The Structure of the TFIIH p34 Subunit Reveals a Von Willebrand Factor A Like Fold

Dominik R. Schmitt, Jochen Kuper, Agnes Elias, Caroline Kisker

Rudolf Virchow Center for Experimental Biomedicine, Institute for Structural Biology, University of Wuerzburg, Wuerzburg, Germany

Abstract

RNA polymerase II dependent transcription and nucleotide excision repair are mediated by a multifaceted interplay of subunits within the general transcription factor II H (TFIIH). A better understanding of the molecular structure of TFIIH is the key to unravel the mechanism of action of this versatile protein complex within these vital cellular processes. The importance of this complex becomes further evident in the context of severe diseases like xeroderma pigmentosum, Cockayne’s syndrome and trichothiodystrophy, that arise from single point mutations in TFIIH subunits. Here we describe the structure of the p34 subunit of the TFIIH complex from the eukaryotic thermophilic fungus Chaetomium thermophilum. The structure revealed that p34 contains a von Willebrand Factor A (vWA) like domain, a fold which is generally known to be involved in protein-protein interactions. Within TFIIH p34 strongly interacts with p44, a positive regulator of the helicase XPD. Putative protein-protein interfaces are analyzed and possible binding sites for the p34-p44 interaction suggested.

Introduction

The TFIIH complex is a multi-subunit protein assembly involved in transcription and DNA repair. It consists of 10 different proteins forming two sub-complexes. The 7-subunit TFIIH core comprises the two helicases XPD and XPB, as well as p62, p52, p44, p34 and p8. In addition, MAT1, CycH and cdk7 form the Cyclin-Activating-Kinase complex (CAK), which is attached to the core via a direct interaction of XPD and presumably XPB with MAT1 [1,2]. Recently, additional subunits like XPG and Tfb6 have been suggested [3,4] and ancillary interacting partners were identified [5], emphasizing the complexity of TFIIH and the high level of dynamics to participate in different processes. While both the TFIIH core and the CAK subcomplex are essential for RNA Polymerase II mediated transcription, the release of CAK from TFIIH promotes the excision of DNA lesions during the nucleotide excision repair (NER) pathway [6]. Mutations in XPD, XPB and p8, affecting these processes, are known to be causative for xeroderma pigmentosum (XP), Cockayne’s syndrome (CS) and trichothiodystrophy (TTD), severe diseases associated with TFIIH function [7–9]. In addition, mutations in XPB, p62 and p52, lead to photosensitivity and XP/TTD like phenotypes in Drosophila [10–12].

The core TFIIH assembly is highly conserved among eukaryotes, with only p8, p52 and p62 lacking in a few species [13]. Of all its subunits XPD and XPB exhibit the highest sequence conservation, while p62 is most divergent [13]. Within TFIIH the helicase and ATPase activities of XPD and XPB are tightly regulated by their associated partners p44 and p52, respectively [14,15], both of which also interact with other subunits. The interaction of p52 with p8, the smallest of the core subunits, adds to the XPB stimulation and is essential for TFIIH stability and NER activity [16]. Moreover, p44 interacts with the p34 subunit and the depletion of this subunit from the complex leads to reduced DNA repair capacity [17]. Over the past two decades the intricate network of regulation within TFIIH has been a major research focus, especially regarding the XPD/p44 and XPB/p52/p8 subunits. However, only very little is known with respect to the other core subunits. Apart from its role in stimulation of the XPD helicase [15], an E3 ubiquitin ligase activity has been proposed for the p44 yeast homologue Ssl1 [18] whereas no enzymatic function is known for the p34 subunit of TFIIH, although it has been implicated in the process of mRNA-splicing, based on phylogenetic considerations [13].

Here we describe the first crystal structure of the TFIIH subunit p34 from C. thermophilum (ct), comprising its N-terminal domain, which is known to strongly interact with the C-terminal C4C4 RING domain of p44 [19,20]. The structure reveals a von Willebrand Factor A (vWA) like fold, typical for other vWA containing proteins involved in a multitude of protein-protein interactions. Structural comparison allowed us to delineate similarities as well as differences to already known vWA domains, providing insight into the role of p34 within TFIIH. In addition, a putative binding site for p44ct could be mapped onto the p34ct vWA domain, based on its electrostatic surface potential and previous studies.
Materials and Methods

Protein Expression and Purification

The genes for p34ct full-length (1–429) and p34ct 1–277 were cloned into the pBADM-11 vector (EMBL) using ligase independent SLIC cloning [21]. Protein expression was carried out in E. coli BL21-CodonPlus (DE3) RIL cells grown in Lennox LB medium. After induction with 0.05% L–(-)–arabinose at an OD600 of 0.8 the cells were further grown for 18 – 20 hours at 15 °C.

