Implication of Tubby Proteins as Transcription Factors by Structure-Based Functional Analysis

Titus J. Boggon,[1] Wei-Song Shan,[2] Sandro Santagata,[3] Samuel C. Myers,[1] Lawrence Shapiro[1]*

Tubby-like proteins (TULPs) are found in a broad range of multicellular organisms. In mammals, genetic mutation of tubby or other TULPs can result in one or more of three disease phenotypes: obesity (from which the name "tubby" is derived), retinal degeneration, and hearing loss. These disease phenotypes indicate a vital role for tubby proteins; however, no biochemical function has yet been ascribed to any member of this protein family. A structure-directed approach was employed to investigate the biological function of these proteins. The crystal structure of the core domain from mouse tubby was determined at a resolution of 1.9 angstroms. From primarily structural clues, experiments were devised, the results of which suggest that TULPs are a unique family of bipartite transcription factors.

The *tub* gene was initially identified by positional cloning of the genetic locus responsible for the autosomal recessive obesity syndrome in the *tubby* strain of obese mice (*1, 2*). *Tub*$^{-/-}$ mice (*3*) also exhibit progressive sensorineural degeneration of the retina and cochlear hair cells, the primary sensory neurons involved in hearing (*4*). Similarly, mutation of the human *TULP1* gene is the genetic origin of retinitis pigmentosa type 14 (RP-14), which can progressively lead to total blindness (*5, 6*). These sensory defects arise as the result of neuronal cell death by apoptosis (*4, 7, 8*). The tubby obesity syndrome may also arise from cell death in the appetite control center of the hypothalamus, where the tubby protein is highly expressed (*5, 9*), but this has not been demonstrated experimentally.

Tubby is the founding member of a multigene protein family. At least four tubby-like proteins (tubby and TULPs 1 through 3) are conserved among different species of mammals, and tubby-like proteins are also found in other multicellular organisms (*9*), including plants (*1*). They have not, however, been identified in single-celled organisms. These proteins feature a characteristic ~270–amino acid "tubby domain" at the COOH-terminus that does not exhibit homology to other known proteins. Most tubby proteins include NH_2-terminal regions that, in general, are not closely related to one another. These NH_2-terminal regions, however, are often similar in cross-species orthologs. They comprise about 180 to 280 amino acids in tubby and TULPs 1 through 3.

Recent technological developments have opened a new possibility for the application of structural biology: to efficiently answer biological questions at an early stage of their investigation, where questions of protein function are more appropriately phrased as "What does it do?" rather than "How does it do it?" (*10*). The revolution in genomics has provided comprehensive lists of potential biological actors in numerous systems and organisms, and the technological revolution in crystallography, particularly in multiwavelength anomalous diffraction (MAD) analysis (*11*), has enabled a quantum leap in the ease and rapidity of structure solution. Our knowledge and understanding of biological structures has expanded to the degree that visualizing the three-dimensional structure of a new protein can often point the way to biological insights that would have been difficult to find by other types of analysis (*12*). We have applied such a "structure-based functional genomics" approach to tubby-like proteins. This approach enabled us to efficiently identify the likely biochemical mechanism of this protein family.

Crystal Structure of the Tubby Core Domain

As the first step in characterizing the biological function of tubby, we expressed its COOH-terminal core domain (residues 243 through 505) (Fig. 1) in a recombinant expression system (*13*) and determined its high-resolution crystal structure by MAD analysis of the selenomethionyl (SeMet) protein (*14*) (Tables 1, 2, and 3). The structure reveals a striking fold in which a central hydrophobic helix at the COOH-terminus wholly traverses the interior of a closed 12-stranded β barrel (Fig. 2). This arrangement is unique among known protein structures.

The tubby β barrel adopts an alternating up-down nearest-neighbor topology, such that hydrogen bonding is in the antiparallel mode for all strands. We have numbered the strands of the barrel as 1 through 12 in sequence order. Several excursions in the loops between these strands are observed. A three-stranded β sheet intervenes in the 9 and 10 connection, and we have thus designated these strands as 9A, 9B, and 9C. Similarly, four helices, H4, H6A, H6B, and H8, are found in the corresponding loop regions between strands of the main barrel. The barrel is slightly oblong, with C_α to C_α widths across it varying between 18 and 22 Å. From top to bottom, the barrel measures ~18 Å. The whole domain has maximum dimensions of 40 Å in the direction parallel to the central helix H12, and 51 Å in the perpendicular plane. Helix H0 at the NH_2-terminus caps the top of the barrel, and the long hydrophobic helix H12 traverses the inside of the barrel from top to bottom. The possibility had previously been suggested that the H12 sequence might constitute a transmembrane domain (*9*).

Helix H12 forms integral contacts in almost every part of the hydrophobic core. The crystallographic temperature factors within this region are, on average, the lowest in the structure (27.8 Å^2 for H12 compared to 41.6 Å^2 for all other regions). It is highly unlikely that this helix could adopt other conformations outside of the core. The mutation in the tubby mouse abolishes a splice donor site in the 3′ exon, resulting in a deletion of the last 44 amino acids, which are replaced with 20 intron-encoded amino acids (*1, 2*). Thus, in the tubby mouse, the entire hydrophobic core of this domain will be disrupted, and it is therefore almost certain that no functional protein can be produced in these animals.

The electrostatic surface of the protein, generated with the program GRASP (*15*), shows two conspicuous features (Fig. 3). First, a groove of highly positive charge runs about one-half of the way around the barrel. This groove is ~50 Å long, varies between ~12 to 20 Å in width, and is up to ~9 Å in depth. There are surface contributions from strands 5, 6, 7, 8, 9, and 10, and the groove is bordered at the top (near to the NH_2-terminal end) by helix H8 and at the bottom by the large 7-8 loop and the three-stranded "extra" 9ABC sheet. The second notable feature is a smaller, primarily convex, patch of negative charge found on the face opposite to the

[1]Structural Biology Program, Department of Physiology and Biophysics, [2]Department of Biochemistry and Molecular Biology, [3]Ruttenberg Cancer Center, Mount Sinai School of Medicine of New York University, New York, NY 10029, USA.

*To whom correspondence should be addressed. E-mail: shapiro@anguilla.physbio.mssm.edu

center of the large groove.

Mapping TULP1 mutations (*6, 16*), identified from RP-14 patients, onto the tubby structure reveals a striking pattern. All surface mutations [with one exception, described in (*17*)] cluster within a relatively small region of the large positively charged groove that wraps the barrel. Some of these disease-causing mutants, including $Arg^{240} \rightarrow Pro^{240}$ and $Arg^{378} \rightarrow His^{378}$ (human TULP1 numbering), convert positively charged side chains to neutral ones, suggesting an important biological function dependent on the maintenance of a positive surface. On the basis of the contiguous arrangement of these mutants, we postulated that this surface might form a protein or nucleic acid binding site.

Nuclear Localization of Tubby Protein

Although the tissue distributions of tubby and other TULPs have been investigated by Northern (RNA) blotting and in situ messenger RNA (mRNA) hybridization (*5, 9*), cell type and subcellular protein localization have not been reported. To address these questions, we generated antibodies against the purified tubby COOH-terminal domain protein and used these, as well as antibodies raised against a peptide from the NH_2-terminal region, for immunocytochemical staining and protein immunoblotting experiments (*18*). First, we tested, with both of these methods, for protein expression in several cell lines and tissues including kidney cells, mouse fibroblasts (L cells), hippocampal neurons, astrocytes, and the Neuro-2A cell line. As expected, tubby is expressed primarily in cells of neural origin and at low or undetectable levels in the nonneural cells that we tested. The polyclonal antibody raised against the tubby COOH-terminal domain and the NH_2-terminal peptide antibody yield similar cell staining patterns and protein immunoblots. However, the COOH-terminal domain antibody gives protein immunoblots containing other bands near the tubby molecular weight. We presume that these are due to cross-reaction with other tubby-like proteins, which are highly related in sequence in the COOH-terminal domain.

