An Extended Winged Helix Domain in General Transcription Factor E/IIEα*

Anton Meinhart, Jascha Blobel, and Patrick Cramer‡

From the Institute of Biochemistry, Gene Center, University of Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

Initiation of eukaryotic mRNA transcription requires melting of promoter DNA with the help of the general transcription factors TFIIE and TFIIH. Here we define a conserved and functionally essential N-terminal domain in TFE, the archaeal homolog of the large TFIIIE subunit α. X-ray crystallography shows that this TFE domain adopts a winged helix-turn-helix (winged helix) fold, extended by specific α-helices at the N and C termini. Although the winged helix fold is often found in DNA-binding proteins, we show that TFE is not a typical DNA-binding winged helix protein, because its putative DNA-binding face shows a negatively charged groove and an unusually long wing, and because the domain lacks DNA-binding activity in vitro. The groove and a conserved hydrophobic surface patch on the additional N-terminal α-helix may, however, allow for interactions with other general transcription factors and RNA polymerase. Homology modeling shows that the TFE domain is conserved in TFIIEα, including the potential functional surfaces.

In eukaryotes, RNA polymerase (pol)1 II requires five general transcription factors, called TFIIIB, -D, -E, -F, and -H, to initiate mRNA transcription. The general transcription factors assemble with pol II into a large initiation complex on promoter DNA (1–3). According to in vitro reconstitution experiments, initiation complex assembly starts with recognition of the promoter DNA by TFIIID (4, 5). The TFIIID subunit TATA box-binding protein (TBP) binds to the TATA box, a frequently occurring promoter element located about 30 bp upstream of the transcription start site. The TBP-DNA complex is stabilized by specific binding of TFIIID to TBP and to promoter DNA sequences adjacent to the TATA box (6). The TBP-TFIIID-DNA complex can be further stabilized by binding of the optional factor TFIIA. A preformed pol II-TFIIID complex is then recruited to the transcription start site. For most parts of the resulting complex, detailed structures are available, including TBP (7, 8), the TBP-DNA complex (9, 10), complexes TFIIID-TFIIIB (11, 12) and TBP-DNA-TFIIIA (13–15), domains of TFIIIB (16) and TFIIIF (17–22), and pol II (23–27). Before RNA synthesis can begin, DNA must be melted around the transcription start site with the help of TFIIIE and TFIIH (28). Binding of TFIIIE to the pol II complex apparently recruits TFIIH (29). TFIIH contains helicase activities that unwind DNA and form a transcription bubble (30, 31). TFIIH also comprises a kinase that phosphorylates the pol II C-terminal repeat domain during the transition from initiation to elongation. Both enzymatic activities of TFIIH are stimulated by TFIIE (32). In keeping with a role of TFIIIE as a bridging factor between pol II and TFIIH, TFIIE has been shown to bind to pol II and to TFIIH (29). Several lines of evidence suggest a role of TFE in promoter DNA binding and melting. First, TFIE can bind single-stranded DNA (33). Second, TFIE function is influenced by the helical stability of promoter DNA, and the requirement for TFIIIE during initiation can be bypassed by premelted promoter sequences (34). Finally, site-specific cross-linking located TFIIIE near DNA around the active center of pol II. One study observed cross-linking of TFIE to DNA within the transcription bubble and just downstream of it (35). Another cross-linking study located TFIIIE adjacent to promoter DNA just upstream of the transcription start site (36).

TFIIIE consists of a large and a small subunit, referred to as α and β, respectively, in higher eukaryotes. From N to C termini, the TFIIIEβ sequence shows a serine-rich region with little predicted secondary structure, a region with sequence similarity to a DNA-binding domain in the small TFIIIF subunit, a region with similarity to the bacterial σ factor, and a presumably helical region with clusters of basic residues. NMR has shown that the TFIIIEβ region with homology to TFIIIF forms a winged helix domain (37). Structural information on TFIIIEα is, however, lacking. The C-terminal half of TFIIIEα is likely unstructured and is dispensable for yeast viability (33). In contrast, the N-terminal half of TFIIIEα is essential for cell growth (33) and contains a leucine-rich region with weak sequence similarity to the bacterial σ factor, followed by a zinc-binding domain (38). In agreement with in vivo data, the N-terminal half of TFIIIEα is required for basal and activated transcription in vitro (39) and interacts with TBP and pol II (29).