The full-length protein as well as the shortened variant were purified using Ni-metal affinity chromatography (Ni-TED, Macherey-Nagel) followed by size exclusion chromatography (HiLoad 16/60 Superdex 200 prep grade, GE Healthcare) in either 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP. The samples were concentrated via Vivaspin filtration units (Sartorius) and the concentration was spectrophotometrically determined based on their theoretical molar absorption coefficients of 28,420 M⁻¹ cm⁻¹ (p34ct) and 26,930 M⁻¹ cm⁻¹ (p34ct 1–277), respectively. After flash freezing in liquid nitrogen, the protein was stored at −80 °C.

Crystallization

Full-length p34ct was crystallized using 0.5 – 2.0 µl of the protein solution in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP, at a concentration of 10 – 12 mg/ml. The solution was equilibrated against the reservoir solution consisting of 100 mM Tris-HCl pH 8.0 and varying amounts of 18 – 32% PEG 550 MME, depending on the volume of protein solution per volume of reservoir in the mixture. Crystals of p34ct usually grew within 3 days at 20 °C to sizes of more than 50 × 200 × 200 µm³.

For data collection the crystals were flash frozen in liquid nitrogen either directly in their mother liquor or by using paraffin oil as a cryo protectant. In order to overcome the phase ambiguity some crystals were soaked in a 1.0 µl drop of 25% PEG MME 550, 100 mM Tris-HCl pH 8.0 and 100 mM KI for 1 – 5 min prior to flash freezing.

Crystals of p34ct 1–277 were grown directly from 2.0 – 4.0 µl of protein solution in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP at a concentration of 10 – 12 mg/ml, added to 0.5 µl of reservoir solution. The mixture was equilibrated against the reservoir solution containing 100 mM Tris-HCl pH 8.0 and varying amounts of 18 – 32% PEG 550 MME, depending on the volume of protein solution per volume of reservoir in the mixture. Crystals of p34ct usually grew within 3 days at 20 °C to sizes of more than 50 × 200 × 200 µm³.

For data collection the p34ct 1–277 crystals were washed in 1.0 µl protein buffer and prepared for flash freezing by adding twice 1.0 µl of a cryo solution containing 10 mM Tris-HCl pH 8.0, 75 mM KCl, 12.5% di-ethylene-glycol, 6.25% ethylene-glycol, 6.25% MPD, 6.25% 1,2-propanediol and 6.25% glycerol to the drop and then transferring them to 1.0 µl of the same cryo solution.

Data Collection and Structure Solution

Data collection of flash frozen crystals was performed at 100 K at beamlines ID 23.2 (ESRF) for the p34ct native data, BL 14.1 (BESSY) for p34ct anomalous KI data and ID 29 (ESRF) for p34ct 1–277 native data at wavelengths of 0.87 Å, 1.60 Å and 0.92 Å, respectively.

Both full-length p34ct and p34ct 1–277 crystallized in the cubic space group F4₁32 with unit cell parameters ranging from a = b = c = 257.1 Å to a = b = c = 257.3 Å. All data sets were indexed and processed with either IMOSFLM and SCALAA [22,23] or XDS [24]. De novo structure solution was achieved using SHELX C/D/E [25] employing a SIRAS approach by combining anomalous p34ct KI data at 4.2 Å with a highly isomorphous p34ct native data set at 3.0 Å. Initial electron density maps allowed to manually build a preliminary p34ct model using the programs O and COOT [26,27] and comprised the N-terminal domain of the protein. Model building was supported by results from secondary structure predictions utilizing the Phyre² algorithm [28].

The preliminary model was used to solve the higher resolution p34ct 1–277 structure by molecular replacement using Phaser [29]. The model of the p34ct N-terminal domain was adjusted and extended in COOT [27] and refined to a resolution of 2.8 Å using Phenix [30]. All Figures containing structures were generated using PyMOL [31].