To determine the subcellular localization of tubby, we performed immunofluorescence microscopy experiments using primary cultures of hippocampal neurons, Neuro-2A cells, and astrocytes, all of which express the tubby protein at high levels. The pattern of immunofluorescent staining suggests that tubby is localized mainly within the nuclei of these cells. These results were confirmed by subcellular fractionation of nuclei followed by protein immunoblotting, which shows that most tubby protein is in the nuclear fraction (Fig. 4). We also performed localization studies on cells synchronized at different points in the cell cycle (*18*), but noticeable changes in the immunostaining patterns were not observed in these experiments.

Mouse tubby has four sequences that fit consensus patterns (*19*) for nuclear localization signals (NLS). These are: $K^{39}KKR$, $P^{56}RSRRAR$, $P^{123}RKEKKG$, and $K^{302}RKK$ (*20*). These potential NLS are mostly conserved throughout the mammalian tubby-like proteins, and some (corresponding to those at positions 39 and 302) are also conserved in *Caenorhabditis elegans*. No bipartite NLS are evident. Only one of the potential NLS is within the COOH-terminal domain. This region ($K^{302}RKK$) is found at the base of β-strand 3, and the K_{302} side chain points into the core of the protein to make a structural salt bridge with D^{499} (*20*). Thus, we think it is unlikely that this corresponds to a functional

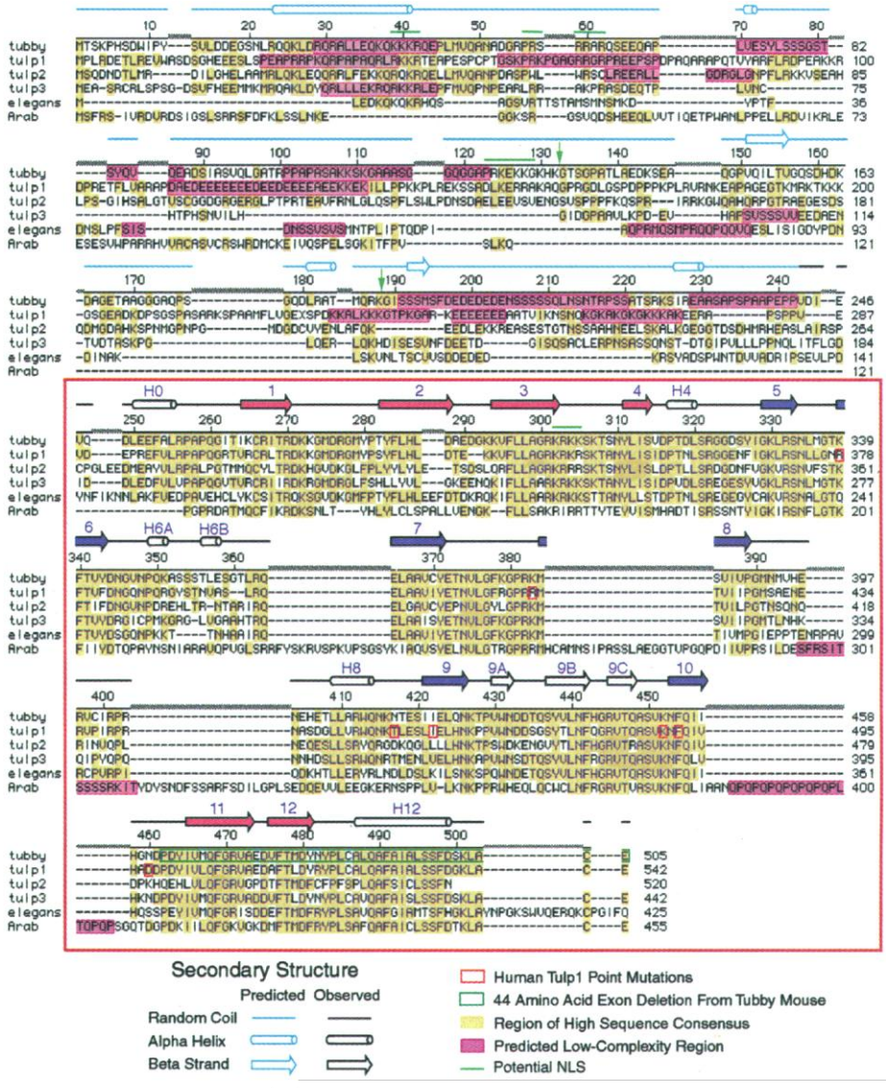


Fig. 1. Sequence alignments (*20*) for mouse tubby; human TULPs 1, 2, and 3; and TULPs from *C. elegans* and the plant *Arabidopsis* [database accession codes are given in (*36*)]. The conserved COOH-terminal "tubby domain" region is boxed in red. Regions of high sequence identity (four or more) are shaded in yellow. Human TULP1 point mutations (*6, 16*) are boxed in red, and the 44–amino acid exon deletion from the tubby mouse (*1, 2*) is boxed in green. The observed secondary structure for this domain is shown in black with the β strands of the barrel colored according to their predominant surface electrostatic potential (positive, blue; negative, red). The predicted secondary structure [PREDICT PROTEIN (*37*)] for the NH_2-terminal region of mouse tubby is colored in cyan. Regions of predicted low complexity (*38*) for all six proteins are boxed in pink. The relatively sparse secondary structure elements predicted for the NH_2-terminal regions of tubby-like proteins, for the most part, are not well conserved among different family members. In combination with the abundance of low-complexity regions, which do not generally adopt stably folded structures, it is likely that these domains are largely unstructured. This is commonly observed in many transcriptional activation domains (*24*). The alternatively spliced exon 5 in the mouse tubby NH_2-terminal segment is indicated by green arrows at the beginning and end of the exon. The green bars indicate potential NLS.

NLS. However, the conservation of the other three NLS in the NH_2-terminal region, which appears to be mostly unstructured (Fig. 1), argues for their functional relevance.

