Understanding a Transcriptional Paradigm at the Molecular Level

**THE STRUCTURE OF YEAST Gal80p**

Received for publication, November 3, 2006, and in revised form, November 15, 2006

**James B. Thoden**, Christopher A. Sellick, Richard J. Reece, and Hazel M. Holden

From the Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706 and the Faculty of Life Sciences, The University of Manchester, The Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom

In yeast, the GAL genes encode the enzymes required for normal galactose metabolism. Regulation of these genes in response to the organism being challenged with galactose has served as a paradigm for eukaryotic transcriptional control over the last 50 years. Three proteins, the activator Gal4p, the repressor Gal80p, and the ligand sensor Gal3p, control the switch between inert and active gene expression. Gal80p, the focus of this investigation, plays a pivotal role both in terms of repressing the activity of Gal4p and allowing the GAL switch to respond to galactose. Here we present the three-dimensional structure of Gal80p from Kluyveromyces lactis and show that it is structurally homologous to glucose-fructose oxidoreductase, an enzyme in the sorbitol-glucanate pathway. Our results clearly define the overall tertiary and quaternary structure of Gal80p and suggest that Gal4p and Gal3p bind to Gal80p at distinct but overlapping sites. In addition to providing a molecular basis for previous biochemical and genetic studies, our structure demonstrates that much of the enzymatic scaffold of the oxidoreductase has been maintained in Gal80p, but it is utilized in a very different manner to facilitate transcriptional regulation.

The GAL genetic switch (supplemental Fig. S1) in both the baker's yeast *Saccharomyces cerevisiae* and the milk yeast *Kluyveromyces lactis* is composed of an activator (Gal4p), an inhibitor (Gal80p), and a ligand sensor (Gal3p in *S. cerevisiae* or Gal1p in *K. lactis*). These proteins in the two yeasts are, at least in part, interchangeable. For example, Gal4p from both *S. cerevisiae* (ScGal4p) and *K. lactis* (KlGal4p) will complement a gal4 mutation in either yeast (1, 2) despite the two proteins sharing comparatively little overall sequence similarity (28% amino acid identity and 57% similarity over their entire length). Gal80p from either yeast are highly related (58% amino acid identity and 82% similarity) and will inhibit the transcriptional activity of either version of Gal4p. However, while KlGal1p can complement both a Sgal1 (galactokinase-defective) and a Sgal3 (ligand sensor-defective) mutation (3), ScGal3p cannot complement the non-inducible phenotype of a Klgal1 deletion mutant unless the KlGAL80 gene is also replaced by ScGAL80 (4).

Studies have suggested that there are differences in the cellular locations of the three GAL regulatory proteins in the two yeasts and consequently potential differences in the mechanism of transcriptional activation. In both cases, Gal4p is presumed to be nuclear, while ScGal80p can be found in both the nucleus and the cytoplasm and is capable of shuttling between the two. ScGal3p is predominately, if not exclusively, cytoplasmic (5, 6). Recently, KlGal80p has been identified as an exclusively nuclear protein while KlGal1p is present both in the nucleus and the cytoplasm (7). On the basis of these and other data it has been suggested that the *S. cerevisiae* GAL switch is activated when galactose and ATP bind to Gal3p in the cytoplasm. This traps Gal80p in the cytoplasm thereby freeing Gal4p from its repressing effects and allowing transcriptional activation to occur (8). In *K. lactis*, the GAL switch appears to be controlled by competition in the nucleus between KlGal1p and KlGal4p for KlGal80p binding (7). In either model, Gal80p must interact with two very different proteins: Gal4p, via its carboxyl-terminal activation domain, and Gal3p (Gal1p in *K. lactis*) when galactose and ATP are bound to the ligand sensor. Here, we describe the molecular structure of KlGAL80p and propose a model for its interaction with the activator, Gal4p, and with the ligand sensor, Gal1p.