Multi Angle Light Scattering

To determine the molecular mass of p34ct and p34ct 1–277 in solution, multi angle light scattering measurements (MALS, DAWN HELOS II, Wyatt Technology) combined with refractive index detection (Optilab t-rEX, Wyatt Technology) were performed [32,33]. The samples were diluted to a concentration of 30 µM in 20 mM CHES-NaOH pH 9.5, 150 mM KCl and 1 mM TCEP and loaded onto a Superdex 200 10/300 GL analytical size exclusion chromatography column (GE Healthcare), which was coupled to the MALS detector. The flow rate was set to 0.5 ml/min with an injection volume of 100 µl and the light scattering signal was collected at 293 K.

DNA Binding Assays

DNA binding of p34ct was investigated using native agarose gels in 25 mM Tris-HCl pH 8.5 and 19.2 mM glycerine. To prepare dsDNA substrates two 50-mer oligonucleotides (NDT, GACTACGTACTGCTTCGCCATGGCGAGATCTGC and NDB, GCCAGATCGCCATTGCGGTAGAGATGGAGAGCCGTACAGCTTGATGTC) were annealed by incubating the mixture in 100 mM KCl at 85 °C for 10 min followed by slowly cooling the samples to room temperature. For interaction studies 1 – 2 µM of ssDNA (NDB) or dsDNA (NDT/NDB) were used with 10 – 20 µM protein in a total reaction volume of 10 µl. The samples were incubated for 20 – 30 min at 4 °C and then supplemented by 10 µl loading dye, followed by an additional incubation for 10 min at 4 °C. Finally 20 µl of each sample were loaded onto 0.8% agarose gels, which were run at 50 V and 4 °C for 4 – 6 hours. Midori Green (Biozyn Scientific) was used to visualize the DNA. As positive control for ssDNA and dsDNA binding we used DNA Polymerase I from Bacillus caldotenax.

Results and Discussion

Crystallization and Structure Solution of p34ct and p34ct 1–277

Crystals of p34ct grew in the highly symmetric space group F4₁32 and diffracted to a resolution of 3.8 Å. To solve the phase problem we initially tried to use the intrinsic zinc signal of the protein’s C-terminal zinc-binding domain. However, despite the clear presence of zinc as inferred from the absorption signal at the zinc edge (not shown), this approach failed due to the low anomalous signal intensities. We finally succeeded in achieving a de novo solution using derivative data of a potassium iodide soaked crystal at 4.2 Å.

The initial low resolution electron density map revealed the presence of approximately half of the protein, with one molecule in the asymmetric unit. Based on secondary structure prediction results using the Phyre² algorithm [28] it was clear that p34ct contains an N-terminal domain and an additional C-terminal.
Table 1. Data Collection and Refinement Statistics.

| Data Collection | p34ct | p34ct KI | p34ct 1–277 |
|-----------------|-------|---------|-------------|
| Space Group     | F 4, 3 2 | F 4, 3 2 | F 4, 3 2 |
| Cell Dimensions | a, b, c [Å] 257.31, 257.31, 257.31 | 257.07, 257.07, 257.07 | 257.11, 257.11, 257.11 |
| α, β, γ [°]     | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution [Å]  | 3.8    | 4.2     | 2.8         |
| Wavelength [Å]  | 0.8726 | 1.6000  | 0.9199      |
| Unique Reflections | 7,694 | 10,090  | 18,512      |
| I/σI            | 16.2 (1.4) | 18.7 (2.1) | 20.9 (2.9) |
| Rmerge [%]      | 14.2 (142.1) | 14.5 (78.8) | 15.3 (259.8) |
| Completeness [%]| 99.8 (100.0) | 99.9 (100.0) | 100.0 (100.0) |
| Redundancy      | 10.1   | 46.4    | 70.5        |
| SigAno          | 1.288  |         |             |
| Refinement      | p34ct  | p34ct KI| p34ct 1–277 |
| Resolution      | 2.8    |         |             |
| Number of Atoms | 3,190 |         |             |
| Rwork (Rfree)   | 21.9 (23.8) |         |             |
| Mean B-Factor [Å²] | 96.5 |         |             |
| R.m.s. Deviations |       |         |             |
| Bond Lengths [Å] | 0.004 |         |             |
| Bond Angles [°] | 0.783 |         |             |
| Ramachandran Statistics | 98.5/1.5/0.0/0.0 | |  |