The Tubby COOH-Terminal Domain Binds Double-Stranded DNA

The predominantly nuclear localization of tubby led us to ask whether the function of this protein might involve interaction with DNA. This possibility was also supported by the observation that harmful mutants cluster within a positively charged putative molecular interaction site. Retrospectively, this positive groove is appropriately dimensioned for DNA interaction. We therefore tested the ability of the recombinant COOH-terminal domain of tubby to bind DNA. We assayed for DNA binding by gel-shift experiments (Fig. 5), using ^{32}P-labeled single- and double-stranded oligonucleotides of determined sequence (*21*). Remarkably, tubby binds avidly to double-stranded DNA but binds very poorly to single-stranded DNA. This specificity indicates that binding is not the result of nonspecific electrostatic interactions but rather is dependent on specific determinants characteristic of double-stranded DNA. Using a simple gel-shift assay, we were not able to determine the sequence specificity of DNA binding. However, the DNA binding behavior exhibited by tubby for noncognate double-stranded oligonucleotides is typical of that observed for sequence-specific DNA

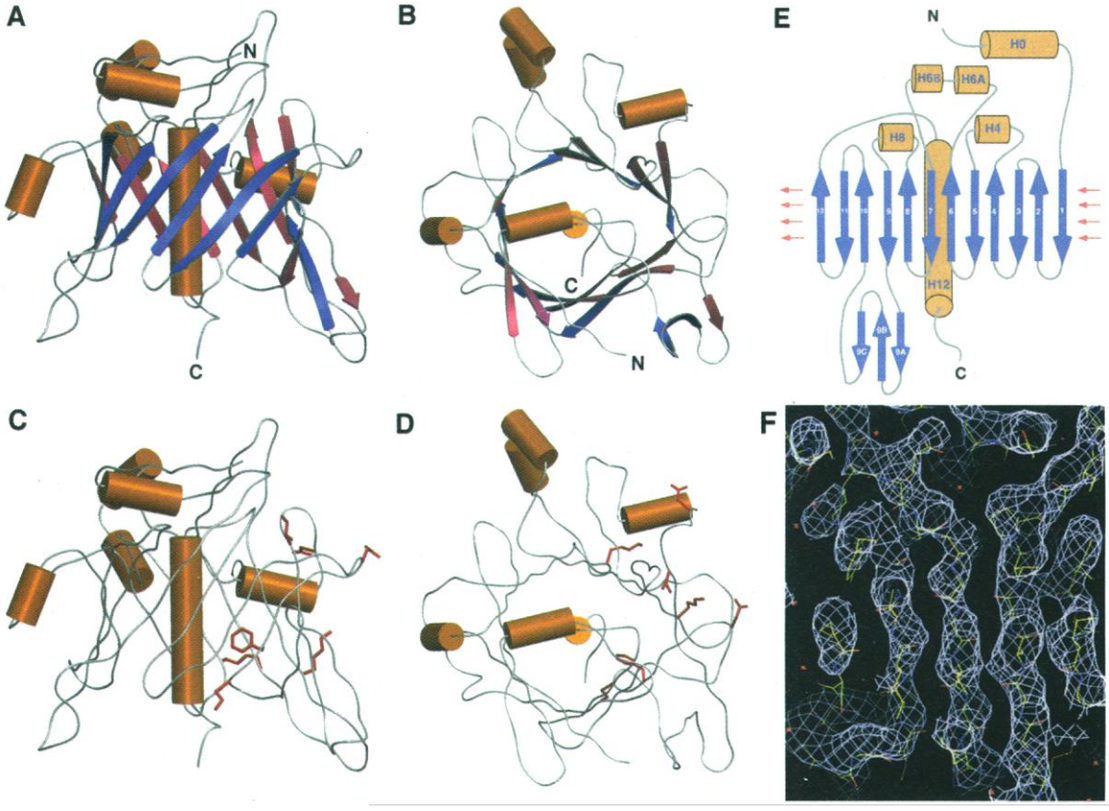


Fig. 2. (**A**) Ribbon diagram of the COOH-terminal domain from tubby (*39*). Helices are shown as cylinders, and the NH_2- and COOH-termini are indicated. The structure adopts a 12-stranded β-barrel conformation, filled by a central hydrophobic α helix (H12) that traverses the entire length of the barrel. (**B**) A 90° rotation from (A) looking down the main barrel from the NH_2-terminus. The helix H0 near the NH_2-terminus caps the top of the barrel. (**C**) Mutations from RP-14 patients, colored in red and mapped from the homologous protein TULP1, reveal a defined contiguous area. The surface of this region forms a large positive groove (see Fig. 3), which we suggest as forming the probable DNA binding surface. (**D**) A 90° rotation from (C). (**E**) Topology diagram of the tubby COOH-terminal domain. The small red arrows indicate continuation of hydrogen bonding between β strands 1 and 12 to form a closed barrel. (**F**) Sample electron density from the MAD experiment contoured at 1σ. This region shows part of the β barrel, with the final refined model. The selenomethionine and native crystals (which were used for refinement) have slightly different unit cell dimensions, accounting for part of the discrepancies between model and density shown here. Yellow and red stick models represent carbon and oxygen atoms, respectively. Red dots indicate water molecules.

Table 1. Data statistics of the crystallographic analysis. A high-resolution native data set, four-wavelength MAD experiment on a SeMet crystal and a single data set on an iridium derivative were collected on beamline X4A of the NSLS with a Quantum 4R charge-coupled device detector. Data were integrated and merged using the HKL software suite (*33*). Difference Fourier maps that were made with phasing from the iridium derivative and calculated with MLPHARE (*40*) were used to find the five selenium sites in the SeMet crystal. Numbers in parentheses refer to the highest resolution shell (3.11 to 3.00 Å for SeMet λ1, λ2, λ3, and λ4 and K_2IrCl_6, and 1.97 to 1.90 Å for the native crystal). The residual data error $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is observed intensity and $\langle I\rangle$ is average intensity. Unique reflections for nonnative data consider Bijvoet mates to be inequivalent.

Crystal	Wavelength λ (Å)	Resolution range (Å)	Reflections		Data statistics		$\langle I\rangle/\langle\sigma I\rangle$	Multiplicity	Number of sites
			Unique	Measured	Coverage (%)	R_{sym}			
Native	0.95370	20.0–1.9	21,512	569,516	98.0 (93.3)	0.044 (0.143)	30.24	4.4	–
SeMet λ1	0.98710	20.0–3.0	10,242	106,410	99.3 (100.0)	0.046 (0.145)	38.92	4.2	5
SeMet λ2	0.97920	20.0–3.0	10,214	103,004	99.2 (100.0)	0.043 (0.125)	44.02	4.2	5
SeMet λ3	0.97880	20.0–3.0	10,110	96,875	98.9 (100.0)	0.046 (0.098)	34.83	4.2	5
SeMet λ4	0.96411	20.0–3.0	10,155	99,561	99.5 (100.0)	0.042 (0.087)	42.55	4.2	5
K_2IrCl_4	0.95369	20.0–3.0	10,469	144,242	97.2 (99.4)	0.076 (0.121)	23.70	5.4	2

binding proteins (*22*). Therefore, a strong possibility exists that tubby will also bind DNA in a sequence-specific manner.

The tubby COOH-terminal domain does not resemble other known DNA binding proteins. It does not fall into any of the characterized groups, such as the helix-turn-helix or b-zip classes. We think that the tubby family defines a new class of DNA binding proteins. The lack of similarity to other DNA binding proteins does not repudiate the idea that DNA binding is a primary function of the tubby COOH-terminal domain; many DNA binding proteins recognize DNA through unique nonclassical structural motifs, including nuclear factor kappa B (NF-κB) (*23*) and many others. A full understanding of the molecular basis of tubby-DNA interaction will require experimental structural studies of protein-DNA complexes.

NH_2-Terminal Regions of Tubby and TULP1 Potently Activate Transcription

We next sought to investigate the function mediated by tubby-DNA binding. The NH_2-terminal segments show considerable variability among tubby-like proteins. These domains show scant secondary structure in sequence-based predictions and contain low-complexity stretches of amino acid sequence. Segments of this character are unlikely to form autonomously folded protein domains. Although there is no definitive sequence relation, these segments are reminiscent of the transactivation domains of many transcription factors (*24*). Thus, we formulated the hypothesis that tubby may act as a transcriptional activator.

