Structure and DNA binding of the human Rtf1 Plus3 domain.

Rob N. de Jong[1,2,3], Vincent Truffault[1,4], Tammo Diercks[2], Eiso AB[2,5], Mark A. Daniels[2], Rob Kaptein[2] and Gert E. Folkers[2,6].

[1] Both authors contributed equally to this work.

[2] Bijvoet Centre for Biomolecular Research, Utrecht University, Faculty of Chemistry, Dept. NMR Spectroscopy, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[3] Current address: Genmab B.V., Yalelaan 60, 3584 CM Utrecht, The Netherlands

[4] Current address: Max-Planck-Institut für Entwicklungsbiologie, Spemannstr. 35/I, D-72076 Tübingen, Germany.

[5] Current address: Leiden University, Leiden Institute of Chemistry, Einsteinweg 55, 2333 CC Leiden, The Netherlands.

[6] To whom correspondence should be addressed (g.folkers@chem.uu.nl).

Running title: Structure and DNA binding of the hRtf1 Plus3 domain

Character count: 47134

Summary

The yeast Paf1 complex consists of Paf1, Rtf1, Cdc73, Ctr9 and Leo1 and regulates histone H2B ubiquitination, histone H3 methylation, RNA polymerase II carboxy-terminal domain (CTD) Ser2 phosphorylation and RNA 3' end processing. We provide the first structural insight into the Paf1 complex with the NMR structure of the conserved and functionally important Plus3 domain of human Rtf1. A predominantly β-stranded subdomain displays structural similarity to Dicer/Argonaute PAZ domains and to Tudor domains. We further demonstrate that the highly basic Rtf1 Plus3 domain can interact in vitro with single stranded DNA via residues on the rim of the β-sheet, reminiscent of siRNA binding by PAZ domains, but did not detect binding to double stranded DNA or RNA. We discuss the potential role of Rtf1 Plus3 ssDNA binding during transcription elongation.

Keywords:

Histone H2B Ubiquitination / Histone H3 Methylation / Hyperparathyroidism Jaw Tumor Syndrome HPT-JT / Parafibromin / Transcription Elongation Regulation.

Introduction

The ‘histone code’ of acetylated, methylated, phosphorylated and ubiquitinated histones regulate the chromatin association of many different proteins that control transcription efficiency [1, 2]. The yeast Paf1 complex of Paf1, Rtf1, Cdc73, Leo1 and Ctr9 was suggested to play a role in transcription elongation based on physical interactions with RNA polymerase II, genetic and physical interactions with known elongation factors, and recruitment to the open reading frame of transcribed genes [3-7]. Subsequently, the Paf1 complex was implicated in histone H2B ubiquitination and histone H3 methylation (Supplementary Figure 1).

The Rad6/Bre1 complex is responsible for the ubiquitination of histone H2B [8, 9], while Set1, Dot1 and Set2 are the histone methyl transferases that modify the histone H3 lysines K4, K79 and K36, respectively (reviewed in Gerber and Shilatifard, 2003). H2B ubiquitination is a prerequisite for progressive di- and tri-methylation of histone H3K4 and H3K79 by Set1 and Dot1 [10-12]. For full activity, the Bre1/Rad6 complex requires the presence of Rtf1 and Paf1 [10, 11, 13]. Rad6 associates transiently with genes upon transcription activation concomitant with the appearance of RNA polII, but deletion of *RTF1* prohibited specifically the recruitment of Rad6 to open reading frames and Rad6 binding to hyperphosphorylated RNA polymerase II [13]. Rad6 and Paf1 complex recruitment are regulated by phosphorylation at Rad6 residue Ser120 by the Bur1/2 cyclin dependent kinase [14, 15].

The Paf1 complex was also shown to affect mRNA 3’ end processing. The deletion of Paf1 complex components caused a decrease in the mRNA poly(A) tail length and read-through at poly-adenylation signals, but also affected the 3’ end formation of non-polyadenylated RNA transcripts [16-18]. Furthermore, the human Paf1 complex copurified with a component of the Ski complex involved in 3’-5’ mRNA degradation [19]. The pleiotropic effects of interfering in the Paf1 complex pathway suggest, that it serves as a binding platform for transcription

elongation factors while traveling with the RNA polymerase II along the open reading frame, but it probably dissociates before the polyadenylation sequence [20-22].

The conservation of Paf1 complex function in higher eukaryotes has been confirmed in human cells, Drosophila and Arabidopsis. The clinical relevance of Paf1 complex function became apparent with the implication of the *HPRT2* gene encoding human Cdc73/parafibromin in the hyperparathyroidism-jaw tumor (HPT-JT) syndrome [23]. Productive transcription and H2B ubiquitination of a RARβ2 chromatin template in vitro required the human Paf1 complex [24]. Interestingly, human Rtf1 does not seem to be a member of the preassembled Paf1 core complex [25, 26], although binding to the other Paf1 complex components and RNA polymerase II was detected [19]. Drosophila Rtf1 does also not stably associate with Paf1 and Cdc73, but H3K4 trimethylation is again disrupted upon depletion of Rtf1 [27, 28].