**EXPERIMENTAL PROCEDURES**

X-ray Structural Analysis—KlGal80p was cloned, expressed, and purified according to standard procedures (described in the supplemental "Experimental Procedures"). The protein utilized for crystallization contained a His-tag at the NH2 terminus with the following sequence: MGSSHHHHHHSEKDELFSQGH. Crystals of both the wild-type and selenomethionine-labeled protein were obtained from 18 to 22% 2-methyl-2,4-pentanediol and 100 mM MES (pH 6.25) at 4 °C. They belonged to the space group P21212 with unit cell dimensions of a = 112.2 Å, b = 137.1 Å, and c = 72.9 Å. The asymmetric unit contained one dimer. X-ray data from flashed-cooled crystals were collected at the Structural Biology Center Beamline 19-BM (Advanced Photon Source, Argonne National Laboratory, Argonne, IL). These data were processed and scaled with HKL2000. The software package SOLVE (9) was used to locate the positions of 10 selenium atoms and to generate initial protein phases. Solvent flattening and averaging with RESOLVE (10) resulted in an interpretable electron density map for the selenomethionine-labeled protein at 3.0 Å resolution. The model based on this electron density map was subsequently used to refine...
against x-ray data collected from a wild-type crystal to 2.1 Å resolution with the software package TNT (11). X-ray data collection and model refinement statistics are presented in supplemental Tables S1 and S2, respectively.

Light Scattering—Protein necessary for light scattering experiments was prepared according to standard procedures (described in the supplemental “Experimental Procedures”). Protein samples (200 µl at 20 °C containing 3.5 nM of each protein) were run on a Superdex 200 10/300 GL column (Amersham Biosciences) equilibrated and run (at 0.7 ml/min) in buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl. Protein was detected using a DAWN EOS 18-angle laser photometer coupled to an Optilab rex refractive index detector with a Wyatt QELS system. For the interaction between KlGal1p and KlGal80p, samples additionally contained 100 mM galactose and 2.5 mM ADP.

RESULTS AND DISCUSSION

The x-ray structure of KlGal80p was solved to 2.1 Å resolution and refined to an overall $R_{work}/R_{free}$ of 19.5%/24.5%. Overall, the electron density for the protein was well ordered except for two short loop regions (Asp-245 to Gly-248 and Asp-309 to Ser-316 in Subunit I and Asn-247 to Arg-250 and Gly-311 to Ser-316 in Subunit II) and two larger regions that are located on the same side of the molecule (Gly-328 to Glu-362 and Leu-394 to Lys-413 in Subunit I and Gly-328 to Glu-361 and Gly-395 to Lys-413 in Subunit II).

The Gal80p subunit adopts a two-domain architecture (Fig. 1a). The NH$_2$-terminal domain is formed by Ala-1 to Leu-151 and consists of a classical Rossmann fold with six strands of parallel β-sheet flanked on one side by two α-helices and by three on the other. An additional α-helix from the COOH-terminal domain (Ser-376 to Phe-393) completes the helix/sheet packing of the Rossmann fold (Fig. 1a). There is a decided β-bulge in the second strand of the dinucleotide-binding motif at Val-50. A cis-alanine at position 124 lies in the random coil region connecting the fifth β-strand to the fifth α-helix of the Rossmann fold. The COOH-terminal domain of Gal80p (Gln-152 to Ile-457) is dominated by a nine-stranded mixed β-sheet with an overall width of ∼37 Å. This domain also contains a small three-stranded mixed β-sheet, four α-helices, and a helical turn (Fig. 1a).

Gal80p packs in the crystalline lattice as a dimer with overall dimensions of $\approx 55 \times 75 \times 110$ Å (Fig. 1b). From light scattering experiments, we have demonstrated that KlGal80p in solution is exclusively dimeric even in the absence of EDTA as opposed to that previously reported in (7) (Table S3). The large mixed β-sheet of the COOH-terminal domain is intimately involved in the formation of the dimer, whereas the Rossmann fold motifs are distributed on opposite sides of the protein. The subunit:subunit interface is extensive with a total buried surface area of 4400 Å$^2$.

