ARTICLE

^{1}H, ^{13}C, and ^{15}N resonance assignments and solution structure of the CID domain of SR-related- and CTD-associated factor 8 (SCAF8)

Weirong Dang[1] · Yutaka Muto[1,2,3] · Fahu He[1] · Mari Takahashi[1,2] · Kengo Tsuda[1] · Takashi Nagata[1,4] · Akiko Tanaka[1,2] · Naohiro Kobayashi[1,5] · Takanori Kigawa[1,2] · Peter Güntert[6,7,8,9] · Mikako Shirouzu[1,10] · Shigeyuki Yokoyama[1,11,12,13] · Kanako Kuwasako[1,2,3]

Abstract

Transcription mediated by RNA polymerase II (RNAPII) involves multiple stages, including initiation, promoter-proximal pausing for capping, elongation, and termination. The C-terminal domain of RNAPII (CTD) contains repetitions of the heptad consensus sequence, such as $Y_1S_2P_3T_4S_5P_6S_7$ with some variety, and the phosphorylated positions in the heptad sequences are altered according to the transcriptional stages. The interaction between several regulatory protein factors and the phosphorylated heptad sequence plays an important role in the accurate progress of transcription. A subset of these regulatory proteins possesses a CTD-interacting domain (CID) that specifically recognizes the phosphorylated CTD and mediates stage-specific transcriptional control. Among them, SCAF8 (RBM16), which also contains a CID, plays a key role in accurate transcriptional termination in conjunction with its paralog SCAF4. Despite their importance, the precise molecular mechanisms through which SCAF8 and SCAF4 coordinate transcriptional termination via their CID domains remain poorly understood. In this study, we report the ^{1}H, ^{15}N, and ^{13}C NMR resonance assignments and solution structure of the human SCAF8 CID domain. The structure exhibits an α1–α2–α3–α4–α5–α6–α7–α8 helical topology, consistent with the previously determined crystal structure. These assignments provide a valuable foundation for understanding how SCAF8 interacts with the RNAPII CTD and contributes to transcriptional elongation and termination.

Keywords RNA polymerase II · CID domain · SCAF8 · RBM16 · Transcriptional termination

Weirong Dang, Yutaka Muto have contributed equally to this work.

✉ Shigeyuki Yokoyama
  yokoyama@a.riken.jp

✉ Kanako Kuwasako
  kanameki@musashino-u.ac.jp

1   RIKEN, Systems and Structural Biology Center, 1–7-22 Suehiro-cho, Tsurumi, Yokohama 230–0045, Japan

2   RIKEN Center for Biosystems Dynamics Research, 1–7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230–0045, Japan

3   Faculty of Pharmacy and Research Institute of Pharmaceutical Sciences, Musashino University, Tokyo 202–8585, Japan

4   Institute of Advanced Energy and Graduate School of Energy Science, Kyoto University, Gokasho, Uji, Kyoto 611–0011, Japan

5   RIKEN Yokohama NMR Facility, 1–7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230–0045, Japan

6   Tatsuo Miyazawa Memorial Program, RIKEN Genomic Sciences Center, Yokohama 230–0045, Japan

7   Institute of Biophysical Chemistry, Goethe-University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

8   Institute of Molecular Physical Science, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

9   Department of Chemistry, Tokyo Metropolitan University, 1–1 Minami-Ohsawa, Hachioji, Tokyo 192–0397, Japan

10 Present address: RIKEN Center for Integrative Medical Sciences, 1–7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230–0045, Japan

11 Department of Structural Biology and Biochemistry, Institute of New Industry Incubation, Institute of Science Tokyo, Yushima 1–5-45, Bunkyo-ku, Tokyo 113–8549, Japan

12 Department of Drug Target Protein Research, Shinshu University School of Medicine, 3–1-1 Asahi, Matsumoto, Nagano 390–8621, Japan

13 RIKEN Cluster for Science, Technology and Innovation Hub, 1–7-22 Suehiro-cho, Tsurumi, Yokohama 230–0045, Japan

Biological context

The C-terminal domain of RNA polymerase II (RNAPII) contains repetitions of a heptad consensus sequence, such as $Y_1S_2P_3T_4S_5P_6S_7$ with some variety (in humans, 52 repetitions) (Sanchez et al. 2018; Eick et al. 2013; Corden 2013; Pineda et al. 2015; Jeronimo et al. 2016; Lyons et al. 2020; LeBlanc et al. 2021; Martinez and Svejstrup 2025). Transcription by RNAPII includes the accurate transcriptional events of initiation, promoter-proximal pausing for capping, elongation, and termination. Several post-transcriptional modifications of the heptad consensus sequence enable interactions between specific regulatory protein factors that control these steps (LeBlanc et al. 2021). In particular, the phosphorylation of Ser, Thr, and Tyr residues, as well as Pro isomerization, are crucial (Noble et al. 2005; Yogesha et al. 2014).