Archaea contain a TFIIIE homolog, called TFE, that corresponds to the essential N-terminal half of TFIIIEα (40, 41). Also found in archaea are homologs of TBP and TFIIIB, but not of TFIIIF and TFIIH (42). Whereas the archaeal TBP and TFIIIB homologs are essential for transcription initiation, TFE is not absolutely required but stimulates transcription in vitro (40, 41). Because mutagenesis of the T box or low concentrations of TBP sensitize a promoter to TFIE in vitro, TFE apparently stabilizes the archaeal TBP-DNA complex (41). TFE also interacts directly with archaeal TBP and RNA polymerase (41).
Here we have defined and crystallized the N-terminal domain of TFE from the archaeon Sulfolobus solfataricus. X-ray analysis reveals that this domain adopts an extended winged helix fold with unusual features. Analysis of surface properties, comparisons with other winged helix domains, and biochemical data suggest that the function of the TFE winged helix domain is not primarily DNA binding. The available data are, however, consistent with a role of this domain as an adaptor between RNA polymerase and general transcription factors.

EXPERIMENTAL PROCEDURES

Cloning and Site-directed Mutagenesis—The TFE DNA fragments TFEΔC and TFEΔCΔL, which encode for TFE residues 1–110 and 1–97, respectively, were subcloned using as template DNA pET-sTFE, which encodes for full-length S. solfataricus TFE (41). Plasmid pET-sTFE was a generous gift of Stephen Bell. For cloning of fragment TFEΔCΔL, oligonucleotide primers TFE1 (5′-TATATCATATGATGGTTAACGCA-GAGGATCTGTATT-3′) and TFE2 (5′-TATCGGCCCGCCTATCATATTCTTTCTTATTACAGC-3′) were used in PCR amplification. Primer TFE1 adds an Ndel restriction site upstream of the start codon. For amplification of TFE fragment TFEΔCΔL, primer TFE1 was replaced by TFE2 (5′-TATCTGGGCGCCTATCATATTCTTTCTTATTACAGC-3′). Primers TFE1 and TFE2 encoded an overhang for a NotI restriction site, which did not introduce a stop codon. Ligation of the restriction digested PCR products into pET21b (Novagen) added to the gene set for a C-terminal hexahistidine tag followed by a stop codon. Cloning was performed according to standard procedures. To introduce additional methionine residues into the polyepitope chain encoded by gene TFEΔCΔL, codons for hydrophobic residues were changed by site-directed mutagenesis using the two-step PCR overlap extension method, as described in a previous study (43). Several mutant forms of TFEΔCΔL have been constructed, including the single mutant F76M, the double mutant I39M/F76M, and the triple mutant I39M/F62M/F76M.

Protein Expression, Selenomethionine Labeling, and Purification—Plasmid DNA encoding for wild-type TFEΔC or TFEΔCΔL protein was transformed into Escherichia coli strain B834(DE3) (Stratagene). Cells were grown at 37 °C in LB medium supplemented with ampicillin and chloramphenicol at concentrations of 100 μg/ml each. Induction was performed during logarithmic growth by addition of 0.5 μM isopropyl-β-D-thiogalactopyranoside, and bacteria were grown for an additional 3.5 h. Cells were harvested by centrifugation and resuspended in buffer A (50 mM Tris-Cl, pH 8.0, 300 mM NaCl, and 10 mM β-mercaptoethanol). For selenomethionine labeling of TFEΔCΔL protein, TFEΔCΔL was expressed in E. coli strain B834(DE3) (Novagen). Bacteria were grown in LB medium with 100 μg/ml ampicillin at 37 °C to an A600 of 0.6. Cells were harvested and resuspended in the same buffer supplemented with seleno-methionine and ampicillin at concentrations of 100 mg/ml each. The optical density was measured, and cells were grown until the optical density at λ = 600 nm increased by 0.2 unit, to deplete the cells from any residual methionine from the standard LB medium. Protein expression was induced by addition of 0.5 μM isopropyl-β-D-thiogalactopyranoside, and cells were grown for an additional 4 h at 37 °C. Cells were harvested and resuspended in buffer A.

