidine results in the active sites being partially closed and inaccessible. Crystal structures of the NikR protein from E. coli (10) and Pyrococcus horikoshii (15) have led to the suggestion that nickel binding may serve to stabilize the structure, reducing its flexibility, and thus may promote DNA binding.

If there is a common theme in the regulation of ACT domain proteins, it may be that ligand binding reduces the flexibility of the structures. However, ACT domain proteins are only beginning to be studied in detail, and there is not yet enough information available.

An apparent third type of ACT domain proteins are those that consist of only the ACT domain, either as stand alone proteins or with multiple copies. Examples of these are the YbeD protein and the YkoF thiamine binding protein. One possible function, as has been postulated for YkoF, is that they may function as ligand transporters. Alternatively, they may function to sequester the ligands until their intracellular levels fall to a point that they are needed.
Summary

ACT domains are found in a wide variety of proteins and in a variety of different arrangements. The majority of these bind amino acids that function to regulate some aspect of amino acid metabolism. Most ACT domain-containing proteins have been discovered in Bacteria and Archaea, indicating that the domain appeared very early on. At least one functional example has been found in a mammalian protein, rat phenylalanine hydroxylase. Interestingly, ACT domains are not found in all organisms nor are they used consistently within species or found in all proteins possessing a common activity. In their original article, Aravind and Koonin (2) suggested that the ACT domain represents a “conserved, evolutionarily mobile module” that when fused to other proteins made them susceptible to regulation by small molecules. The large variety of ACT domain-containing proteins is consistent with this idea, but they do not appear to be used universally within a single species. The preponderance of ACT domains that bind amino acids suggests that this was its original function. However, there are obvious examples that it is evolving to bind other ligands and perhaps participate in the regulation of other processes. One should anticipate that the realm of the ACT domain as a regulatory element will continue to expand as more structures are determined and more is learned about their respective functions. What was once just a very short story is developing into a drama with many ACTs.

Acknowledgment—I thank Dr. Peter Chivers for critical reading of the manuscript and for helpful comments.


REFERENCES