Until now, structural information on the Paf1 complex was missing completely. We provide the first structural insight into the Paf1 complex with the elucidation of the solution structure of the human Rtf1 Plus-3 domain, the most conserved Rtf1 region. It forms a compact globular domain with a novel fold, but a predominantly β-stranded subdomain displays similarities to the Argonaute and Dicer siRNA binding PAZ domains and to the Tudor/Chromo/PWWP domain family found in chromatin associating proteins. It also reveals a previously unnoted relationship with transcription elongation factors. We demonstrate that the Rtf1 Plus3 domain can interact with single stranded DNA via amino acids at the rim of the β-sheet, partly resembling the Argonaute/siRNA interaction. We discuss implications of the Rtf1 Plus3 structure and DNA binding for its role in transcription regulation.

Results

In an ongoing structural genomics effort, we screen conserved protein domains without detectable sequence homology to structures deposited in the Protein Data Bank (PDB), that are present in human proteins involved in transcription for their suitability to structure determination using Nuclear Magnetic Resonance (NMR) by recording fingerprint [^{15}N,^{1}H]-HSQC spectra (Supplementary Figure 2) [29]. One such domain, the Plus3 domain, spans the most conserved regions of the Saccharomyces cerevisiae protein Rtf1 and is surrounded by regions of low complexity and coiled-coil propensity. A representative set of aligned sequences (Fig. 1A) shows only a limited number of highly conserved amino acids, among which are three positively charged residues that gave the Plus3 domain its name. The N-terminus of the human Rtf1 (hRTF1) protein has not yet been defined experimentally, but was numbered according to the EnsEMBL annotation that indicates it precedes the largest deposited human mRNA by 9 amino acids, and a mouse mRNA by one amino acid. The structure of the human Rtf1 Plus3 domain was solved by NMR spectroscopy following our standard structural genomics protocol [30], and deposited in the PDB under accession code 2BZE.

The Plus3 domain structure consists of six α-helices intervened by a sequence of six β-strands in a mixed α/β topology (Fig. 1). β-strands 1, 2, 5 and 6 compose a four-stranded anti-parallel β-sheet with a β-hairpin insertion formed by strands 3 and 4. The NMR ensemble (Fig. 1C) is well defined except in the poorly conserved loop between β-strands 1 and 2: the mean global backbone rmsd is 0.93Å including and 0.62Å excluding the loop (Table I). The N-terminal helices α1-3 and C-terminal helix α6 pack together to form an alpha subdomain, while the β-strands and the small 3_{10} helix α4 form a beta subdomain. The two subdomains pack together to form a compact, globular protein (Fig. 2).

Conserved residues playing a structural role

Remarkably, all three fully conserved amino acids R367, R389 and S444 cluster at the interface between the alpha and beta subdomains, together with the highly conserved Y401, N445, F448 and E452 (Fig. 2C). R389 in β1 projects outwards from the convex side of the β-sheet, exposing only the guanidinium head group to solution. The guanidinium group forms a salt bridge with E452 in α5 and stacks on the aromatic ring of Y401 in β2. The R389 backbone carbonyl forms a hydrogen bond with the amide proton of S444, while the aliphatic moiety of the R389 side chain packs against F448. In hRTF1, the F448 backbone carbonyl and amide proton are hydrogen bonded to the poorly conserved N361 sidechain in α1 (not shown).

A highly curved, four-stranded anti-parallel β-sheet is formed by strands β1, β2, β5 and β6 (Fig. 3A). Residues 440 to 443 form a short 3_{10} helix that crosses the β-sheet. The curvature of the β-sheet is stabilized by a hydrophobic core consisting of the highly conserved V388 (β1), I406 (β2) and L427 (β5) and the hydrophobically conserved residues I390 (β1), L429 (β5) and V443 (α4). A short β-hairpin, consisting of the three-residue strands β3 and β4, protrudes from strands β2 and β5 and contains the highly conserved Y416, which forms two hydrogen bonds with the K425 carbonyl and L440 amide groups (Fig. 3A).

While the four N-terminal amino acids in our expression construct (V354-P357) proved essential to prevent protein aggregation, we only obtained highly expressed, soluble protein after extending the Plus3 domain defined in the SMART database [31] C-terminally by 23 amino acids. This C-terminal region forms helix α6 that participates in a four-helical cluster with helices α1-3, stabilized by hydrophobic interactions between the conserved residues L365 in the α1-α2-loop, F379 and V383 in α3 and the hydrophobically conserved amino acids W373 in α2 and I480 in α6 (Fig. 3B). In addition, I473 in helix α6 packs against the aliphatic moieties of E359, R362 and V363 in α1, and a non-conserved salt bridge is formed in hRtf1 between R362 and E477

(α6). Since the hydrophobicity of the residues interacting with α6 is conserved throughout the Plus3 family, we expect that the C-terminal helix in human Rtf1 is an integral part of other Plus3 domains as well.

The point mutations V274D and M289K in the Saccharomyces cerevisiae Plus3 domain cause temperature sensitive, conditional 6-azauracil and Spt$^-$ phenotypes [17]. The hRtf1 residue I390 in β1, which corresponds to V274 in yeast Rtf1, participates in the hydrophobic packing of the β-sheet core (Fig. 3C). V403 in β2 (yeast M289) is surrounded by aromatic residues from α5 and the β-sheet (Fig. 3D). Both I390 and V403 are completely buried, suggesting that the observed phenotypes are caused by destabilization of the β-sheet as a result from the charges introduced in the hydrophobic cores. Interestingly, both mutations neighbor highly conserved, surface exposed arginine residues, that could therefore be functionally important (Fig. 3).