Native and selenomethionine-labeled proteins were purified according to the same procedure. Cells were broken with a French press, and the suspension was clarified by centrifugation. The cleared supernatant was incubated for 20 min at 72 °C, to denature the cell envelope of E. coli proteins. After centrifugation, the cleared lysate was loaded onto a 1-ml nickel-nitrilotriacetic acid column (Qiagen), equilibrated with buffer A. The column was washed with 20 ml of buffer A, followed by 4 ml of buffer A containing additionally 40 mM imidazole. Bound proteins were eluted with buffer A containing 250 mM imidazole and were judged to be ≥99% pure. The proteins were then purified to homogeneity on a Superose-12 HR gel filtration column (Amersham Biosciences) equilibrated with buffer B (50 mM Hepes, pH 7.5, 200 mM NaCl, 1 mM EDTA, 10 μM ZnCl2, 5 mM dithiothreitol). Peak fractions were pooled and concentrated to 15 mg/ml. Protein concentrations were measured using Bradford reagent (Bio-Rad). Whereas TFEΔCΔL and its variants remained soluble during purification, TFEΔCΔL was found in the insoluble pellet after centrifugation.

Crystallization and Data Collection—Purified protein samples were subjected to commercial crystal screens (Hampton Research) using the hanging drop vapor diffusion method. In initial screens, TFEΔCΔL formed large clusters of thin needles within a week under three different conditions. The reservoir solutions were (i) 100 mM Tris-Cl, pH 8.5, 200 mM LiSO4, 30% (v/v) polyethylene glycol (PEG) 4000, (ii) 100 mM Tris-Cl, pH 8.5, 200 mM NaOAc, 30% (v/v) PEG40000, and (iii) 20% (v/v) ethanol, 10% (v/v) glycerol. Optimization around the second condition yielded thicker needles (smallest dimension, ∼200 μm). Crystals of selenomethionine-labeled proteins were grown under the same conditions. Crystals were harvested in mother liquor and were transferred to a solution containing additionally 15% (v/v) PEG400. Crystals were incubated for 10 min, flash-frozen in liquid nitrogen, and stored for data collection at the synchrotron. Multwavelength anomalous diffraction (MAD) experiments were performed on crystals from both native and selenomethionine-labeled proteins. X-ray diffraction data from protein crystals of native TFEΔCΔL and selenomethionine-labeled TFEΔCΔL variant I39M/F76M were collected at the Swiss Light Source (Table I). Further MAD experiments on protein crystals of TFEΔCΔL variants F76M and I39M/F62M/F76M were performed at beam line ID14-4 at the European Synchrotron Radiation Facility (ESRF) in Grenoble (Table I).

Crystal Structure Determination—X-ray diffraction data of wild type and native protein crystals and crystals of selenomethionine-labeled protein from TFEΔCΔL variants F76M and I39M/F76M were integrated and scaled using the programs DENZO and SCALEPACK (45). MAD data from protein crystals of variant I39M/F62M/F76M were reduced using the program denSolve (46). Selenium sites were found by using the programs XDS (47, 48), and the solution was confirmed with program SnB (49). The obtained selenium sites were used in phasing with SHARP (50) and were refined with RESOLVE (47, 48) with automated model building turned off.

Additional methionines were introduced at positions of conserved hydrophobic residues, some of which are methionines in TFE homologs. After introduction of one methionine in addition to the native M1 and M34 (TFEΔC variant F76M), we could consistently identify two selenium peaks with SnB, but no interpretable electron density was obtained after refinement with SHARP. Structure determination was tried in both enantiomorph space groups (P6, 22 and P6, 22). Introduction of a second additional methionine (double mutant I39M/F76M) did not help. Further MAD experiments led to the detection of three selenium peaks by SnB and SOLVE (Z-score 13.3), of which two matched the peaks in variant F76M. After phasing with SHARP, a noisy electron density was obtained that showed segments of α-helices. However, no continuous trace could be built into this density. Finally, a third additional methionine was introduced. MAD data from crystals of the triple mutant I39M/F62M/F76M gave rise to an additional selenium peak but still resulted in a poor electron density map. The known three selenium sites could be identified, the Z-score increased to 23.5 (mean figure of merit = 0.55, CCNnot_Fourier = 0.17), and the electron density improved dramatically. However, a fourth selenium site was still missing, and no continuous trace could be built into the density. Refinement was achieved when a fourth selenium site was identified with SHARP, and the obtained phases were further improved with RESOLVE.