In mammals, when transcriptional activators bind to enhancer sequences on DNA, they also interact with the transcriptional co-activator complex, Mediator. The pre-initiation complex (PIC), consisting of RNAPII and general transcription factors, is recruited to the promoter sequence on DNA by the interaction between PIC and Mediator (Soutourina 2018). First, the C-terminal domain (CTD) of RNAPII has no modification. Then, S_5 and S_7 in the heptad repeats of the CTD of RNAPII are transiently subjected to O-GlcNAcylation by Mediator. Successively, S_5 and S_7 are phosphorylated just upon the removal of O-GlcMAc modification. Phosphorylation of the CTD causes RNAPII to escape from the PIC complex. Then RNAPII undergoes a transient promoter-proximal pausing step before transitioning to productive elongation. During the pausing step, the S_5-phosphorylation of CTD is necessary for the recruitment of the capping enzyme complex, which acts on the nascent 5'-end of the transcribed mRNA. Then, additional Y_1-phosphorylation of the CTD by Abl1 promotes p-TEFb/CDK9 to induce successive phosphorylation at the S_2 position, which initiates transcriptional elongation (Mayfield et al. 2019; Lyons et al. 2020). At the early stage of transcriptional elongation, the phosphorylation level of S_5 is rather high. Thus, phosphorylation at both S_2 and S_5 can be observed. During the progression of RNAPII toward the 3'-end of the transcript, the level of S_5 phosphorylation in the CTD decreases, in contrast to an increase in S_2 phosphorylation. Near the 3'-end position of the transcript and poly-A site, the level of the phosphorylation for S_2 is highest at and downstream of the poly-A, which promotes the recruitment of 3' RNA processing factors.

Several regulatory proteins distinguish S_2, S_5, or S_2/S_5-phosphorylated heptad sequences to control transcriptional stages, and CTD-interacting domains (CIDs) are known as binding modules for the phosphorylated heptad sequence. Pcf11, Rtt103, Nrd1, SCAF4, and SCAF8 (SR-related- and CTD-associated factors 4 and 8) were identified as such protein factors with a CID domain.

The CID domain is composed of approximately 140 amino-acid residues and principally recognizes phosphorylated heptad sequences. Each CID has its own preference for the phosphorylated positions. The CIDs of SCAF4 and SCAF8 prefer the S_2/S_5-phosphorylated sequence. In contrast, Pcf11 and Rtt103 are for S_2-phosphorylated ones, and that of Nrd1 for S_5-phosphorylated ones, respectively. Several structural studies using nuclear magnetic resonance (NMR) and/or X-ray crystallography have been performed for solo CID domains and those with target phosphorylated sequences, revealing their recognition mechanisms at atomic resolution (Meinhart and Cramer 2004; Meinhart et al. 2005; Becker et al. 2008; Kubíček et al. 2012; Jasnovidova et al. 2017a, b; Nemec et al. 2017; Zhou et al. 2022).

Human SCAF8 was first identified as an RNAPII binder in two-hybrid screening experiments (Yuryev et al. 1996; Patturajan et al. 1998). It is composed of 1271 amino-acid residues, which contain CID spanning residues 1–137 and an RNA recognition motif (termed the RRM domain) spanning residues 479–545. Recently, it was revealed that SCAF8 can prevent the premature termination of early polyadenylation sequences in human genes, cooperating with its paralog, SCAF4 (Gregersen et al. 2019, 2022). Intriguingly, SCAF8 alone functions as an RNAPII elongation factor; thus, without SCAF4, incorrect read-through transcripts are produced. Therefore, SCAF4 is simultaneously necessary for correct transcript termination. SCAF4 and SCAF8 are co-associated with actively elongating RNAPII. As described above, the CIDs of both SCAF8 and SCAF4 preferentially bind to the S_2/S_5-diphosphorylated CTD of RNAPII, and the structural information about recognition has been elucidated (Becker et al. 2008; Zhou et al. 2022). However, how the interaction mode between the CTD of RNAPII and SCAF8 is affected by the presence of the SCAF4 CID domain to produce an accurate transcript has not yet been elucidated.