Values in parentheses refer to the highest resolution shell. Rmerg = \[\frac{\sum_{hkl} |I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl} I(hkl)}\] and Rpim = \[\frac{\sum_{hkl} |F_o(hkl) - F_c(hkl)|}{\sum_{hkl} F_c(hkl)}\] where \(F_o\) and \(F_c\) are the observed and calculated structure factors, respectively. \(R_{\text{merge}}\) is the same as \(R_{\text{merge}}\), calculated for the 5% of the data that was randomly omitted from refinement. The Ramachandran statistics indicate the fraction of residues in the most favored, allowed and disallowed regions of the Ramachandran plot, respectively, in addition to the fraction of rotamer outliers observed.

doi:10.1371 journal.pone.0102389.t001

domain including a C4 zinc-finger motif, the latter of which could not be located in the electron density maps. Analysis of the p34ct crystals by mass spectrometry, however, indicated that the crystals were composed predominantly of the full-length protein (data not shown). This led us to conclude, that the p34ct C-terminal domain is disordered within the crystal lattice and likely not involved in any crystal contacts, which is in line with the fact that zinc was present in the crystal, but did not give rise to a significant anomalous signal. Using C2-traces of structural homologues, suggested by the Phyre² analysis [28], we were able to manually built a preliminary model of p34ct’s N-terminal domain at 3.8 Å resolution. The anomalous derivative data used for phasing revealed a total of 8 iodide ions within the electron density map, bound to the N-terminal domain of the protein (Figure S1). The iodide ions were found at six different sites, with 4 out of the 8 ions being present in closely spaced pairs, showing a distance of 3–7 Å between the single ions (Figure S1, D and E).

Based on the knowledge gained through the preliminary p34ct structure, we designed a construct comprising residues 1 to 277 of p34ct, corresponding to its N-terminal domain and thus removing the flexible and unstructured part of the protein. The p34ct 1–277 variant also crystallized in space group F4132 with almost exactly the same unit cell dimensions of a = b = c = 257.1 Å (Table 1). However, diffraction was significantly improved and a highly redundant data set to a resolution of 2.8 Å could be obtained. The structure of the p34ct 1–277 N-terminal domain was solved by molecular replacement with the truncated preliminary p34ct model serving as template for phasing.

Overall, 17 amino acids at the N-terminus of p34ct 1-277, in addition to the hexa-histidine tag, as well as 2 long linker regions (residues 86 to 105 and 169 to 197) were not accounted for in the electron density map (Figures 1 and 2). The final model thus contains 209 out of the 277 residues and was refined to Rwork and Rfree values of 0.2187 and 0.2380, respectively, with good overall stereochemistry (Figure S2), 98.5% of all the residues are located in favored regions in the Ramachandran plot and no outliers are observed (Table 1).

Structure of p34ct 1–277

The N-terminal domain of p34ct consists of a central 6-stranded β-sheet with 5 parallel (β1, β2, β4, β5 and β6) and one shorter anti-parallel β-strand (β3), which is surrounded by a total of 3 α-helices on either side (α2, α3, α8 on one side and α5, α6, α7 on the other). Two additional, much shorter α-helices are located at the C-terminal end of the central β-sheet of the domain (α1 and α4, Figure 1). Surprisingly, the overall fold reveals a von Willebrand Factor A (vWA) like architecture (Figure 1B), with high structural similarity to the A1 domain of the von Willebrand Factor protein, a blood glycoprotein involved in hemostasis and platelet aggregation [34]. A vWA like domain has previously not been reported for p34. In line with this result our searches using the primary sequence of both human and C. thermophilum p34 did not return
a vWA like feature via BLAST and the SMART domain annotation database, respectively [35,36]. However, a Phyre2 secondary structure analysis did predict the vWA like fold and suggested several vWA containing structural homologues to p34 [28]. Interestingly, also the N-terminal half of p44, the binding partner of p34 within TFIIH, is predicted to contain a vWA like domain.