Structural homology suggests a nucleic acid binding function

To understand the function of the Rtf1 Plus3 domain, we searched for structural homologues using DALI [32]. Using the entire Plus3 domain, we only detected significant homology for the beta subdomain. Separate searches using only the beta subdomain returned over 40 significant hits, while the alpha subdomain yielded none. The strongly curved β-sheet is reminiscent of the SH3 domain fold, but since the essential SH3 ligand binding loop between strands β1 and β2 is missing in Rtf1, we do not consider this to be functionally relevant. More interesting are the homologies detected with a wide variety of nucleic acid binding proteins, including both Dicer and Argonaute PAZ domains (closest stuctural homologue, Z-score 3.4, rmsd 3.5 over 71 aligned residues), the Tudor domains of Sm and SMN proteins (not discussed), and bacterial transcription elongation factors.

Figure 4A shows the superposition of the Rtf1 Plus3 domain with the Drosophila melanogaster Ago2 PAZ domain (dark blue) complexed to a RNA pentanucleotide [33]. The 3’

terminal nucleotides (gold) are bound by residues in the β-sheet and a β/β/α insertion between strands β2 and β5 that forms a hydrophobic pocket recognizing the unpaired 3' terminal siRNA bases. This crucial insertion is absent in Rtf1, where the β3/β4 hairpin instead turns in opposite direction. The structural homology includes helices α3 and α4 and the β-sheet in Rtf1, but sequence homology is low and some fully conserved residues important for Ago2 siRNA binding are absent in Rtf1.

Since Rtf1 was shown to have diverse functions during the transcription elongation phase and in RNA processing, we were interested by the Plus3 structural similarities with the bacterial transcription elongation factor NusG (Fig. 4B) [34]. NusG displays functional similarity to the yeast Spt5 protein and sequence similarity in the β-strands that correspond to Rtf1 β1 and β2. This KOW motif (colored gold in Fig. 4B) has been implicated in nucleic acid binding and shows considerable sequence conservation at structurally important positions with the Rtf1 residues at the equivalent positions (Fig. 4C). In conclusion, the structural homology with a number of nucleic acid binding proteins, the generally high positive charge of Plus3 domains and the fact that solvent exposed, basic residues are amongst the most conserved residues suggest that the Rtf1 Plus3 domain might be a nucleic acid binding domain.

The Rtf1 Plus3 domain can bind single stranded DNA

Since structurally homologous proteins also have functions other than nucleic acid binding, we tested a number of potential ligands by NMR ^{15}N-HSQC titration experiments. Homology with chromodomains suggested putative histone H3 tail binding, but we did not detect Rtf1 Plus3 binding to the 20 N-terminal residues of histone H3 either un- or di-methylated at K4 (not shown). Rtf1 was also implicated in the recruitment of Rad6 from promoters to open reading frames during transcription elongation [13], but we did not observe binding of either of the closest human Rad6 homologues E2-A or E2-B to the human Rtf1 Plus3 domain (not shown). Addition of double stranded DNA did not result in changes in the ^{15}N-HSQC spectrum either. We

therefore tested Plus3 domain DNA binding in electro-mobility shift assays (EMSA) using a variety of different DNA structures.

We incubated increasing amounts of purified, His-tagged Rtf1 Plus3 with double stranded DNA (dsDNA), single stranded DNA (ssDNA) and a DNA probe representing a transcription bubble with two base-paired stems separated by 10 nucleotides of unpaired ss DNA (Fig. 5A). DNA binding to dsDNA was barely detectable, but ssDNA was bound as judged by the disappearance of unbound DNA and the appearance of a protein/DNA complex band. The addition of Plus3 to the bubble DNA resulted in the formation of a protein/DNA complex with an apparent K_D of approximately 3 μM. Protein aggregation at 6 μM Plus3 caused significant loss of radioactvity for the bound ssDNA and bubble probes, but not for the poorly bound dsDNA.

To demonstrate that the mobility shift was caused by Rtf1/DNA complex formation, we first incubated purified Glutathione S-transferase (GST) tagged Plus3 domain with the bubble probe (Fig. 5B). The band corresponding to the His-Plus3/DNA complex was absent; instead, the higher molecular weight of the GST-tag resulted in the formation of a less mobile protein/DNA complex. The apparently stabilized binding for GST-Plus3 compared to His-Plus3 might be caused by the expected dimerization of the GST tag. As an alternative test to verify the observed protein/DNA complex, we incubated the bubble DNA and Rtf1 with an α-His Horseradish Peroxidase (HRP) conjugate that binds the His-tag (Fig. 5C). As expected, this resulted in the formation of a lower mobility band representing a DNA/Plus3/HRP complex; it also partially inhibited DNA binding by Rtf1 Plus3.

We confirmed substrate specificity by showing efficient competition of His RTF1 Plus3 bound to a bubble substrate using probes that contain conformationally restricted ssDNA sequences (bubble and hairpin20) (Figure 5D). Slightly less efficient competition was observed for ssDNA (BCTC) and a splayed arm substrate (fork). No competition was detectable for a 20bp dsDNA substrate, which indicates that the Plus3 domain has at least 25 fold weaker affinity for

this substrate. We did not observe any evidence for sequence specific DNA recognition, as various equal-length oligonucleotides showed comparable efficiency in competing Rtf1 Plus3 binding from the ssDNA probe BCTC (Supplementary Figure 3) or bubble substrate (data not shown). We conclude that the Rtf1 Plus3 domain can bind to DNA substrates that are at least partially single stranded with significant affinity.