Polymerase chain reaction (PCR) amplification of the NH_2-terminal region of tubby (residues 1 through 242) from a mouse brain complementary DNA (cDNA) library (*25*) revealed the presence of two different alternatively spliced forms of the tubby NH_2-terminal segment. One of these corresponds to the full-length NH_2-terminal region, and the other corresponds to an alternative splice form in which exon 5 is deleted. These are the same two splice forms that were identified in an earlier study (*1*).

To test whether these domains might act as transcriptional modulators, we constructed chimeric proteins in which the alternatively spliced forms of the tubby NH_2-terminal segment were fused to the DNA binding domain of GAL4 (*26*). We also made an analogous construct for the apparently unique splice form of the TULP1 NH_2-terminal region. We then tested the ability of these constructs, transfected into the Neuro-2A neuronal cell line, to activate transcription of a reporter gene situated downstream from a GAL4 DNA binding site (*18*). Remarkably, the full-length form of the tubby NH_2-terminal region potently activates transcription of the chloramphenicol acetyltransferase (CAT) reporter over 20 times the background level (Fig. 6), corresponding to ~30% of the activation seen for a GAL4 fusion with the exceptionally strong VP16 viral activator (*18, 26*). In addition, the TULP1 NH_2-terminal region also induces transcription activation, though less potently. The alternatively spliced form of the tubby NH_2-terminal segment lacking exon 5, however, shows much weaker activity, which is not clearly distinguishable from the background amount.

Alternative splicing within the activation domain is a common mechanism of control in

Table 2. Phasing statistics for the MAD experiment. MAD phases were calculated with the program SHARP (*34*), with the SeMet λ1 data set used as the pseudo-native. For details of structure solution and refinement, see (*14*). A total of 5608 acentric pairs and 1283 centric reflections were used for the best phase calculations, yielding an overall figure of merit of 0.65 for acentrics and 0.46 for centrics. After solvent flattening and phase extension to 2.8 Å with the program SOLOMON (*35*), the overall figure of merit was 0.88. Figure of merit = $\langle \Sigma P(\alpha)\exp(i\alpha)/\Sigma P(\alpha)\rangle$, where $P(\alpha)$ is the probability distribution for the phase α and i is the unit imaginary number. Phasing power = $\Sigma[F_{H(calc)}{}^2/(F_{PH(obs)} - F_{PH(calc)})^2]^{1/2}$, where $F_{H(calc)}$ is the calculated heavy-atom structure factor and $F_{PH(obs)}$ and $F_{PH(calc)}$ are the observed and calculated derivative structure factors, respectively. $R_{cullis} = (\Sigma|F_{PH(obs)} - F_{P(obs)}| - |F_{H(calc)}|)/(\Sigma| F_{PH(obs)} - F_{P(obs)}|)$, where $F_{P(obs)}$ is the observed protein structure factor.

Crystal	Resolution range (Å)	R_{cullis}		Phasing power	
		Centric iso	Acentric iso/ano	Centric iso	Acentric iso/ano
SeMet λ2	20.0–3.0	0.71	0.60/0.66	0.71	1.24/2.76
SeMet λ3	20.0–3.0	0.77	0.64/0.45	0.97	1.43/4.29
SeMet λ4	20.0–3.0	0.75	0.65/0.48	1.27	2.02/3.94

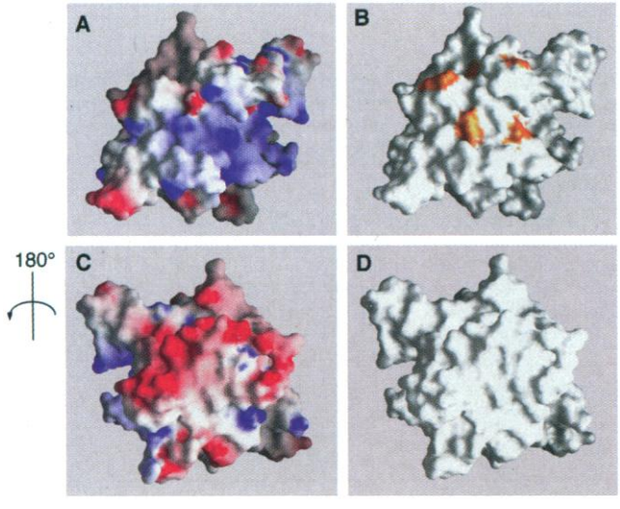


Fig. 3. (**A**) Electrostatic surface of the tubby COOH-terminal domain generated with the program GRASP (*15*). The axis of H12, the central helix in the center of the β barrel, is vertical in all panels, with the NH_2-terminal at the top. The surface potential is colored, representing electrostatic potentials from $-10kT$ (red) to $+10kT$ (blue), where k is the Boltzmann constant and T is the temperature. This analysis reveals a groove, ~50 Å long, ~12 to 20 Å wide, and up to ~9 Å deep, that wraps around the center of the barrel. (**B**) Mutations from RP-14 patients, mapped from the homologous protein TULP1, displayed on the surface of the tubby structure in yellow. These mutations cluster within and bordering the positively charged groove, which we suggest as forming the probable DNA binding surface. (**C** and **D**) Corresponding diagrams rotated 180° about the long axis of the central helix. A negatively charged surface is seen on this face of the molecule, but disease-causing mutations are not.

Table 3. Statistics for refinement of molecular model. Of the residues in the structure, 70.3% lie in the most favorable regions of the Ramachandran plot as defined in the program PROCHECK (*41*), and no nonglycine residues are in disallowed regions. R factor = $100 \times \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, where F_{obs} are the observed structure factors and F_{calc} are the calculated structure factors. The crystallographic R factor R_{cryst} is based on 95% of the data used in refinement, and the free R factor R_{free} is based on 5% of the data withheld for the cross-validation test.

Parameter	Value
Resolution range (Å)	8.0–1.9
Sigma cutoff (F)	3σ
Number of working set reflections ($\|F\| > 3\sigma$)	19,244
Number of R_{free} set reflections	937
R_{cryst} (%)	21.0
R_{free} (%)	26.5
Number of atoms (protein/water)	2,066/509
Root-mean-square deviations from ideal geometry	
Bond lengths (Å)	0.005
Bond angles (degrees)	1.4
Dihedral angles (degrees)	25.0
Improper angles (degrees)	0.64

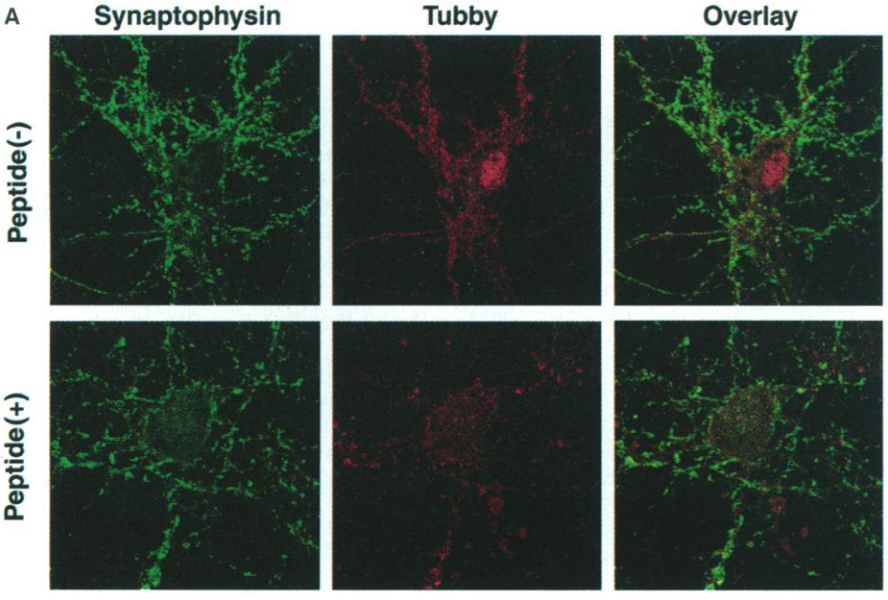
transcriptional modulators (*27*). Exon 5 of tubby contains sequence elements that are similar to those found in glutamine-rich transcriptional activators, such as cyclic adenosine 3′,5′-monophosphate response element–binding protein, Sp1, and Oct-2. In addition, this segment (as well as other sequences in the NH_2-terminal region) displays an abundance of serine and threonine residues, which are also overrepresented in many activation domains (*24*). Although exon 5 accounts for only a small portion of the NH_2-terminal region, its excision results in a drastic reduction in the ability of the fusion protein to activate transcription. It therefore seems possible that alternative splicing of these regions could provide an additional layer of regulation for transcriptional activation by tubby of its potential target genes.