While the XPD, XPB and p44 subunits of TFIIH are relatively well conserved between different organisms, with mean residue identities of 52%, 50% and 35%, respectively, there seems to be higher variability in p34 [13]. Here, the mean sequence identity is reduced to 30% within a set of 63 different species [13] and reaches 28% when comparing the human to *C. thermophilum* p34. Interestingly, the *C. thermophilum* p34 is also larger than its human orthologue, with extensions at both N- and C-termini and flexible linker insertions. In total the size of p34 increases by 39% and reveals the presence of one variable linker and three linker insertions when compared to its human, mouse, yeast and plant homologues (Figure 2). The variable linker is located in a region connecting β-strand β3 to helix α3, which is disordered in our crystal structure (Figure 2, Linker 1). The three insertions encompass residues 166 – 199, 275 – 320 and 391 – 414 and are not present in the other species. The first of these is located in the vWA domain (Figure 2, Insertion 1). It connects helix α5 back to the β-sheet core and is again disordered in the p34ct structure. The second insertion bridges the N-terminal vWA domain and the highly conserved C-terminal region (Figure 2, Insertion 2). It spans more than 40 residues and most likely introduces high flexibility between the N-terminal and C-terminal parts of the protein. The third insertion is located at the far C-terminus beyond the C4 zinc-binding motif (Figure 2, Insertion 3).

**The p34ct 1–277 Structure in Light of other vWA Domains**

With the N-terminal domain of p34ct assuming a vWA like fold it was not surprising that a closer investigation using the DALI server yielded 140 similar structures with a Z-score higher than 10.0 and r.m.s.d. values between 2.1 and 3.3 Å [37]. Most of the molecules with the highest scores either correspond to the name giving human von Willebrand Factor (vWF) A1 domain itself [38] or complexes of this domain from mouse or human origin with a variety of different proteins [39–42].

---

**Figure 1. Overall structure of p34ct and analysis of the MIDAS region.** (A) Front view of the p34ct 1–277 vWA domain (left) and view from the top (right) with the color coding corresponding to the N-terminal (green), central (yellow) and C-terminal (orange) part of the domain, respectively. The location of the vWA domain within full-length p34ct is indicated in the equally colored bar representation. (B) The topology diagram of the vWA fold shows that its architecture can be distinguished from a typical Rossmann fold by β-strand β3, which is anti-parallel in all vWA domains.

doi:10.1371/journal.pone.0102389.g001
Superposition of the p34ct 1–277 structure with proteins suggested by the DALI search, revealed that the overall fold is highly conserved, even among rather distant homologues involved in different cellular pathways, such as the I-domain of the cell adhesion molecule Integrin α [43], the von Willebrand Factor A1 domain [38] and the 26S proteasome regulatory subunit Rpn10 [44], which all share the central 6-stranded β-sheet core, surrounded by 6 α-helices (Figures 3A, 3B and 3C, respectively).

However, there are also some remarkable structural features in p34ct 1–277 that have not been observed in the vWA domains of most other proteins. Helix α3 is significantly longer in p34ct 1–277, while helix α8 is relatively short and often found to be longer in other vWA containing proteins, except for Rpn10 and the vWA domain of the DNA dependent helicase Ku70 [45]. Furthermore, the two helices at the C-terminal end of the central β-sheet (α1 and α4, Figure 3) are often replaced by short loops in other vWA folds. The most striking feature, however, is helix α5, which is about twice as long compared to the corresponding helix in other vWA proteins, considerably protruding outward from the top side of the domain (Figure 3 A–C). The drastic extension of this helix might also explain the presence of the rather long linker required to provide the connection back to the β-sheet core, as observed in p34ct 1–277 (Figure 2, Insertion 1). However, as this linker is absent in p34 sequences of other organisms, it remains unclear, if helix α5 is similarly prominent in these species. Based on our p34ct vWA structure, multiple sequence alignments and secondary structure predictions, we propose helix α5 to be at least one or two turns shorter in human and mouse p34 and the p34 homologue Tfb4 in S. cerevisiae, respectively. Another possibility would be a shift of helix α5 more to the center of the domain to permit the connection to β4 (Figure 2).

While no vWA domain is known to bind nucleotides, many of them are capable of coordinating metal ions via a noncontiguous sequence motif termed metal-ion-dependent adhesion site (MIDAS). For example, all vWA domains in integrin α subunits and most integrin β representatives contain a perfectly conserved MIDAS motif, which is formed by 3 loops at the C-terminal end of the central β-sheet of the vWA fold [46]. The MIDAS motif consists of 5 residues in the order D-x-S-x-…T…D, which are in the following referred to as A1 to A5. In p34 from C. thermophilum (ct), homo sapiens (h) and several other species the aspartic acid at position A1 is perfectly conserved. However, in both p34ct and p34h the MIDAS motif is disrupted by a helical insertion (α1) directly after A1, which is often rich in one or two tryptophans, leading to non-conservative substitutions in A2 and A3 (Figure 3D). While A4 seems to be conserved, the aspartic acid at position A5 is replaced by a serine in p34. Hence, p34 is not likely to bind a metal ion via the MIDAS motif, which is consistent with our p34ct vWA structure where no bound ligands were observed.