The Rtf1 Plus3 beta sheet participates in ssDNA binding

To identify Rtf1 amino acids involved in DNA binding, we recorded ^{15}N-HSQC spectra of Rtf1 Plus3 in the presence of increasing amounts of bubble DNA. In accordance with the EMSA results, we observed chemical shift changes for a number of ^{15}N bound protons (Supplementary Fig. 4), summarized in Figure 5E. Poor solubility of the DNA/protein complex required relatively stringent salt conditions (>400 mM) and limited the maximum protein/DNA concentration to approximately 100 μM, which prevented us from observing some residues with very low signal intensity in the flexible β1/β2 loop. Residues sensitive to the addition of DNA include the backbone amides of R402, R430 (buried), Q435, R436 and the Hε proton of the R436 guanidium group, all on the concave side of the β-sheet (Fig. 6A). The side chains of R430 and Q435, located on the convex side of the β-sheet, were not affected by addition of DNA, suggesting that it is the concave β-sheet side exlusively that contacts DNA.

To corroborate the Rtf1 Plus3 surface highlighted by HSQC shift changes as the DNA binding site, we performed site specific mutagenesis. Rtf1 Plus3 mutants were incubated with the bubble DNA probe and displayed varying levels of DNA binding (Fig. 6C). Mutation of the poorly conserved side chains of E411 or Q435 had little effect, while mutation of R402, R430 or R436 largely or completely inhibited DNA binding. Consistent with the NMR data, mutation of the side chain of Q435 on the convex side of the β-sheet did not affect DNA binding, while mutation of the R436 side chain on the concave side blocked binding to DNA. The convex side

R402E mutation also inibihited DNA binding, but we cannot exclude a structural disruption caused by charge repulsion between the E405 and R402E sidechains that are in close proximity.

The Plus3 domain displayed homology to RNA binding proteins and recent data suggested a function for Rtf1 in the regulation of RNA processing[17, 18], so we also tested the RNA binding abilities of Rtf1 Plus3 domain. Neither in EMSA (Supplementary Fig. 5), nor in HSQC titration (not shown) did we detect RNA binding by Rtf1. In summary, these results suggest that the capacity to bind single stranded DNA is at least one function of the Rtf1 Plus3 domain.

Discussion

Plus3 is structurally and functionally related to other nucleic acid binding domains

The globular Plus3 domain as a whole adopts a new fold, but the subdomain containing the highly curved β-sheet followed by a 3_{10} helix displays structural similarity to a wide variety of proteins. The β-sheet is reminiscent of SH3 domains, but lacks the hallmark ligand binding loop between strands β1 and β2. Closer related are the Tudor domains, which are members of a domain family including PWWP and chromodomains that can interact with nucleic acids, chromatin and histone proteins. The closest structural homologues to Rtf1 Plus3 are the siRNA binding PAZ domains of Argonaute2 and its Dicer family members.

A comparison of the DNA binding results (Fig. 6) with the superposition of these proteins onto the Plus3 domain (Fig. 3) suggests structural and functional similarities. In the Argonaute and Dicer PAZ domains, β-sheet residues close to Rtf1 R430 in β5, and Q435 and R436 in β6, are involved in nucleic acid binding (Fig. 3). A similar β-sheet surface in a Tudor domain of the p53 binding protein (53BP1) was also implicated in DNA binding [35]. Because of poor signal intensity due to low Rtf1/DNA complex solubility, we could not determine whether the Rtf1 β1/β2 loop participates in DNA binding, as was the case for 53BP1. The DNMT3B PWWP domain, superimposing on Plus3 with a Z-score of 2.1, is related to Tudor domains and involved in chromatin targeting. A S282P mutation in this domain causes ICF syndrome and disrupts the chromatin targeting of DNMT3B [36]. DNMT3B S282 is located at the position corresponding to R436 in Rtf1, which is among the residues most sensitive to addition of DNA and, mutation of which obliterated DNA binding in EMSA (Fig. 6). Strands β1 and β2 align with the nucleic acid binding KOW motif in the bacterial transcription elongation factor NusG and in

the eukaryotic elongation factor Spt5 (Fig. 3). Mutation of R402 in β2 blocked Plus3 DNA binding (Fig. 6).

The role of the Rtf1 Plus3 domain in transcription

The presence of the Paf1 complex in the mRNA coding regions of actively transcribed genes has been firmly established, and deletion of yeast *RTF1* disrupts the recruitment of Paf1 complex components, Set1, Set2 and the Chd1 helicase to transcribed genes as well as the transfer of Rad6 from promoters to coding regions [7, 13, 16, 37-40]. Since the Paf1 complex binds RNA pol II, recruitment of its binding partners to coding regions was suggested to take place via a Paf1 interaction "platform" that travels along with RNA pol II during transcription up to, but not beyond the polyadenylation site [20, 21]. The human core Paf1 complex was shown to interact with RNA pol II as well, apparently in the absence of the human Rtf1 homologue. This indicates that its interaction with RNA polymerase II does not depend on Rtf1 [24], although yeast *RTF1* deletion also diminished RNA pol II binding of Paf1. This raises questions on the role of human Rtf1 within the Paf1 complex, but one could imagine for example a more gene-specific role, a transient contribution or a regulated association of Rtf1.