A search with DALI (12) reveals that the closest structural relatives of Gal80p include Zymomonas mobilis glucose-fructose oxidoreductase (13), rat biliverdin reductase (14), and Leuconostoc mesenteroides glucose-6-phosphate dehydrogenase (15). A comparison of the α-carbon traces for Gal80p and Z. mobilis oxidoreductase (Fig. 2a) demonstrates that these two proteins superimpose with a root-mean-square deviation of 2.0 Å for 307 structurally equivalent α-carbons, which is remarkable given their limited amino acid sequence identity of ∼13% (16). It has been postulated that Lys-129 and Tyr-217 are involved in the catalytic mechanism of the oxidoreductase (13). These residues correspond to Trp-123 and His-214 in KlGal80p. In addition to the
above-mentioned enzymes of known function, there are several putative oxidoreductases whose models have been deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank with similar three-dimensional architectures. Most of these proteins have a cis-proline in the same position as the cis-alanine (Ala-124) in Gal80p. In those proteins with bound NAD(P), the cis-proline is typically preceded by a lysine residue whose ε-nitrogen forms a hydrogen bonding interaction with the 2′-hydroxyl of the nicotinamide ribose. There are other examples, such as porcine malate dehydrogenase (17), where the cis-proline is preceded by an asparagine residue whose side chain, again, interacts with the nicotinamide ribose. In Gal80p the corresponding residue is a tryptophan, which precludes a similar interaction.

One of the characteristic signature sequences for a Rossmann fold is Gly-X-Gly-X-Gly/Ala, which connects the first β-strand to the first (or dinucleotide-binding) helix. It is the second glycine of this sequence that packs against the phosphoryl groups of the NAD(P). A superposition of these regions in the seven closest structural relatives of Gal80p is presented in Fig. 2b. In Gal80p, there is a three-residue insertion in this loop, which results in a markedly different conformation. Indeed, in Gal80p the second glycine of the signature sequence is replaced with Thr-26. Given that the side chain of Thr-26 is too bulky to lie against the phosphoryl groups of NAD(P) and that the typical lysine residue that hydrogen bonds to the nicotinamide ribose is replaced with Trp-123 in Gal80p, it is unlikely that it binds NAD(P), at least in the orientation observed in other family members.

Several recent and insightful mutational analyses of ScGal80p have been performed (16, 18). In one of these studies, variants of the protein from *S. cerevisiae* were uncovered that were defective only in Gal4p- or in Gal3p binding (16). Given the amino acid sequence homology between the *S. cerevisiae* and *K. lactis* proteins, we have mapped these mutations onto the KlGal80p structure (Fig. 2a). Those mutations that give rise to defective Gal4p binding to the ScGal80p and that are visible in the KlGal80p model are located at positions Gly-153, Gly-184, Asp-261, His-262, Gly-283, and Leu-320. Most of these are located at the subunit-subunit interface. The mutations at Gly-153 and Gly-184, however, are particularly interesting because they are separated by ~17 Å and lie on either side of a large cleft formed by the COOH-terminal end of the β-sheet in the Rossmann fold and an α-helix defined by Ser-211 to Ile-222. Two additional mutations have been identified in ScGal80p that result in defective Gal4p-binding: A309T and G310D (16). The corresponding amino acids in...
KI-Gal80p, namely Ala-310 and Gly-311, reside in a disordered surface loop of six residues that connects two anti-parallel $\beta$-strands and that is situated at the top of the cleft (Fig. 3b). In glucose-fructose oxidoreductase there is a three-residue deletion in the loop, which folds in toward the nicotinamide ring of the NADP. Indeed, it is this cleft region that in glucose-fructose oxidoreductase and similarly related enzymes harbors both the NADP and the catalytic machinery. Strikingly, the cleft in the oxidoreductase is not nearly as wide as that observed in Gal80p. As an example, the $\alpha$-carbons of Lys-41 and Ala-196 of the oxidoreductase are separated by $\sim 6 \, \text{Å}$, and the side chain of Asp-194 forms a salt bridge with Lys-41, which further closes off the cleft. In Gal80p that gap is much wider with the $\alpha$-carbons of Trp-31 and Ser-194, for example, being $\sim 14 \, \text{Å}$ apart. Additionally there are no apparent salt bridges to close the gap. Given the location of the Gal4p-only defective mutants in Gal80p and the three-dimensional characteristics of the cleft, we would suggest that this region forms the binding site for Gal4p.