The resulting electron density, in combination with the selenium atom positions used as sequence markers, allowed for building of an atomic model with program O (51) that included TFE residues 3–88 except residues 70–72 in a flexible loop region (Fig. 1). After one round of refinement against the remote dataset, the initial model was repositioned and reoriented in the slightly different unit cell of the native crystal by rigid body refinement with CNS (52). After positional and B factor refinement, model-phased maps allowed for building of the N-terminal residues 1–2, but still there was no electron density observed covering the flexible residues 70–72 and for the C-terminal region. Several cycles of model rebuilding and refinement with CNS were performed. Refinement by simulated annealing was used in the beginning, but later only positional refinement was carried out. Before each refinement cycle, all atomic temperature factors were set to 50 Å2 and were refined again before calculation of a new electron density map. During refinement, the automated solvent correction in CNS failed, most likely due to the high Rmerge values at low resolution. We therefore used only data in the resolution range 6.5–2.9 Å during refinement. The model refinement converged with a free R factor of 32.1% (33.9% if the problematic low resolution data are included). The relatively high model R factors may be explained by the moderate resolution of our data, a high positional B factor, and by the high percentage of disordered residues (29%). Except residues Val-2 and Thrp-75, which have high B factors, all residues fall within the allowed regions of the Ramachandran plot. The coordinates have been deposited with the Protein Data Bank (PDB accession code 1Q1H).
To ensure that the relatively high R factors are not due to low quality of the native data set, we have also refined the structure against the remote wavelength of the MAD data set. To this end, the free R factor flags were transferred and three residues were mutated to methionines. Several cycles of slight manual adjustment of the structure and refinement of atomic positions and B factors with CNS quickly converged on a refined model that has very similar R factors (Table I). A superposition of the two refined models revealed that they are essentially identical within the experimental error, except for minor differences around the sites of mutation.

DNA-binding Assay—Band shift assays were carried out essentially as in a previous study (53). Briefly, purified TFEΔC was added to 100 pmol of DNA in a 100- and 200-fold molar excess in buffer TGMEM (50 mM Tris-Cl, pH 7.5, 380 mM glycine, 2 mM EDTA, 4% (w/v) PEG6000, 20% (v/v) glycerol, and 16 mM magnesium acetate). The reaction mixture was incubated at 20 °C for 30 min before 10 µl of sample loading buffer (TGMEM with 30% glycerol and bromphenol blue) was added. The mixtures were then separated on a 5% native polyacrylamide gel (95 V for 90 min at 20 °C). The gels were stained with the highly sensitive DNA stain SYBR Gold (Molecular Probes, Leiden, Netherlands).

RESULTS AND DISCUSSION

TFE Contains a Stable N-terminal Domain—The modular and flexible nature of TFIIE has hampered crystallization so far. To obtain structural information for the essential N-terminal region of TFIIEα we turned to the homolog TFE from the archaeon S. solfataricus. Full-length S. solfataricus TFE was expressed in E. coli and purified as described (41), with minor modifications (see “Experimental Procedures”). Because initial crystallization screens with the full-length protein were unsuccessful, we conducted partial proteolysis of purified TFE with four proteases of different specificity (trypsin, chymotrypsin, elastase, and subtilisin). These experiments always resulted in a stable N-terminal fragment (verified by Edman sequencing, not shown), which lacks a C-terminal region that includes a potential zinc-binding domain that is apparently not well structured in S. solfataricus TFE (Fig. 2A). Although binding of zinc ions by this domain is essential for function of TFE (40) and TFIIE (54), several TFE sequences, including S. solfataricus TFE, lack one of the four cysteine residues involved in zinc coordination (55).

Crystal Structure Determination—Based on the proteolysis results, we constructed the TFE variant TFEΔC, which comprises residues 1–110 (see “Experimental Procedures”). TFEΔC was expressed in E. coli with high solubility. The purified TFEΔC variant crystallized in two morphologies. Individual cryo-cooled crystals of both morphologies diffracted synchrotron radiation to 2.8-A resolution at best and turned out to have the same symmetry and comparable unit cell dimensions.

Trials to derivatize TFEΔC crystals with heavy metals were unsuccessful, most likely due to a lack of cysteine residues. Structure determination by selenomethionine incorporation and multiwavelength anomalous diffraction (MAD) was also hampered, because TFEΔC comprises only two methionine residues, M1 and M34, that did not provide sufficient signal for phasing. To increase the signal for MAD phasing, we introduced additional methionines in a sequential, non-disruptive manner (see “Experimental Procedures”). Sufficient signal for MAD phasing was only obtained after three additional methionines were introduced. MAD phasing with the selenomethionine-labeled triple mutant I39M/F62M/F76M led to an electron density map that allowed building of an initial model (Fig. 1). This initial model was used to phase native data from a wild-type TFE crystal, leading to an improved electron density map. A model comprising TFE residues 1–88 (except a disordered loop containing residues 70–72) and a total of 701 non-hydrogen atoms could be built and refined at 2.9-A resolution (Table I).