Recently, Warner et al. reported that internal Rtf1 deletions that would undoubtedly disrupt the Plus3 domain structure, blocked the association of Rtf1 with transcribed open reading frames as well as Ctr9 co-recruitment to these genes [41]. Furthermore, these deletions in the Plus3 domain compromised telomeric silencing and histone H3 K4 trimethylation [41] and caused phenotypes associated with transcription elongation defects. The chromatin dependent association of the Plus3 domain with transcriptionally active genes in the context of the intact protein is consistent with the nucleic acid affinity and structural preferences we observed for the isolated Rtf1 Plus3 domain. We did not detect binding of the Plus3 domain to either histone H3 termini or RNA, other substrates that could recruit Rtf1 to chromatin.

The ability of the Rtf1 Plus3 domain to interact with single stranded DNA could suggest a role for Rtf1 in the structural organization of the elongating transcription bubble. Transcription elongation through chromatin is stimulated by H2B ubiquitination and Spt16/Pob3 (FACT in humans), which can cause displacement of a H2A/H2B dimer from the histone octamer [24, 42]. The vital role for yeast Rtf1 in H2B ubiquitination and the Rtf1-dependent promoter escape of Rad6 could suggest a model in which the Plus3 domain is located in front of the polymerase to facilitate proper positioning of the H2B ubiquitinating complex Rad6/Bre1. During processive elongation, Rad6/Bre1 could then transfer ubiquitin to H2B while scanning the incoming nucleosomes with the help of a Rtf1 Plus3 mediated DNA interaction. Whether and to what extent the observed Rtf1 affinity for nucleic acids contributes to chromatin regulation in vivo, remains to be determined.

Experimental Procedures

Protein expression and purification

A DNA fragment encoding amino acids 354 to 485 of human Rtf1 (Swiss-Prot entry Q92541 aa 345-476) was cloned into the His-tag expression vector pET15B (Novagen). Cloning, recombinant protein expression and isotopic labeling in the *Escherichia coli* strain BL21 DE3 RIL (Stratagene), lysis of bacteria and protein purification were performed essentially as described in our standard stuctural genomics protocol [29], with the following modifications. The protein was purified over Ni-MC POROS (PerSeptive Biosystems), buffer exchanged to 10 mM Tris/Bis-Tris pH7.0,1 180 mM NaCl, diluted to 100 mM NaCl, and purified over a HS-POROS column (PerSeptive Biosystems) in 10 mM Tris/Bis-Tris pH7.0 with elution at 200 mM. The protein was buffer exchanged to 150 mM NaCl/50 mM Na_2HPO_4 / NaH_2PO_4 at pH 6.0, and concentrated by ultrafiltration to approximately 1.5 mM using 3 kDa cutoff Centricon (Amicon) spinconcentrators. For NMR samples, 10 % D_2O and 0.02% NaN_3 were added. Further details are provided as Supplementary Information.

Rtf1 Plus3 point mutants were produced using a megaprimer mutagenesis strategy amplifying the human Rtf1 aa 354-485 and cloned into pLICHIS, a pET15B derived expression vector suitable for enzyme-free cloning [43]. Mutant proteins from 100 ml L-Broth extracts were purified only on the MC-POROS column, concentrated and washed with 150 mM NaCl/50 mM Na_2HPO_4/NaH_2HPO_4 pH 6.0. Rtf1 Plus3 domain tagged at the N-terminus by a consecutive His and GST tag was expressed from the homemade pLICHISGST expression vector [43]and purified like the point mutants.

DNA binding

Electromobility shift assays (EMSA) were perfomed in an EMSA buffer (EB) containing 10 mM Tris pH 7.0, 1mM DTT, 100 μg/ml BSA and 10% glycerol; 75 mM NaCl was present during the experiment described in Fig. 6. Indicated amounts of purified His-Rtf1 Plus3 or His-GST-Rtf1 Plus3 were diluted in EB and incubated for 30 minutes on ice with appr. 2.5 fmol ^{32}P radiolabeled DNA probes in 20 μl end volume of EB. Free DNA and protein/DNA complexes were separated on 8% polyacrylamide gels (0.25xTBE, acrylamide:bis acrylamide 37.5:1) at 6 V/cm. Where indicated, 0.5 ul α-His HRP conjugate (4 mg/ml His-PROBE HRP, Pierce, Rockford, USA) was added. To produce DNA probes with different structures, we used 30-mer GGGCGGCGGG-$(T)_{10}$-GGCGGGGCGG (BGTG), CCGCCCCGCC-A_{10}-CCCGCCGCCC (BCAC) or CCGCCCCGCC-T_{10}-CCCGCCGCCC (BCTC) oligonucleotides. The double stranded probe consisted of BCAC hybridized to BGTG, the single stranded probe was BGTG and the bubble consisted of BGTG hybridized to BCTC. For competition experiments, the indicated amounts of non-labeled competitor probes [44] were added after complex formation between 3μM His-RTF1 Plus3 and the indicated amount of radiolabeled bubble substrate. Thirty minutes after addition of the competitor, the reaction mixture was loaded on gel.