The ability of a protein segment to activate transcription in the GAL4 fusion assay is by no means an absolute proof of its natural function. However, we think that the sum of circumstantial evidence in the case of tubby-like proteins is highly convincing. Several factors demonstrate their similarity to other families of transcription factors. First, tubby and other TULPs contain DNA binding domains that are always situated at a protein terminus (the COOH-terminus in this case). Second, disruption of the positively charged surface on this domain results in disease phenotypes, indicating the probable biological importance of DNA binding. Third, the localization of tubby is primarily nuclear. Fourth, the NH_2-terminal regions of tubby-like proteins contain low-complexity regions, including acidic and glutamine-rich regions, reminiscent of transcription activation domains. Fifth, these regions have the ability to potently stimulate RNA transcription. And sixth, these regions are alternatively spliced in ways that radically alter their ability to activate transcription, while not affecting the overall character of the protein segment. The combination of these elements strongly suggests that tubby-like proteins constitute a unique family of transcriptional regulators.

Conclusions

The molecular architecture of tubby proteins is well suited for function in transcriptional modulation. The long positively charged putative DNA binding groove and the spatial clustering of disease mutants suggest that tubby proteins might bind DNA sequences of ~20 base pairs in length. Accomplishing this, however, would require substantial bending of the DNA to "wrap" partway around the tubby barrel. This leads us to speculate that the unique filled-barrel structure of the tubby DNA binding domain may be to impart rigidity for bending DNA. Many transcription factors bend DNA as a required aspect of their function. It is possible that tubby proteins will ultimately fall into this class of "architectural" transcription factors.

Our data implicate tubby-like proteins as bipartite transcription regulators that bind double-stranded DNA through COOH-terminal DNA binding domains and have transcription modulation segments at their NH_2-



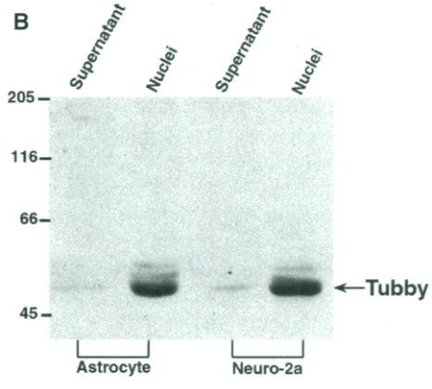


Fig. 4. Tubby protein is found primarily in the nucleus. (**A**) Immunofluorescence staining of a primary culture of hippocampal neurons (*18*). Tubby staining is shown in red, and synaptophysin staining, which marks the cell body and synaptic processes, is shown in green. The tubby antibody used in this experiment was raised against an NH_2-terminal 19–amino acid peptide from tubby that does not show similarity to other proteins in the sequence database. In protein immunoblotting, this antibody stains a single 55-kD band, which corresponds to the correct molecular weight for tubby. The lower panels show immunofluorescence staining in the presence of the immunogen peptide. No staining above background is observed, thus demonstrating the specificity of the antibody. (**B**) Protein immunoblotting of subcellular fractions from Neuro-2A cells with the same antibody. Tubby protein is found almost exclusively in the nuclear fraction.

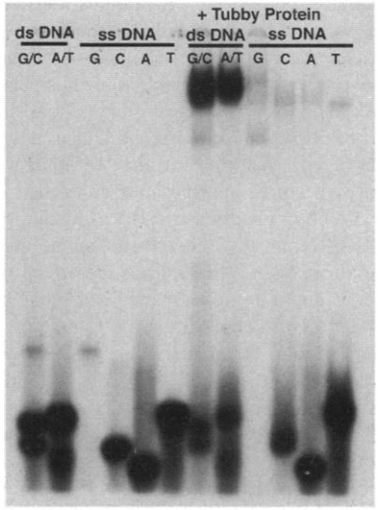


Fig. 5. Tubby COOH-terminal domain binding to DNA. The tubby COOH-terminal domain, as a fusion protein with GST, was incubated in the presence of ^{32}P-labeled single-stranded (ss) and double-stranded (ds) oligonucleotides, and binding was assayed by gel mobility shift analysis. The data shown are for the homomeric 20-mers $d(C_{20})$, $d(G_{20})$, $d(A_{20})$, and $d(T_{20})$. Tubby demonstrates a marked preference for binding double-stranded DNA. The $d(G_{20})$ probe forms an anomalous secondary structure and migrates out of the gel. Similar DNA binding behavior is observed for the isolated tubby COOH-terminal domain. Other double-stranded oligonucleotides of varying sequence were also bound efficiently, and the preference for double-stranded DNA was maintained for all sequences tested. Experimental details are described in (*21*).