There has been considerable speculation concerning the structure of the activation domain of Gal4p, which is known to be coincident with the Gal80p interaction site (19–21). Experiments have demonstrated that in ScGal4p, the carboxyl-terminal 30 residues are recognized by Gal80p (22). Some structural predictions have suggested that these residues lie in an $\alpha$-helix, and it was inferred that the activation domain of Gal4p, and hence the region that interacts with Gal80p, may be helical. Other studies, in marked contrast, have suggested that the activation domain of Gal4p is $\beta$-sheet at low pH and is essentially unstructured at physiological pH (23). Using current structure prediction algorithms for peptides, the corresponding residues in KlGal4p (amino acids 836–865: TNNFLNPSTQQLNFTTMDDVYNYIFDNDNE), are predicted to have $\alpha$-helical character. Employing an $\alpha$-helix in Gal4p for binding to Gal80p makes structural and chemical sense. In an $\alpha$-helix the hydrogen bonding capacity of the backbone carbonyl groups and amide nitrogens is mostly satisfied. On the other hand, in a $\beta$-hairpin motif, for example, the backbone hydrogen bonding pattern would not be completely satisfied if it were to bind into the type of cleft observed in Gal80p which is devoid of $\beta$-sheet. On the basis of both secondary structural predictions and the nature of the Gal80p putative binding cleft, we predict that the COOH-terminal 30 residues of KlGal4p most likely bind into the Gal80p cleft (Fig. 3b) as an $\alpha$-helix. Co-crystallization experiments with Gal80p and a peptide representing the COOH terminus of KlGal4p are in progress to address this issue.

The mutations in ScGal80p that are defective in only Gal3p binding are clustered and correspond to Gly-302, Gly-324, Glu-362, and Val-368 in KlGal80p. These mutations map to the structure at the edge of the mixed $\beta$-sheet in the COOH-terminal domain and toward the surface (Fig. 3, a and b). Importantly, these residues are located near the large disordered region between Gly-328 and Gly-362. We propose that these residues mark the binding surface for KlGal1p and that the disordered region in Gal80p becomes ordered upon complexation with the ligand sensor. Note that both of the proposed binding sites for Gal4p and Gal1p (or ScGal3p) are on the same side of the Gal80p dimer (Fig. 3b), and this is in keeping with the recent and elegant experiments of Anders et al. (7), which suggest that the binding sites on Gal80p for Gal1p or Gal4p are overlapping.

Both the crystal structure (Fig. 1b) and biochemical assays in solution (Fig. 4) indicate that KlGal80p is dimeric. The complex between KlGal80p and Gal4p (Fig. 4a) is most easily interpreted as dimers of each protein interacting with one another. From our model, we predict that the activation domains of Gal4p pack into the Gal80p clefts (Fig. 4c). Our data, and in contrast to that previously published (7), suggest that KlGal1p may exist in
both a monomeric and dimeric state (Fig. 4b). However, the interaction between KlGal80p and KlGal1p would appear to occur via a 2:1 complex, although it is possible that, at high concentration, a dimer of Gal80p interacts with two molecules of Gal1p (7). The proposed binding sites for both Gal4p and Gal1p in KlGal80p are distinct but are located on the same side of the molecule. We believe that it is therefore possible that the binding of one partner excludes the binding of the other.

It has become increasingly apparent in recent years that the molecular scaffolds employed by enzymes of sugar metabolism are ideally suited to function as transcriptional regulators as well. The structure of Gal80p, even with its limited amino acid sequence homology, is remarkably similar to glucose-fructose oxidoreductase. Likewise, Gal3p, the transcriptional inducer in S. cerevisiae, demonstrates a ~90% sequence similarity to Gal1p, the bona fide galactokinase of the Leloir pathway that has, itself, function as a ligand sensor (24–30). The molecular architecture of UDP-galactose 4-epimerase, another enzyme in galactose metabolism (31), has now been observed in NmrA, the structure presented here provides new molecular details regarding the architecture of UDP-galactose 4-epimerase, another enzyme in a fundamental cellular property. The structure of external conditions and signals at the transcriptional level is analyzed at the genetic and biochemical level for nearly 50 years. Indeed, the ability of an organism to respond to a variety of external conditions and signals at the transcriptional level is a fundamental cellular property. The structure presented here provides new molecular details regarding the GAL transcriptional switch, which, to date, is the best understood system for eukaryotic transcriptional control.

Acknowledgments—We thank R. Zhang and F. Rotella for their assistance at APS and T. Jowitt and M. Howard at the University of Wisconsin-Madison. The insightful comments of A. Steinberg, D. Timson, and I. Rayment are gratefully acknowledged.

REFERENCES