NMR spectroscopy

All NMR measurements on [U-^{15}N] and [U-^{13}C,^{15}N] labeled hUSP15 were performed at 298 K on BRUKER Avance spectrometers operating at 700 and 900 MHz magnetic field strength. The spectra for resonance assignment and structure determination were recorded using a standard set of NMR experiments [45] essentially as described previously [30]. DNA titrations were performed with 100 μM His-Rtf1 Plus3 in a buffer containing 20 mM Bis-Tris / Tris at pH 6.0 and 400 mM NaCl. Where indicated, bubble DNA in the same buffer with pH calibrated after

dissolving was added to a final concentration of 100 μM. Further details are available in the Supplementary Information.

NOE analysis and structure calculations

Automatic NOE assignment and structure calculations were performed using the CANDID module of the program CYANA [46] essentially as described previously [30]. The assignment completeness for the final run was 99% for non-labile protons excluding the His-tag. The final set of NOE distance restraints determined by CANDID, together with restraints for 42 H-bonds and dihedral restraints for 47 residues from TALOS, were used in a water refinement run using CNS [47]. The structure was validated using WHATCHECK [48] and PROCHECK [49]. Structural coordinates of the 25 lowest energy structures were deposited in the Protein Data Bank under access code 2BZE. NMR chemical shift assignments were deposited in the Biological Magnetic Resonance Bank (BMRB) database under entry 7351. Further details are available in the Supplementary Information.

Acknowledgements

The authors would like to thank Pim Pijnappel for H3 peptides, Rolf Boelens, Alexandre Bonvin and Marc Timmers for stimulating discussions, and H. van Aken, M. Hilbers and J. van der Zwan for technical support. This work was supported by the EU SPINE (Structural Proteomics in Europe) grant QLG2-CT-2002-00988.

References

1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.
2. Kouzarides, T. (2002). Histone methylation in transcriptional control. Curr Opin Genet Dev *12*, 198-209.
3. Costa, P.J., and Arndt, K.M. (2000). Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation. Genetics *156*, 535-547.
4. Porter, S.E., Washburn, T.M., Chang, M., and Jaehning, J.A. (2002). The yeast pafl-rNA polymerase II complex is required for full expression of a subset of cell cycle-regulated genes. Eukaryot Cell *1*, 830-842.
5. Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. Mol Cell Biol *22*, 6979-6992.
6. Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G.A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. Embo J *21*, 1764-1774.
7. Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol Cell *9*, 799-809.
8. Hwang, W.W., Venkatasubrahmanyam, S., Ianculescu, A.G., Tong, A., Boone, C., and Madhani, H.D. (2003). A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell *11*, 261-266.
9. Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., Johnston, M., and Shilatifard, A. (2003). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell *11*, 267-274.
10. Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem *278*, 34739-34742.
11. Ng, H.H., Dole, S., and Struhl, K. (2003). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. J Biol Chem *278*, 33625-33628.
12. Shahbazian, M.D., Zhang, K., and Grunstein, M. (2005). Histone H2B Ubiquitylation Controls Processive Methylation but Not Monomethylation by Dot1 and Set1. Mol Cell *19*, 271-277.
13. Xiao, T., Kao, C.F., Krogan, N.J., Sun, Z.W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol *25*, 637-651.
14. Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2005). The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. Mol Cell *20*, 589-599.
15. Laribee, R.N., Krogan, N.J., Xiao, T., Shibata, Y., Hughes, T.R., Greenblatt, J.F., and Strahl, B.D. (2005). BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. Curr Biol *15*, 1487-1493.

16. Mueller, C.L., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2004). The Paf1 complex has functions independent of actively transcribing RNA polymerase II. Mol Cell *14*, 447-456.
17. Sheldon, K.E., Mauger, D.M., and Arndt, K.M. (2005). A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. Mol Cell *20*, 225-236.
18. Penheiter, K.L., Washburn, T.M., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. Mol Cell *20*, 213-223.
19. Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev *19*, 1668-1673.
20. Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. Embo J *23*, 354-364.
21. Gerber, M., and Shilatifard, A. (2003). Transcriptional elongation by RNA polymerase II and histone methylation. J Biol Chem *278*, 26303-26306.
22. Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell *113*, 429-432.
23. Carpten, J.D., Robbins, C.M., Villablanca, A., Forsberg, L., Presciuttini, S., Bailey-Wilson, J., Simonds, W.F., Gillanders, E.M., Kennedy, A.M., Chen, J.D., Agarwal, S.K., Sood, R., Jones, M.P., Moses, T.Y., Haven, C., Petillo, D., Leotlela, P.D., Harding, B., Cameron, D., Pannett, A.A., Hoog, A., Heath, H., 3rd, James-Newton, L.A., Robinson, B., Zarbo, R.J., Cavaco, B.M., Wassif, W., Perrier, N.D., Rosen, I.B., Kristoffersson, U., Turnpenny, P.D., Farnebo, L.O., Besser, G.M., Jackson, C.E., Morreau, H., Trent, J.M., Thakker, R.V., Marx, S.J., Teh, B.T., Larsson, C., and Hobbs, M.R. (2002). HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. Nat Genet *32*, 676-680.
24. Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.
25. Yart, A., Gstaiger, M., Wirbelauer, C., Pecnik, M., Anastasiou, D., Hess, D., and Krek, W. (2005). The HRPT2 tumor suppressor gene product parafibromin associates with human PAF1 and RNA polymerase II. Mol Cell Biol *25*, 5052-5060.
26. Rozenblatt-Rosen, O., Hughes, C.M., Nannepaga, S.J., Shanmugam, K.S., Copeland, T.D., Guszczynski, T., Resau, J.H., and Meyerson, M. (2005). The parafibromin tumor suppressor protein is part of a human Paf1 complex. Mol Cell Biol *25*, 612-620.
27. Tenney, K., Gerber, M., Ilvarsonn, A., Schneider, J., Gause, M., Dorsett, D., Eissenberg, J.C., and Shilatifard, A. (2006). Drosophila Rtf1 functions in histone methylation, gene expression, and Notch signaling. Proc Natl Acad Sci U S A.
28. Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Mol Cell Biol *26*, 250-260.
29. Folkers, G.E., van Buuren, B.N., and Kaptein, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. J Struct Funct Genomics *5*, 119-131.
30. de Jong, R.N., Ab, E., Diercks, T., Truffault, V., Daniels, M., Kaptein, R., and Folkers, G.E. (2006). Solution Structure of the Human Ubiquitin-specific Protease 15 DUSP Domain. J Biol Chem *281*, 5026-5031.

31. Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P., and Bork, P. (2004). SMART 4.0: towards genomic data integration. Nucleic Acids Res *32 Database issue*, D142-144.
32. Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.
33. Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol *11*, 576-577.
34. Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R., and Wahl, M.C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. Embo J *21*, 4641-4653.
35. Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quemeneur, E., Guerois, R., Callebaut, I., Gilquin, B., and Zinn-Justin, S. (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure *12*, 1551-1562.
36. Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem *279*, 25447-25454.
37. Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. Embo J *22*, 1846-1856.
38. Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., and Greenblatt, J. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol *23*, 4207-4218.
39. Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.
40. Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell *11*, 721-729.
41. Warner, M.H., Roinick, K.L., and Arndt, K.M. (2007). Rtf1 is a multifunctional component of the paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Mol Cell Biol *27*, 6103-6115.
42. Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.
43. de Jong, R.N., Daniels, M.A., Kaptein, R., and Folkers, G.E. (2006). Enzyme Free Cloning for high throughput gene cloning and expression. J Struct Funct Genomics *7*, 109-118.
44. Singh, S., Folkers, G.E., Bonvin, A.M., Boelens, R., Wechselberger, R., Niztayev, A., and Kaptein, R. (2002). Solution structure and DNA-binding properties of the C-terminal domain of UvrC from E.coli. Embo J *21*, 6257-6266.
45. Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy *34*, 93-158.
46. Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol *319*, 209-227.

47. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.
48. Hooft, R.W., Vriend, G., Sander, C., and Abola, E.E. (1996). Errors in protein structures. Nature *381*, 272.
49. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR *8*, 477-486.
50. DeLano, W.L. (2002). The PyMol Molecular Graphics System, DeLano Scientific: San Carlos, CA, USA.
51. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res *33*, W299-302.

Legends

Figure 1. The structural organization of the Rtf1 Plus3 domain. Color coding: red: α-helices; cyan: β-strands; green: 3_{10} helix. (A) Sequence alignment of ten proteins representative of the Plus3 domain sequence space. Rtf1_human (Swiss-Prot Q92541; top), Rtf1_yeast (S. cerevisiae P53064; bottom) and Swiss-Prot accession numbers for the other entries are indicated. Asterisks indicate the three conserved positively charged amino acids after which the Plus3 domain was named. Species from top to bottom: Petroselinum c., Neurospora c., Arabidopsis t., Schizosaccharomyces p., Oryza s., Drosophila m., Caenorhabditis e., Drosophila m.. (B) Topology diagram of the novel Plus3 domain fold. (C) Superimposed NMR ensemble of the 25 lowest energy structures. Structural figures were created using PyMol [50].

Figure 2. The solution structure of the Rtf1 Plus3 domain. (A) Two orthogonal ribbon presentations, rotated by 90° around the y-axis. (B) Stereoview of the backbone trace with every tenth Cα atom marked by a black sphere. Numbers indicate amino acid positions. (C) The fully conserved residues R367, R389 and S444 cluster on the α/β subdomain interface (same orientation as B).

Figure 3. Conserved residues in the Rtf1 Plus3 domain. For clarity, most hydrogens were omitted from the figures. (A) Core of the beta subdomain. The Y416 OH group in the β-hairpin hydrogen-bonds to the L440 backbone amide proton and K425 backbone carbonyl group. (B) The hydrophobic core of the four-helical bundle, color-coded as in Fig. 1. Side chains colored according to the CPK scheme with carbon: grey; hydrogen: white; oxygen: red; nitrogen: blue. (C) The S. cerevisiae mutations V274D (Hs I390, in green) and (D) M289K (Hs V403, in green), which cause temperature sensitive phenotypes, disrupt packing of the β-sheet core and β-sheet/α5

interface, respectively. Other amino acids are color coded by conservation scores, calculated using ConSurf [51] with the alignment in Fig. 1A as input, ranging from white (not conserved) to red (highly conserved).