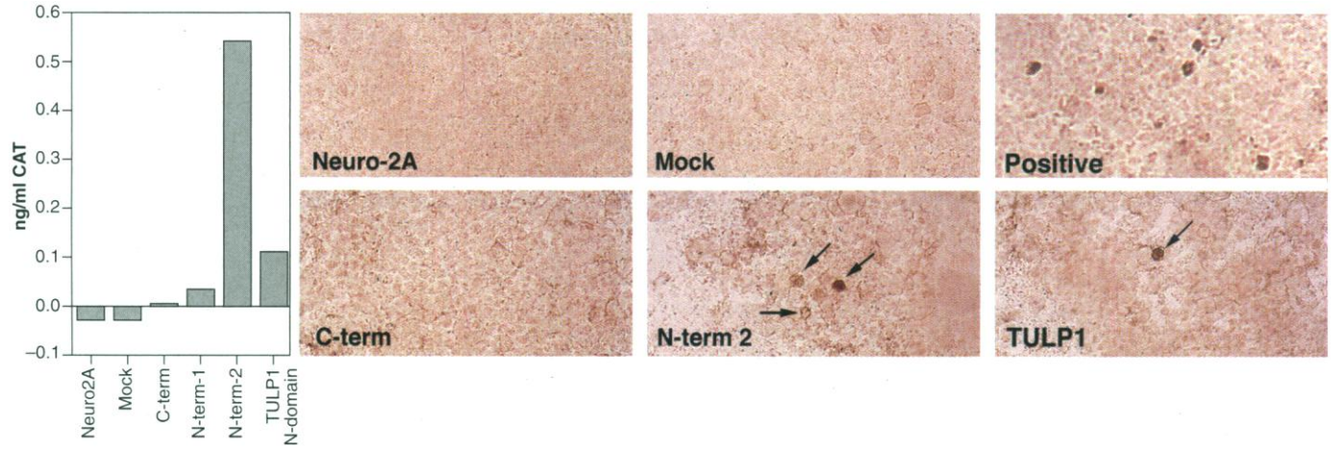


Fig. 6. Transcription activation assay. Both splice forms of the NH_2-terminal region of tubby and the unique NH_2-terminal region of TULP1, as well as appropriate controls (*18*), were fused to the DNA binding domain of GAL4. These constructs were cotransfected into Neuro-2A cells with a reporter vector in which a GAL4 DNA binding site is situated upstream of a sequence encoding CAT. These were developed for microscopy by an enzymatic staining method (*18*). The Neuro-2A panel shows staining of nontransfected cells; Mock was transfected with a GAL4 DNA binding domain without a fusion partner; Positive was transfected with a fusion protein that included the activation domain from the viral protein VP16; C-term included the COOH-terminal DNA binding domain from tubby; N-term 2 included the full-length form of the tubby NH_2-terminal region; and TULP1 included the NH_2-terminal region from TULP1. Activation can be seen by the presence of dark stain in the cell cytoplasm. Whereas GAL4 fusions including the full-length NH_2-termini of tubby and TULP1 clearly induce transcriptional activation, a fusion containing the DNA binding COOH-terminal domain of tubby does not. To quantitate the degree of activation, we performed ELISA assays to detect the production of CAT in these transfected cells (*18*). CAT expression is graphed in the left panel, as calibrated against a standard curve. Whereas the NH_2-terminal segment from TULP1 and the full-length NH_2-terminal region of tubby activate transcription, the alternatively spliced form of the tubby NH_2-terminal segment lacking exon 5 shows little or no activation.

termini. Their apparent function as transcription factors, coupled with the positions of disease mutants in the probable DNA binding groove, reveals the probable molecular basis of RP-14: DNA binding of mutant TULP1 is likely to be compromised, and hence, downstream genes required for photoreceptor survival will not be appropriately controlled.

In tubby mice, the amounts of mutant *tub* mRNA are much higher than that of wild-type mice (*1, 2*). This suggests the possibility of a feedback regulation mechanism in the production of tubby. Similar up-regulation of a transcription factor in another disease syndrome has been seen as well (*28*). The alternative splicing observed in the NH_2-terminal activation region of tubby is also a common mechanism for the regulation of transcription factors (*27*). The form of the tubby protein lacking exon 5 will presumably bind to the same DNA sites as full-length tubby. However, the shorter protein appears to be unable to activate transcription. This is reminiscent of similar transcriptional control systems in which a regulated balance is maintained between silencers and activators with DNA binding specificity for the same site. A well-known example of such a system is the silencer NF-κB p50 homodimer and the activator NF-κB p65 (*23*). Other control mechanisms are likely to exist; for example, it has recently been reported that tubby may be tyrosine-phosphorylated in response to insulin (*29*), but the biological importance of this observation is not yet clear.

The key remaining question is, What are the genes that tubby and other TULPs activate? The disease phenotypes resulting from mutation of tubby proteins have been attributed to neuronal cell death. In the case of TULP1, retinitis pigmentosa arises as the result of the death of photoreceptor neurons, and in tubby mice, both photoreceptors and cochlear hair cells die (*4, 5*). These phenotypes were thought to have their onset at maturity (*3, 4, 7*); however, one study suggests that RP-14 in humans may arise early in life (*8*), and abnormal vision and hearing responses in tubby mice can also be detected at an early age (*30*). In either case, severe phenotypic defects appear to arise as the result of neuronal apoptosis (*31*). Thus, it appears that the loss of function of tubby-like proteins can result in a failure of the viability of neurons. The genes activated by tubby family members may therefore be of critical importance for neuronal survival. The work presented here may provide the intellectual framework for the subsequent identification of these genes.

The structure-based functional genomics approach described here demonstrates the power of applying structural methods to the study of proteins that do not have assigned biochemical functions, but are implicated through genetics or genomic screens in specific biological processes. In general, structures of such proteins will fall into two categories: (i) structures that reveal similarity to proteins of known function and (ii) structures that are unrelated to known proteins. The former type may suggest shared functional properties with their structural relatives. In the second case, of which tubby is an example, functional characteristics are more likely to be deduced not by structural similarity to other proteins, but rather through analysis of structural features (for example, the spatial clustering of mutants). This analysis can then be used as a cornerstone to guide the application of other biological techniques. Although this latter case may present greater difficulties, the case of tubby-like proteins suggests that these difficulties can be surmounted. As more protein structures become available, the power of this method should grow accordingly.

References and Notes

1. P. W. Kleyn *et al.*, *Cell* **85**, 281 (1996).
2. K. Noben-Trauth, J. K. Naggert, M. A. North, P. M. Nishina, *Nature* **380**, 534 (1996).
3. D. L. Coleman and E. M. Eicher, *J. Hered.* **81**, 424 (1990).
4. K. K. Ohlemiller *et al.*, *Neuroreport* **6**, 845 (1995); K. K. Ohlemiller *et al.*, *Audiol. Neurootol.* **2**, 175 (1997).
5. M. A. North, J. K. Naggert, Y. Yan, K. Noben-Trauth, P. M. Nishina, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3128 (1997).
6. P. Banerjee *et al.*, *Nature Genet.* **18**, 177 (1998); S. A. Hagstrom, M. A. North, P. L. Nishina, E. L. Berson, T. P. Dryja, *Nature Genet.* **18**, 174 (1998).
7. K. K. Ohlemiller *et al.*, *Cell Tissue Res.* **291**, 489 (1998).
8. C. A. Lewis *et al.*, *Invest. Ophthalmol. Visual Sci.* **40**, 2106 (1999).
9. P. M. Nishina, M. A. North, A. Ikeda, Y. Yan, J. K. Naggert, *Genomics* **54**, 215 (1998).
10. S. H. Kim, *Nature Struct. Biol.* **5** (suppl), 643 (1998); T. C. Terwilliger *et al.*, *Protein Sci.* **7**, 1851 (1998); L. Shapiro and C. D. Lima, *Structure* **6**, 265 (1998).
11. W. A. Hendrickson, *Science* **254**, 51 (1991).
12. C. D. Lima, M. G. Klein, W. A. Hendrickson, *Science* **278**, 286 (1997); L. Shapiro and P. E. Scherer, *Curr.*

Biol. **8**, 335 (1998); T. I. Zarembinski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15189 (1998).