The observed difficulties during phase determination may have been caused by relatively high Rsym values of about 6% in the lowest resolution shell. The elevated Rsym values were independent of the x-ray source and detector type and were not due to incorrect assignment of space group, because they remained high after processing the data in space groups of lower symmetry. The data intensity distribution exactly followed theoretical Wilson statistics, suggesting that the elevated Rsym values are also not due to partial crystal twinning. The phenomenon may, however, be explained by the high percentage of disordered residues (29%, see below) and the low solvent content of the crystals (34%).

TFE Forms an Extended Winged Helix Domain—The crystal structure shows that the N-terminal TFE domain adopts a winged helix fold (Fig. 2C), which is frequently found in transcription factors and other nucleic acid-binding proteins (56–58). To detect similarities of the TFE structure with known molecular structures we submitted the atomic coordinates to the DALI server (59). This analysis revealed that most proteins with significant structural similarity to TFE are DNA-binding winged helix proteins, including transcriptional regulators and enzymes involved in nucleic acid metabolism. Also among the 40 proteins with the highest DALI score were histone H5, the catabolite activator protein, domains 2 and 4 of the bacterial σ factor, and the winged helix domains of the TFIIF Rap30 subunit and TFIIEβ. The winged helix fold of TFE comprises three α-helices and three β-strands in the canonical order a1-β1-a2-a3-β2-β3 (Fig. 2, B and C). Conserved residues within helices
Fig. 2. Structure of the TFE winged helix domain. A, domain architecture of *S. solfataricus* TFE and human TFIIEα. A potential zinc-binding region (Zn), a helix-turn-helix motif (HTH), an alanine-rich region (Ala), a region rich in serines, threonines, aspartates, and glutamates (STDE), and an acidic region (Acidic) are indicated. Regions in TFIIEα that are required for basal transcription, interaction with...
The extended strands form a protruding hydrophobic core of the winged helix domain. Among these residues are several leucines that are repeated in the TFE sequence with a spacing of seven amino acid residues, previously suggesting a possible leucine zipper structure that could confer dimerization (38).

A specific feature of the structure is the extension of the canonical winged helix fold at the N and C termini by the additional helices α0 and α4, respectively (Fig. 2C). Hydrophobic residues from the additional helix α0 extend the hydrophobic core of the winged helix domain, and helix α0 is tightly packed against the canonical winged helix fold. Helix α4 comprises only one turn and is ordered only until residue 88, although residues 89–110 are present in the crystals and secondary structure prediction suggests that the helix extends until residue 110. Because the crystal packing would allow for extension of helix α4 and its protrusion into the solvent region, it is likely that residues 89–110 extend helix α4 but that they are not observed due to disorder in the crystal. Truncation of the TFE C terminus until residue 97 results in an insoluble protein variant, possibly because helix α4 is disrupted. A protruding helix α4 could form a semirigid linker to the C-terminal zinc-binding domain of TFE, or it could be involved in TFE dimerization, although there is no evidence for dimerization in the crystals.

Unusual Surface Features and Potential Interaction Sites—The extended strands β2 and β3 form a protruding β-hairpin. The β2–β3 connecting loop is partially disordered and corresponds to wing 1 of the canonical winged helix fold (Fig. 2C). The wing 1 protrusion in TFE is unusually long and has a basic and an aromatic face, formed by residues in strands β2 and β3, respectively. The aromatic residue Trp-75 in strand β3 is fully solvent-exposed and may play a functional role. Residue Trp-75 is involved in crystal packing and influences the position of the wing 1 protrusion. The aromatic face of the wing 1 protrusion and helix α3 line a negatively charged groove (Fig. 2, D and E). The molecular surfaces of the groove, the wing 1 protrusion, and the central part of helix α3 are conserved among archaeal TFE proteins (Fig. 2, B and D), indicating that this side of the domain is involved in functionally relevant interactions. Several salt bridges are formed on the domain surface, which could contribute to protein stability at high temperature and low pH, typical for the growth environment of S. solfataricus.

There are two hydrophobic surface patches that may serve as protein interaction sites. The larger of these hydrophobic patches is formed by three conserved residues on the exposed side of the additional helix α0, located about 30 Å away from the groove (Fig. 2, D and E). A small hydrophobic patch is found on the side of helix α3 opposite the groove. This patch is formed by Leu43 and Ile45, which are partially buried and also contribute to the hydrophobic core (Fig. 2D).