Figure 4. The Plus3 beta subdomain displays structural similarity to nucleic acid binding proteins. The Rtf1 Plus3 domain is color coded as in Fig. 2A, while α1, α2 and α6 have been omitted for clarity. Structural homologues retrieved and aligned using DALI [32] are displayed in dark blue, with discussed features highlighted in gold. (A) Overlay with the Argonaute PAZ domain (PDB entry 1T2R) complexed to a pentanucleotide RNA, of which two are visible in the structure (gold) [33]. (B) Overlay with the bacterial transcription elongation factor NusG (1M1G), with the KOW motif highlighted in gold [34]. (C) Sequence similarity between the KOW motif in transcription elongation factors and the Rtf1 Plus3 domain. Spt5 from Saccharomyces cerevisiae (Spt5_Sc) and Homo sapiens (Spt5_Hs) contain multiple KOW motifs numbered from N to C terminus. RFAH_Ec: Escherichia coli transcriptional activator rfaH. NusG_Aa: Aquifex aeolicus transcription antitermination protein NusG. RL24_Aa: Aquifex aeolicus 50S ribosomal protein L24. Yeast and human Rtf1 are indicated at the bottom. Asterisks indicate the position of thermosensitivity causing mutantions with elongation phenotypes (Fig. 2 C,D).

Figure 5. The Rtf1 Plus3 domain binds single stranded DNA. (A) Electromobility shift assays (EMSA) with a double stranded, single stranded and bubble DNA consisting of two ds stems separated by two opposing unpaired 10 nucleotide single strands. From left to right, stepwise two-fold increasing concentrations up to 6 μM of Rtf1 Plus3 were incubated with DNA; "-" indicates no protein. Closed arrowheads indicate DNA/protein complex. (B) EMSA with the bubble DNA substrate as in Fig. 5A, but with increasing amounts of His-GST-Rtf1 Plus3 (open arrowhead) and a His-Rtf1 Plus3 control (closed arrowhead). (C) EMSA with bubble DNA and 3 or 6 μM

His-Rtf1 Plus3 incubated with an α-His HRP conjugate that binds to the His-tag. The closed arrowhead indicates DNA/His-Rtf1 Plus3 complex, the open arrowhead the ternary DNA/protein/α-His HRP conjugate complex. (D) EMSA with bubble DNA and 3μM His-Rtf1 Plus3 complex in the absence (+) or presence of 0.008, 0.04, 0.2, 1μM of the indicated oligonucleotides as competitor: ss: BCTC; fork: splayed arm with 10bp dsDNA and 10 unpaired nucleotides (dT10); hairpin: stem-loop substrate with 20 unpaired bases (dT20); ds: 20 bp dsDNA (E) Compound chemical shift deviations in ppm for ^{15}N-bound protons (deviation = $[(6.5*\delta HN)^2 + (\delta N)^2]^{1/2}$) upon addition of 100 μM bubble DNA. The dotted line indicates the minimal threshold chosen for the color scale in Fig. 6. Secondary structure elements are indicated below the figure, colored as in Fig. 1C. Missing bars indicate unobservable protons for which the remaining signals were too weak at this protein concentration (100 μM) or disappeared upon DNA addition.

Figure 6. Mapping of Rtf1 Plus3 amino acids implicated in ssDNA binding. (A) Ribbon presentation of Plus3 domain, oriented as in Fig. 3. Chemical shift perturbation is color coded from white (below treshold of 0.12 ppm) to red (maximal). Grey indicates missing or disappearing signals during titration. (B) SDS PAGE of purified Plus3 mutants. Approximately 1μg of single column purified His-tagged mutant proteins were separated by SDS-PAGE and stained by Coomassie Brilliant Blue. wt: wildtype Rtf1 Plus3. (C) EMSA of 0.8 μM Rtf1 Plus3 proteins on the bubble ss/ds DNA probe as described in Fig. 5. (-) indicates extract purified from empty vector. A filled arrowhead indicates the position of the Plus3/DNA complex.

Table 1. Structural statistics of the human Rtf1 Plus3 domain (aa 354-485)	
NOE based distance restraints	**number of restraints:**
Intraresidual	927
Sequential	941
Medium range (2 to 4)	617
Long range (5 or more)	1080
Total	3565
Other restraints	
phi + psi dihedral restraints	92
H-Bond restraints	42
Mean global RMSD (Å):	
Overall: amino acids 354-485	0.93±0.20 (bb); 1.58±0.24 (heavy)
Structured: amino acids 354-391;400-485	0.62±0.11 (bb); 1.28±0.16 (heavy)
WHATCHECK	**Structure Z-scores:**
1st generation packing quality	-0.375
2nd generation packing quality	-0.964
Ramachandran plot appearance	-3.781
chi-1/chi-2 rotamer normality	-2.848
Backbone conformation	-1.692
Number of bumps per 100 residues	8.36
PROCHECK	**Ramachandran plot (%):**
Most favoured regions	90.3
Allowed regions	9.2
Generously allowed regions	0.3
Disallowed regions	0.2

Table 1. Structural statistics of the human Rtf1 Plus3 domain. Unambiguous restraints are listed by range: intra-residual (within the same residue), sequential (between neighboring residues), medium-range (between residues 2 to 4 positions apart) or long-range (5 or more residues apart). The average RMSD of superimposed backbone (bb) and heavy atoms (heavy) was calculated for

the indicated amino acids. Structure Z-scores: WHATCHECK analysis of the deposited ensemble over the whole protein (residues 354-485). The first deposited structure in the ensemble is the structure closest to the ensemble average. Ramachandran plot: PROCHECK analysis of backbone torsion angles statistics of the deposited ensemble.