13. The tubby COOH-terminal domain was expressed as a fusion protein with glutathione S-transferase (GST) in the pGEX-2T expression vector. The fusion protein was purified by affinity chromatography on glutathione sepharose. The tubby COOH-terminal domain was released by proteolysis with thrombin and purified to homogeneity by ion exchange chromatography on mono-S followed by gel filtration on a Superdex-200 column. Analytical gel filtration shows that the protein exists in solution as a monomer.
14. Crystallization conditions used protein at a concentration of 10 mg/ml in 10 mM tris (pH 8.0), 0.15 M NaCl, and 5 mM dithiothreitol (DTT) (1.2 μl), mixed with well solution (1.2 μl) of 2% polyethylene glycol 4000, 0.1 M Hepes (pH 7.5), 4% 2-propanol, and 5 mM DTT. Crystals formed in space group $P2_12_12_1$ with a = 43.5 Å, b = 51.0 Å, and c = 121.1 Å and contained one tubby molecule per asymmetric unit. Potential heavy-atom derivatives were screened by native polyacrylamide gel screening and mass spectrometry (*32*). Heavy-atom soaks were performed in a stabilization buffer identical to the well solution but lacking DTT. Native crystals were soaked for a period of 4 hours in a solution of 4 mM K_2IrCl_6. Selenomethionyl protein crystals grew in the same conditions as the native. Data were collected on three crystals: a native, a SeMet, and a K_2IrCl_6-soaked crystal at beamline X4A of the National Synchrotron Light Source (NSLS). Crystals were flash-cooled to 100 K in stabilization buffer supplemented with 30% ethylene glycol. Four data sets were collected around the selenium K absorption edge on the SeMet crystal. Data were processed and merged with the programs DENZO and SCALEPACK (*33*). The structure of COOH-terminal domain of tubby was determined by MAD phasing (*11*) of the SeMet crystal. Strong peaks in the Bijvoet difference Patterson map from the iridium derivative allowed two sites to be found and initial phases to be generated by the single isomorphous replacement with anomalous scattering method. Difference Fourier maps using these phases were then calculated in X-PLOR [A. T. Brünger, *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992)] and used to determine the positions of the five ordered selenium atoms. Phases were generated from the selenium data sets, using λ1 as the pseudo-native, to 3.0 Å with the program SHARP (*34*) and were improved by solvent flattening and phase extension to 2.8 Å with the program Solomon (*35*). An initial model of 221 residues was built into the resultant maps with the program O [T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)], rigid body refined, and refined against the high-resolution native data using a maximum likelihood target function with the program CNS [A. T. Brünger *et al.*, *Acta Crystallogr.* **D54**, 905 (1998)]; 265 amino acid residues (19 of which with alternative conformations), 509 water molecules, and two phosphate ions were found and built into the structure.
15. A. Nicholls, K. Sharp, B. Honig, *Proteins* **11**, 281 (1991).
16. S. Gu *et al.*, *Lancet* **351**, 1103 (1998).
17. One TULP1 RP-14 mutation is not localized within the putative DNA binding region. This mutant, $Lys^{489} \rightarrow Arg^{489}$ (human TULP1 numbering, corresponding to mouse tubby Lys^{452}), is situated at the base of the barrel near the COOH-terminus. The side chain of this residue is in close proximity to the COOH-terminal K^{501}LACE sequence (*20*), conserved in all mammalian tubby proteins, which protrudes out of the barrel from the central helix. The small chemical modification of the Lys to Arg mutation and the spatial arrangement near conserved residues led us to investigate the potential of this region as an enzyme active site. Although we could not identify structural analogs of this site, or an enzyme activity, the possibility remains that this region serves an important function.
18. Cells were cultured as follows. Neuro-2A cells were cultured in a humidified atmosphere of 10% CO_2 and in 90% Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, and astrocytes were grown in 5% horse serum. Plates of astrocytes were intermittently trypsinized in 0.05% trypsin and 0.05 mM EDTA, and plated onto poly-L-lysine–coated cover slips for immunostaining. Neurons were prepared from hippocampi of embryonic day 18 Sprague-Dawley rats as described previously [K. Goslin, H. Asmussen, G. Banker, in *Culturing Nerve Cells*, G. Banker and K. Goslin, Eds. (MIT Press, Cambridge, MA, 1991), pp. 251–282]. Neurons were plated at a density of 3600 cells/cm² on poly-L-lysine–coated cover slips in DMEM containing 10% horse serum. After 4 hours, when neurons had attached, cover slips were transferred to dishes containing a monolayer of cortical astroglia, where they could be maintained for up to 2 weeks in DMEM containing N_2 supplement, 1 mM sodium pyruvate, and 0.1% ovalbumin. Immunocytochemistry was determined as follows. Cells were fixed in 4% paraformaldehyde, delipidated in 100% methanol, permeabilized with 0.1% Triton X-100, and blocked with 5% normal goat serum in phosphate-buffered saline. After incubation for 1 hour at 37°C with primary antibody to tubby, the cells were then incubated for 30 min with a fluorescent secondary antibody (Jackson ImmunoResearch Laboratories, West Grove, PA). Cover slips were then mounted and examined with confocal laser microscopy. For cell cycle dependence experiments, cells were synchronized at different points in the cell cycle by incubation overnight with each of the following reagents before fixation and staining: 300 μM L-mimosine, 1 mM hydroxyurea, or nocodazole (1 μg/ml). The immunoblotting process is described as follows. To obtain a fraction of nuclei, we incubated astrocytes and Neuro-2A cells hypotonic buffer (2mM $NaHCO_3$, 1 mM $MgCl_2$, 1 mM EDTA, and protease inhibitors) for 5 min on ice. The swollen cells were disrupted by Dounce homogenization. The pellet was collected by centrifugation at 1000*g* for 10 min. The crude nuclear fraction lysates were extracted in lysis buffer [150 mM NaCl, 20 mM tris-HCl (pH 7.5), 1% NP-40, 1% Triton X-100, 1% SDS, 2mM $CaCl_2$, leupeptin (0.5 μg/ml), pefabloc 0.1 mM, and aprotinin (1 μg/ml)] for 30 min. Before immunoblotting, protein concentration was determined with the bicinchoninic acid assay (Pierce, Rockford, IL), and samples were run on 7.5% SDS–polyacrylamide gel electrophoresis (PAGE), transferred to nitrocellulose, blocked with 5% milk protein, and incubated overnight with primary antibody to Tubby. After secondary antibody incubation and washing, blots were developed with the enhanced chemiluminescence system (NEN Life Science Products, Boston, MA). The activation assay was performed as follows. Different domains of tubby genes were fused to the GAL4 DNA binding domain in the pM vector (Clontech). The orientation and reading frame of fusions were correct for hybrid proteins to be expressed, and sequences were confirmed by di-deoxy DNA sequencing. The plasmids were cotransfected into Neuro-2A cells with the Superfect reagent (Qiagen, Chatsworth, CA). CAT activity was detected 72 hours after transfection by CAT staining [M. J. Donoghue *et al.*, *J. Cell. Biol.* **115**, 423 (1991)] (Boehringer Mannheim, Mannheim, Germany) and microscopy and enzyme-linked immunosorbent assay (ELISA) [P. Porsch *et al.*, *Anal. Biochem.* **211**, 133 (1993)] (Boehringer Mannheim), and the activity was analyzed by absorbance in a multiwell spectrophotometer.
19. J. Garcia-Bustos, J. Heitman, M. N. Hall, *Biochim. Biophys. Acta* **1071**, 83 (1991).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. GST-tubby COOH-terminal domain fusion protein (0.5 μg) was incubated in 25 mM Mops (pH 8.0), 2 mM DTT, 120 mM KOAc, and 2 mM EDTA with 0.5 pmol of either double-stranded or single-stranded oligonucleotides ^{32}P labeled on the 5′ end. After 25 min at 30°C, complexes were resolved on a 6% polyacrylamide native gel and visualized by autoradiography. The following oligonucleotides were used in the binding assays: $d(C_{20})$, $d(G_{20})$, $d(A_{20})$, and $d(T_{20})$. Similar results were observed with purified COOH-terminal domain on the above oligonucleotides as well as 5′-TCCAGTATATATATAAGTCAAG-3′ and 5′-TCCAGTAGATCCAGAAGTCAAG-3′ annealed to their complementary strands.
22. O. G. Berg, in *The Biology of Nonspecific DNA-Protein Interactions*, A. Revzin, Ed. (CRC Press, Boca Raton, FL 1990), pp. 71–85.
23. S. Ghosh, M. J. May, E. B. Kopp, *Annu. Rev. Immunol.* **16**, 225 (1998).
24. S. J. Triezenberg, *Curr. Opin. Genet. Dev.* **5**, 190 (1995).
25. PCR amplification of the tubby NH_2-terminal region was performed with a mouse brain cDNA library (Clontech, Palo Alto, CA) and primers complementary to amino acid residues 1 through 8 and 239 through 246. Two distinct bands were observed, and these were inserted into the vector pM (Clontech) and sequenced. The sequence revealed that the these two bands corresponded to the full-length NH_2-terminal region and an alternative splice form lacking exon 5 (amino acids 133 through 188).
26. I. Sadowski, J. Ma, S. Triezenberg, M. Ptashne, *Nature* **335**, 563 (1988).
27. C. Roman, L. Cohn, K. Calame, *Science* **254**, 94 (1991); T. Tanaka *et al.*, *EMBO J.* **14**, 341 (1995); N. S. Belaguli, W. Zhou, T. H. Trinh, M. W. Majesky, R. J. Schwartz, *Mol. Cell. Biol.* **19**, 4582 (1999).
28. E. Steingrimsson *et al.*, *Nature Genet.* **8**, 256 (1994).
29. R. Kapeller *et al.*, *J. Biol. Chem.* **274**, 24980 (1999).
30. J. R. Heckenlively *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11100 (1995).
31. S. Ikeda *et al.*, *Invest. Ophthalmol. Visual Sci.* **40**, 2706 (1999).
32. Free protein in solution (1 μl of 3 mg/ml) was reacted, on ice, with excess amounts of heavy-atom reagents (1 μl of 10 mM) for 2 hours. Native PAGE of the reaction mixtures revealed several conditions in which the protein had been denatured, and these heavy-atom reagents were then discarded. The remaining mixtures were then analyzed with mass spectrometry, which revealed two covalently bonded protein–heavy-atom complexes, K_2IrCl_6 and ethylmercurithiosalicyclic acid (EMTS). After soaking native protein crystals with these reagents (4 mM for 4 hours) and collecting x-ray diffraction data, strong derivatives were found for both. However, the EMTS-soaked crystal form was very highly nonisomorphous with the native, so this derivative was not used in the structure determination.
33. Z. Otwinowski and W. Minor, *Methods Enzymol.* **276**, 307 (1997).
34. E. de La Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997).
35. J. P. Abrahams and A. G. W. Leslie, *Acta Crystallogr.* **D52**, 30 (1996).
36. The sequences used for alignment correspond to the following GenBank accession codes: Tubby (*Mus musculus*), 1717822; TULP1 (*Homo sapiens*), 4507735; TULP2 (*H. sapiens*), 4507737; TULP3 (*H. sapiens*), 4507739; elegans (*Caenorhabditis elegans*), 3875712; and Arab (*Arabidopsis thaliana*), 2829918.
37. B. Rost, *Methods Enzymol.* **266**, 525 (1996).
38. J. C. Wootton and S. Federhen, *Methods Enzymol.* **266**, 554 (1996).
39. Ribbon diagram figures in this manuscript were made with the program SETOR [S. V. Evans, *J. Mol. Graphics* **11**, 134 (1993)], and the sequence alignment was made with DNASTAR [DNASTAR, *Molecular Biotechnology*, **5**, 185 (1996)].
40. Z. Otwinowski, in *Isomorphous Replacement and Anomalous Scattering, Proceedings of the Daresbury Study Weekend*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (CLRC Daresbury Laboratory, Daresbury, UK, 1991).
41. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
42. We are grateful to C. Ogata and the staff of the NSLS beamline X4A for help with data collection. We thank W. A. Hendrickson, T. Harris, P. Kwong, and P. Scherer for many helpful discussions. T.J.B. is the recipient of a Wellcome Trust International Prize Travelling Research Fellowship (056509/Z/98/Z). L.S. is the recipient of a Career Scientist Award from the Irma T. Hirschl Foundation. This work was supported in part by a pilot study grant from Structural Genomix. Beamline X4A at the NSLS, a U.S. Department of Energy facility, is supported by the Howard Hughes Medical Institute. Coordinates have been deposited in the Protein Data Bank (accession code 1C8Z).