TFIIIEβ, and stimulation of C-terminal repeat domain kinase activity (39) are indicated with gray lines. B, alignment of the TFE winged helix domain amino acid sequence with corresponding regions in other archaeal TFE sequences, and in the human and yeast TFIIIE α subunit. Secondary structure elements are indicated above the sequences (α-helices, magenta cylinders; β-strands, cyan arrows). Marked are residues in the hydrophobic core (V), residues forming the hydrophobic patch on helix α0 (I), arginine 51 (●), and three conserved acidic residues (X°). Residues that were mutated to methionines for structure determination are marked below the alignment (○). The amino acid sequence alignment of archaeal TFE homologs was produced with ClustalW (65). Sequences of human and yeast TFIIIEα were adjusted manually, aided by secondary structure prediction (PHD, 64). The graphic was prepared with ALSRIPT (71). C, ribbon representation of the structure of the TFE winged helix domain. Secondary structure elements are colored and labeled according to B. The graphics were prepared with programs BOBSCRIPT (72), MOLSCRIPT (73), and DINO (Visualizing Structural Biology (2002), www.dino3d.org). D, surface conservation. Surface residues conserved among archaea are colored according to the scheme in B. Arginine 51, a patch of three conserved acidic surface residues, and the negatively charged groove are marked according to B. Surfaces were generated with the program MSMS (74). E, surface charge distribution. The domain surface is colored according to the electrostatic potential from negative (red) to positive (blue). Electrostatic surfaces potentials were calculated with MEAD (75). The top and bottom views are related by a 180° rotation around the vertical axis.
ture to known structures of winged helix protein-DNA complexes (one is shown in Fig. 3). Most DNA-binding winged helix domains studied to date interact with DNA via the “recognition” helix α3 and via wing 1 and generally show a continuous positively charged surface on the DNA binding side (57). Examples for such typical DNA-binding winged helix proteins are HNF-3γ (60), E2F-4 fragment (61), and BmrR (62). The winged helix domain of hRFX1 also uses helix α3 and wing 1 for DNA contacts, but the DNA is positioned differently (63).

The side of the TFE domain that corresponds to the DNA-binding face of typical winged helix proteins shows considerable surface conservation, in agreement with a possible role in DNA binding (Fig. 2D). Helix α3 is positively charged, as required for DNA binding (Fig. 2E). Within this helix, arginine Arg-51 is conserved in all TFE and TFIIEα sequences except for the yeast sequence (Fig. 2B), and a DNA-binding arginine is found at a corresponding location in the winged helix domain of the transcriptional activator BmrR (62). However, the groove between helix α3 and the wing 1 protrusion, which typically accommodates the DNA backbone, is negatively charged in the TFE domain. In particular, the entrance to the groove is formed by a conserved acidic patch in loop β1-α2 (Fig. 2D). Furthermore, wing 1 in TFE is unusually long in comparison with typical DNA-binding winged helix domains, and it is even longer in TFIIEα domains (Fig. 2B). These comparisons predict that the extended wing 1 and the acidic groove interfere with DNA binding and make it unlikely that TFE binds DNA like a typical DNA-binding winged helix protein, despite some features that suggest this.

To test directly whether the TFE winged helix domain binds to nucleic acids in vitro, we carried out electrophoretic mobility shift assays with purified recombinant TFE (see “Experimental Procedures”). We tested various types of synthetic nucleic acid constructs for TFE binding, including single-stranded DNA, a double-stranded blunt end DNA, a double-stranded DNA with a single-strand overhang, and a DNA mismatch bubble. In these experiments we could not detect significant DNA binding by the TFE winged helix domain, even when a 200-fold molar excess of TFE was added to 100 pmol nucleic acids (not shown). It is however possible that cooperative interactions within the transcription initiation complex allow for a contribution of the TFE domain to promoter DNA binding.

Conservation of the Winged Helix Domain in Eukaryotic TFE—Archaeal TFE sequences are highly homologous and can be aligned easily, but it is difficult to align these with the corresponding eukaryotic TFIIEα sequences, because the sequence conservation is weak. However, the TFE structure and secondary structure prediction of eukaryotic TFIIEα (PHD (64)) suggested how to modify a ClustalW-generated (65) alignment of TFE with human and yeast TFIIEα sequences (Fig. 2B). The modified alignment was then used to produce homology models of the human and yeast TFIIEα winged helix domains. Inspection of the modeled structures revealed reasonably packed hydrophobic cores and no severe side chain clashes, indicating that the alignments are correct. Residues of the TFE winged helix domain that are identical in human TFIIEα include Leu-24, Ile-26, Leu-27, Leu-43, Leu-54, and Phe-62, which participate in the extended hydrophobic core.