Oligomeric State and DNA Binding Properties of p34ct in Solution

Although many studies, including Cryo-EM data [47,48], suggest the 7-subunit core of the general transcription factor TFIIH to consist of single subunits, the true oligomeric state of each subunit within TFIIH has not been resolved. Recently, Kainov et al. reported dimeric forms of both Tfb2 (p52) and Tfb5.
that undergo a dimer to hetero-tetramer transition upon interaction with each other [16,49]. As many proteins are known to self-associate in the absence of their respective binding partners [50], we analyzed the oligomeric state of full-length p34ct and p34ct 1–277 in solution using multi angle light scattering coupled to size-exclusion chromatography. The obtained data yielded molecular weights of 47 kDa and 32 kDa, respectively, closely matching the predicted values of 46,767 Da for p34ct and 31,941 Da in the case of p34ct 1–277 (Figure 4C and 4D). Hence, under the conditions tested, p34ct is monomeric in solution. In addition, an analysis of crystal packing interactions using the PDBePISA service [51] revealed no significant dimerization sites, which is in support of our solution data.

Given the DNA oriented role of the TFIIH complex in transcription and DNA repair, we also investigated the DNA binding capability of both p34 variants in solution. In case of the yeast homologue of p34, Tfb4, nucleic acid binding via a basic Lys/Arg-rich C-terminal region has recently been suggested [49]. Our data, however, could not confirm this observation since neither p34ct 1-277 (data not shown) nor the p34ct full-length protein were able to bind to single stranded or double stranded 50-mer DNA substrates (Figure 4B). However, it cannot be ruled out that this differs among species. In p34ct the Lys/Arg-rich stretch at the remote C-terminus is disrupted by an insertion and is not as basic as in p34/Tfb4 from S. cerevisiae (Figure 2).

Elucidation of Putative p44 Binding Sites

With p34 being the natural binding partner of the p44 subunit in TFIIH we investigated how the interaction of both proteins could be envisioned. From previous studies it is known that residues 1–242 of human p34 and 321–395 of human p44 (corresponding to residues 1–285 and 375–534 in C. thermophilum, respectively) are sufficient for the formation of a tight p34–p44 complex [19]. In addition, an NMR structure of the C-terminal domain of human p44 revealed the fold of its C4C4 RING domain [20] (Figure 5A). The authors also investigated, which parts of the domain are important for interaction with p34 and observed, that the conservation of the first zinc-binding site is required for the formation of a stable complex. A single coordinating cysteine residue mutated at that position seemed to disrupt the local fold of the domain and abolished binding [20]. In addition, Phe-374, located in helix \( \alpha_2 \) close to the first zinc-binding site, seems to be important for the interaction with p34 since its mutation to a serine disrupts complex formation (Figure 5A). Overall, the authors suggested, that the interaction is more likely to be of hydrophobic than electrostatic character [20].

Interestingly, however, the overall surface charge of the p44h C4C4 domain is highly acidic, with a total of 10 aspartic and glutamic acid residues, all of which are solvent exposed (Figure 5A), and only 2 positively charged residues (Arg and Lys).
With most of the residues around the first zinc-binding site and especially helix a₂, including the phenylalanine, being highly conserved in p44, we thus propose that these two sites contact p34 at a mainly hydrophobic and potentially slightly basic region. Two predominantly hydrophobic patches can be found on the front and left side of the p34 vWA domain (β₃, α₅, α₆, α₇, Figure 1A and 5B). Binding of p44ct C4C4 to the back of the vWA domain (β₆, α₈) is unlikely, as there is a stretch of solvent exposed, highly negatively charged residues. Similarly we exclude binding to the right face of the domain (α₂, α₃), as most residues in helix α₃ and the hydrophobic part of α₂ are not very well conserved (Figures 2 and 5B–C). Hence, the most likely interface is found at the front and/or left side of the vWA domain (β₃, α₅, α₆, α₇), with most of the surface exposed residues in helices α₆ and α₇ being strongly hydrophobic and highly conserved (Figures 2 and 5B–C). Along that line also the lower portion of helix α₅ shows a high degree of conservation and provides a slightly positive contribution to the overall surface potential, which could favorably accommodate part of the negative contribution by p44 C4C4. Taken together, these considerations make the front and left side of the vWA domain (β₃, α₅, α₆, α₇) the most attractive target for binding of the p44 C4C4 domain (Figures 1 and 5B–C).