6 October 1999; accepted 29 October 1999

References and Notes

[5] **Molecular Characterization of TUB, TULP1, and TULP2, Members of the Novel Tubby Gene Family and their Possible Relation to Ocular Diseases**

Michael A. North; Juergen K. Naggert; Yingzhuo Yan; Konrad Noben-Trauth; Patsy M. Nishina

Proceedings of the National Academy of Sciences of the United States of America, Vol. 94, No. 7. (Apr. 1, 1997), pp. 3128-3133.

Stable URL:

http://links.jstor.org/sici?sici=0027-8424%2819970401%2994%3A7%3C3128%3AMCOTTA%3E2.0.CO%3B2-5

[11] **Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation**

Wayne A. Hendrickson

Science, New Series, Vol. 254, No. 5028, Special Issue: Instrumentation. (Oct. 4, 1991), pp. 51-58.

Stable URL:

http://links.jstor.org/sici?sici=0036-8075%2819911004%293%3A254%3A5028%3C51%3ADOMSFA%3E2.0.CO%3B2-3

[12] **Structure-Based Analysis of Catalysis and Substrate Definition in the HIT Protein Family**

Christopher D. Lima; Michael G. Klein; Wayne A. Hendrickson

Science, New Series, Vol. 278, No. 5336. (Oct. 10, 1997), pp. 286-290.

Stable URL:

http://links.jstor.org/sici?sici=0036-8075%2819971010%293%3A278%3A5336%3C286%3ASAOCAS%3E2.0.CO%3B2-I

LINKED CITATIONS

[12] **Structure-Based Assignment of the Biochemical Function of a Hypothetical Protein: A Test Case of Structural Genomics**

Thomas I. Zarembinski; Li-Wei Hung; Hans-Joachim Mueller-Dieckmann; Kyeong-Kyu Kim; Hisao Yokota; Rosalind Kim; Sung-Hou Kim

Proceedings of the National Academy of Sciences of the United States of America, Vol. 95, No. 26. (Dec. 22, 1998), pp. 15189-15193.

Stable URL:

http://links.jstor.org/sici?sici=0027-8424%2819981222%2995%3A26%3C15189%3ASAOTBF%3E2.0.CO%3B2-I

[27] **A Dominant Negative Form of Transcription Activator mTFE3 Created by Differential Splicing**

Christopher Roman; Lauren Cohn; Kathryn Calame

Science, New Series, Vol. 254, No. 5028, Special Issue: Instrumentation. (Oct. 4, 1991), pp. 94-97.

Stable URL:

http://links.jstor.org/sici?sici=0036-8075%2819911004%293%3A254%3A5028%3C94%3AADNFOT%3E2.0.CO%3B2-X

[30] **Mouse Model for Usher Syndrome: Linkage Mapping Suggests Homology to Usher Type I Reported at Human Chromosome 11p15**

John R. Heckenlively; Bo Chang; Lawrence C. Erway; Chen Peng; Norman L. Hawes; Gregory S.Hageman; Thomas H. Roderick

Proceedings of the National Academy of Sciences of the United States of America, Vol. 92, No. 24. (Nov. 21, 1995), pp. 11100-11104.

Stable URL:

http://links.jstor.org/sici?sici=0027-8424%2819951121%2992%3A24%3C11100%3AMMFUSL%3E2.0.CO%3B2-K