The TFIIEα homology models also reveal that key surface features of archaeal TFE are apparently conserved in eukaryotic TFIIEα. First, the surface charge distribution of S. solfataricus TFE is conserved in yeast and human TFIIEα from helix α3 to strand β3 (not shown). Second, residues Glu-37, Arg-51, and Asn-55, located in the groove between wing 1 and helix α3, are identical between TFE and human TFIIEα. Third, the two hydrophobic surface patches of TFE are conserved in eukaryote TFIIEα, including the three exposed hydrophobic residues in helix α0 and the invariant Gly-17 in the subsequent loop. Because conservation of the hydrophobic patch on helix α0 can not result from a structural requirement, this patch is likely a protein docking site in all TFE and TFIIEα winged helix domains. Taken together, this analysis shows that the TFE structure is a valid model for the corresponding domain in eukaryotic TFIIEα and suggests that the described key surface features are functionally relevant.

Comparison with Other Winged Helix Domains in General Transcription Factors—Our finding extends the number of winged helix domains in general transcription factors to a total...
The winged helix domain does not directly correspond to that of a DNA. It is thus unlikely that TFE contacts promoter DNA in an hydrophobic core of the domains (not shown). The preservation is generally limited to important residues in the reported that a similarly weak sequence homology between the interaction sites of all four winged helix domains in general transcription factors also differ. The DNA-binding faces of TFIIA Rap30 and TFIIIE are located on opposite sides (22, 37). The site for interaction of the Fcp1 peptide with the TFIIA Rap74 winged helix domain is the groove between helix α3 and wing 1 (Fig. 3). The same groove in TFE may be used for protein interactions, but additional putative interaction sites exist (see above). Taken together, this comparison demonstrates that the winged helix fold is very versatile, even among the four domains in general transcription factors.

Comparison with the Bacterial σ Factor—It was noted that TFIIIE shows weak sequence similarity to regions 2 and 4 of the bacterial σ factor (38). Structures of σ domains have been solved (66, 67), and structures of σ bound to the bacterial polymerase core and to the polymerase core and promoter DNA have also been determined recently (68–70). Superposition of the TFE domain onto σ domains 2 and 4 shows that the conservation is generally limited to important residues in the hydrophobic core of the domains (not shown). The σ domains do not contain a wing and no equivalent to helix α0. In the context of the bacterial polymerase-σ-DNA complex, σ domain 2 binds to DNA near the point of melting, and domain 4 binds to the promoter element at position −35. Superposition of the TFE domain on the two σ domains in the context of the polymerase and DNA shows clear differences in residues in the vicinity of DNA. It is thus unlikely that TFE contacts promoter DNA in an equivalent manner in the archaeal and eukaryotic systems.

This comparison suggests that the function of the TFE winged helix domain does not directly correspond to that of a σ domain. Instead, the sequence homology between the TFE domain and the two σ regions apparently reflects the conservation of structural residues in a related helical fold. We recently reported that a similarly weak sequence homology between σ and the pol II subunit Rpb4 does not reflect a functional equivalence either (26). Despite differences in structure, key functional aspects of protein factors involved in transcription could nevertheless be similar between eukaryotes and bacteria.

Overall Structure of TFIIIE in the Initiation Complex—Our results are consistent with a model of the TFE/TFIIIEα winged helix domain as a bridging factor or adapter between TBP, the polymerase, and possibly promoter DNA. TFE interacts physically both with TBP and with archaeal RNA polymerase (41). These protein interactions may be achieved by the conserved surface features of the TFE winged helix domain, including the two hydrophobic surface patches and a negatively charged groove. TFE can promote transcription apparently due to stabilization of the TBP-DNA interaction (41), suggesting that TFE could contact DNA. Although our results apparently exclude a canonical DNA interaction, the TFE domain shows some features of DNA-binding winged helix domains that may contribute to context-dependent promoter DNA binding.

The adapter function of the winged helix domain may be the basic conserved function of TFE/TFIIIE, because the winged helix domain is common to archaea and eukaryotes. In contrast, the C-terminal region of TFIIIEα and the TFIIIEβ subunit only occur in eukaryotes, and appear to play eukaryote-specific roles, as reflected by their cooperation with TFIIIF and TFIIH, factors that do not exist in archaea. Whereas the TFE/TFIIIEα winged helix domain could bind between TBP and pol II, the TFIIIEβ subunit may bind to promoter DNA in and around the early transcription bubble (35–37). Open questions on the TFIIIE structural mechanism include the basis of TFIIIE subunit dimerization and the exact location of the TFIIIE domains with respect to other components of the transcription machinery.