Implications for Protein-Protein Interactions within TFIIH

The typical vWA domain is known for its capability to mediate protein-protein interactions. It is thus tempting to speculate how both p44 and p34 function within the context of the general transcription factor TFIIH. It has been shown that the p44 vWA domain contacts a C-terminal region in XPD, thereby stimulating its helicase activity [15]. In turn p34 vWA binds strongly to the C-terminal C4C4 RING domain of p44 [19], which is essential for transcriptional activity [32], potentially forming a regulatory triad with XPD and p44. However, only very little is known about the function of p44’s central and p34’s C-terminal zinc finger domain,
respectively. Based on the knowledge that vWA domains are capable to assume multiple protein-protein interfaces [41,42,53,54] it is very likely that the vWA domains of both p44 and p34 are involved in multiple contacts as well, presumably providing stability and serving as a platform and mediator for other subunits, while their zinc-finger motifs could enable them to perform additional roles, which are so far unknown.

Recently, it was shown that a stable 5-subunit minimal complex of Rad3, Tfb1, Tfb2, Tfb4 and Ssl1 (corresponding to the human XPD, p62, p52, p34 and p44 proteins) can only be obtained, when Tfb4 (p34) is included during co-expression [55]. If Tfb4 was omitted, the complex lacked Tfb2 (p52), which suggests that Tfb4 (p34) is required for integration of Tfb2 (p52) into the minimal core complex in yeast. This data supports the notion that p34 is involved in multiple protein-protein interactions within TFIIH. Further analysis will be required to decipher this network of interplay and regulation between TFIIH subunits, especially in the context of transcription and nucleotide excision repair, to gain insight into the molecular requirements for a fully functional TFIIH complex.

Accession Numbers

The data have been deposited and the PDB Code 4PN7 has been assigned.

Supporting Information

Figure S1 Environment of the iodide ions used for SIRAS phasing. The local environment of the 8 iodide ions used for SIRAS phasing is depicted in A – F, with the anomalous electron density for each iodide ion contoured at 5.0 sigma (orange mesh). The iodide ions are depicted as grey spheres, while the
coordinating residues at each site are shown in stick representation. Secondary structure elements of symmetry related molecules are indicated by a prime symbol following the α or β labeling, respectively.

Figure S2 Final model and 2F₀–F₂ electron density map of the p34ct structure. The final model of p34ct is shown in stick representation for a portion of helix α5 (A) and the α7β6 transition (B) with the 2F₀–F₂ electron density map contoured at 1.0 sigma.

Acknowledgments
We kindly thank Ed Hurt, from the University of Heidelberg, for providing C. thermophilum DNA. In addition, the staff of both ID 23.2 and ID 29 of the ESRF in Grenoble as well as BL14.1 of the BESSY II in Berlin deserve many thanks for their excellent beamline support.

Author Contributions
Conceived and designed the experiments: DRS JK AE CK. Performed the experiments: DRS JK AE. Analyzed the data: DRS JK CK. Contributed to the writing of the manuscript: DRS JK CK.

References
1. Chen J, Suter B (2003) Npd1, a structural bridge and a functional link. Cell Cycle 2 (6): 503–506.
2. Busso D, Keriel A, Sandrock B, Peterzunski A, Gileadi O, et al. (2000) Distinct regions of MAT1 regulate cell kine and TIHH transcription activities, J. Biol. Chem. 275 (30):22115–22123. Available: doi:10.1074/jbc.M002570200
3. Egly J, Coin F (2011) A history of TFIIH. Two decades of molecular biology on a pivotal transcription/repair factor. DNA repair. Available: doi:jbrep.2011.04.021
4. Murakami K, Gibbons BJ, Davis RE, Nagai S, Liu X, et al. (2012) TB6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Sae2 helicase after transcription initiation. Proc. Natl. Acad. Sci. U.S.A. 109(13): 4016–4021.
5. Moungues S, Gauthier V, Lagarou A, Bordier C, Mourret A, et al. (2013) ELL, a novel TFIIH partner, is involved in transcription restart after DNA repair. Proc. Natl. Acad. Sci. U.S.A. 110(44): 17927–17932.
6. Coin F, Okemeny V, Moncot V, Grob S, Blattner C, et al. (2000) Nucleotide excision repair driven by the dissociation of CAK from TFIIH. Mol. Cell 19: 337–347.
7. Kraemer KH, Sander M, Bohr VA (2007) New areas of focus at workshop on human diseases involving DNA repair deficiency and premature aging. Med. Aging Dev. 12(8): 229–235.
8. Hashimoto S, Egly JM (2009) Trichothiodystrophy view from the molecular basis of DNA repair/translation factor TFIIH. Hum. Mol. Genet. 18: R224–R230. Available: doi:10.1093/hmg/ddp390
9. Scalella J, Lederer N, Magalhães PM, Poterszman A, Egly J, et al. (2008) Structural basis for group A trichothiodystrophy. Nat. Struct. Mol. Biol. 15(9): 980–984.
10. Aguilar-Fuentes J, Fregoso M, Herrera M, Reynaud E, Braun C, et al. (2008) p52/TTDA overexpression enhances UV-irradiation resistance and suppresses premature ageing. Mutat. Res. 577(1–2): 179–194.
11. Fribourg S, Romier C, Werten S, Gangloff YG, Poterszman A, et al. (2001) Solution structure of the C-terminal domain of TFIIH P44 subunit. Cell 104 (5): 509–520. Available:doi:10.1016/S0092-8674(01)00342-2
12. Fregoso M, Laine J, Aguilar-Fuentes J, Mocquet V, Reynaud E, et al. (2007) XPB/p52 and XPD/p34 subunits of the BTF2/TFIIH transcription factor have homologies with snake venom proteins. J. Biol. Chem. 278(39): 37777–37781.
13. Bedez F, Linard B, Brochet X, Ripp R, Thompson JD, et al. (2013) Functional basis for group A trichothiodystrophy. Nat. Genet. 45(11): e1000253.
14. Coin F, Oksenych V, Egly J (2007) Distinct roles for the XPB/p52 and XPD/p34 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol. Cell 26: 245–256. Available:doi:10.1016/j.molcel.2005.03.007
15. Coin F, Marini JC, Roodzol C, Fribourg S, Pedrini AM, et al. (1998) An XPD allele that causes the basis for group A trichothiodystrophy. Nat. Genet. 19(1): 95–97.
16. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly J (2008) Structural basis for p34 subunit of TFIIH to the writing of the manuscript: DRS JK CK. ex-periments: DRS JK AE. Analyzed the data: DRS JK CK. Contributed to the writing of the manuscript: DRS JK CK.
46. Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/ integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. Mol. Biol. Cell 13(10): 3369–3387.
47. Schultz P, Fribourg S, Poterszman A, Mallolh V, Moras D, et al. (2000) Molecular structure of human TFIIH. Cell. 102(5): 599–607.
48. Gibbons BJ, Brigugio EJ, Azbel M, Murakami K, You NR, et al. (2012) Subunit architecture of general transcription factor TFIIH. Proc. Natl. Acad. Sci. U.S.A. 109(6): 1949–1954.
49. Kainov DE, Sethi LA, Svejstrup JQ, Egly J, Poterszman A (2010) Interacting partners of the Tfb2 subunit from yeast TFIIH. DNA Repair (Amst.) 9(1): 33–39.
50. Marianayagam NJ, Sunle M, Matthews JM (2004) The power of two: protein dimerization in biology. Trends Biochem. Sci. 29(11): 618–625.
51. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372(3): 774–797.
52. Tremeau-Bravard A, Perez C, Egly JM (2001) A role of the C-terminal part of p44 in the promoter escape activity of transcription factor IIIH. J. Biol. Chem 276 (29):27693–27697. Available doi:10.1074/jbc.M102457200.
53. Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, et al. (2012) Complete subunit architecture of the proteasome regulatory particle. Nature 482 (7384): 186–191.
54. Sledz P, Unverdorben P, Beck F, Pfeifer G, Schweitzer A, et al. (2013) Structure of the 20S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. Proc. Natl. Acad. Sci. U.S.A. 110(18): 7264–7269.
55. Takagi Y, Komori H, Chang W, Hudmon A, Erdjument-Bromage H, et al. (2003) Revised subunit structure of yeast transcription factor IIH (TFIIH) and reconciliation with human TFIIH. J. Biol. Chem 278(43): 43997–43900.
56. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5): 1792–1797.
57. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9): 1189-1191.