Acknowledgments—We thank Clemens Schulze-Briese for help at the protein crystallography beamline X06SA at the Swiss Light Source. We thank Edward Mitchell (ESRF Grenoble) and Kornelius Zeth (Max-Planck-Institute, Martinsried) for help at beamline ID14-4 at the ESRF Grenoble. We thank S. Bell (Medical Research Council, Cancer Cell Unit, Cambridge) for the initial TFE expression plasmid. We thank Karl-Peter Hopfer (Gene Center, Munich) for help with mass spectrometry and data collection. We thank Frank Siedler (Max-Planck-Institute Biochemistry, Martinsried) for help with mass spectrometry. We thank Hubert Kettenberger for help with DNA binding assays. We thank members of the Crammer laboratory for help and comments on the manuscript.

REFERENCES

1. Kornberg, R. D. (1999) Trends Cell Biol. 9, 345–349
2. Lee, T. I., and Young, R. A. (2000) Annu. Rev. Genet. 34, 77–137
3. Conaway, R., and Conaway, J. (1997) Prog. Nucleic Acids Res. Mol. Biol. 56, 177–246
4. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989) Cell 56, 549–561
5. Buratowski, S. (1994) Cell 77, 1–3
6. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ehrlich, R. H. (1998) Genes Dev. 12, 34–44
7. Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horkislohi, M., Cramer, P., Bushnell, D. A., and Kornberg, R. D. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 8174–8178
8. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Nature 365, 512–519
9. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) Nature 365, 520–527
10. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisataka, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) Nature 377, 119–128
11. Koss, P., Ghosh, G., DeDecker, B. S., and Sigler, P. B. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 6042–6047
12. Littlefield, O., Korkhov, Y., and Sigler, P. B. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 13668–13673
13. Tan, S., Hunziker, S., Sargent, D. F., and Richmond, T. J. (1996) Nature 381, 127–151
14. Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) Science 272, 830–836
15. Zhu, W., Zeng, Q., Colangelo, C. M., Lewis, M., Summers, M. F., and Scott, S. G. (1996) Nat. Struct. Biol. 3, 122–124
16. Nguyen, B. D., Abbott, K., Potemka, K., Kober, M. S., Arambauj, J., Greenblatt, J., Legault, P., and Omichinski, J. G. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 5688–5693
17. Kamada, K., Roeder, R. G., and Burley, S. K. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 2296–2299
18. Nguyen, B. D., Chen, H. T., Kober, M. S., Greenblatt, J., Legault, P., and Omichinski, J. G. (2003) Biochemistry 42, 1460–1469
19. Kamada, K. De Angelis, J., Roeder, R. G., and Burley, S. K. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 3115–3120
20. Gaiser, F., Tan, S., and Richong, T. J. (2000) J. Mol. Biol. 302, 1119–1127
21. Cramer, P., Bushnell, D. A., Fr., J., Gnant, A. L., Mairer-Davis, B., Thompson, N. E., Burgess, R. B., Kornberg, R. D., and David, P. R., and Kornberg, R. D. (2000) Science 288, 640–649
22. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Science 292, 1483–1486
23. Cramer, P. (2002) Curr. Opin. Struct. Biol. 12, 89–97
24. Armache, J.-K., Kettnerberger, H., and Cramer, P. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 6964–6968
25. Bushnell, D. A., and Kornberg, R. D. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 6969–6972
26. Ohkuma, Y. (1997) J. Biochem. 122, 481–489
27. Green, M. E., Goodrich, J. R., and Yoo, J. (1998) Genes Dev. 8, 515–524
28. Svedrup, J. Q., Vichi, P., and Egly, J. M. (1996) Trends Biochem. Sci. 21, 346–350
29. Cox, F., and Egly, J. M. (1998) Cold Spring Harbor Symp. Quant. Biol. 63, 105–110
30. Ohkuma, Y., and Roeder, R. G. (1994) Nature 368, 160–163
31. Kuldell, N. H., and Buratowski, S. (1997) Mol. Cell. Biol. 17, 5288–5298
32. Holstege, F. P., Tantin, D., Carey, M., van der Vliet, P. C., and Timmers, H. T. M. (1995) EMBO J. 14, 810–819
33. Kim, T. K., Ehrlich, R. H., and Reinberg, D. (2000) Science 288, 1418–1422