NMR experiments have revealed that the CIDs of Pcf11 and Rtt103 cooperatively bind to neighboring CTD repeats for 3'-RNA processing and transcription termination (Lunde et al. 2010; Jasnovidova et al. 2017a, b). Furthermore, the issue is rather complicated because there is a possibility that some CID domains could interact with factors other than the phosphorylated heptad sequences. In *Saccharomyces cerevisiae*, Nrd1 is involved in terminating the transcription of non-coding RNAs, such as snoRNAs, along with Nab3 and Sen1. The CID of Nrd1 reportedly binds to the S_5-phosphorylated CTD and interacts with the C-terminal segment of Sen1, which contains the Nrd1 interacting motif composed of approximately 13 amino-acid residues (Zhang et al. 2019; Chaves-Arquero et al. 2023; Han et al. 2020). Furthermore,

in the case of Pcf11, the CID reportedly could interact with RNA molecules that compete with the phosphorylated heptad sequence. This competition plays an important role in regulating the interactions between the CTD and CID (Hollingworth et al. 2006). Furthermore, SCAF8 plays an important role other than in transcriptional stability. For example, SCAF8 is reportedly involved in determination of the alternative splicing sites (Schmok et al. 2024). Thus, it is necessary to consider interactions between CIDs and RNA/protein factors other than heptad sequences to elucidate the transcriptional termination system.

Herein, we report the ¹H, ¹³C, and ¹⁵N chemical shift assignments and solution structure of the CID of human SCAF8, as determined by heteronuclear NMR methods. Previously, the NMR chemical shift data of CIDs for the recognition of S_2- and S_5-phosphorylated heptad sequences (Pcf11 and Nrd1, respectively) have been reported (Xu et al. 2015; Kubíček et al. 2011). Thus, NMR data for the SCAF8 CID that recognizes the S_2/S_5-phosphorylated heptad are important for the comparison of different CID domains. The structural information and assignments obtained in this study provide information for the elucidation of the behavior of SCAF8 during transcriptional elongation and termination.

Methods and experiments

Sample preparation

The DNA encoding the CID domain (Met1–Gly137) of human SCAF8 (UniProt accession no. Q9UPN6) was subcloned from a full-length human cDNA clone using polymerase chain reaction. This DNA fragment was cloned into the expression vector pCR2.1 (Invitrogen) as a fusion protein with an N-terminal native His affinity tag and a tobacco etch virus protease cleavage site preceding the GSSGSS-GDN linker sequence. The ¹³C/¹⁵N-labeled fusion protein was synthesized using a cell-free protein expression system (Kigawa et al. 2004; Matsuda et al. 2007). The lysate was clarified by centrifugation at $16,000 \times g$ for 20 min and filtered through a 0.45-mm membrane (Merck Millipore). The clarified lysate was applied to a His-Trap column (Cytiva), eluted with a 12–500 mM imidazole gradient, and the tag was removed by incubation with tobacco etch virus protease for 1 h at 30 °C. The tag-free samples were further purified by Superdex-75 gel filtration chromatography (Cytiva). For NMR measurements, the resulting samples were concentrated to approximately 1.0 mM in 20 mM d-Tris–HCl buffer (pH 7.0), containing 100 mM NaCl, 1 mM 1,4-ᴅʟ-dithiothreitol-d₁₀, and 0.02% NaN₃ (in 90% H_2O/10% D_2O), using an Amicon Ultra-15 filter (3000 MWCO, Merck Millipore).

Consequently, the fusion protein contained an artificial tag-derived sequences, (GSSGSSGDN) at the N-terminus and (SGPSSG) at the C-terminus derived from the expression vector. The folding state of the ¹³C/¹⁵N-labeled protein, composed of 137 residues of the human SCAF8 CID, was verified by 2D ¹H–¹⁵N heteronuclear single quantum coherence (HSQC) experiments (Kigawa et al. 2004), which exhibited well-dispersed resonances (Fig. 1).

NMR spectroscopy and structure calculations

All NMR data were acquired at 298 K on Bruker 700 MHz and Bruker 800 MHz spectrometers and processed using *NMRPipe* software (Delaglio et al. 1995). Two-dimensional ¹H–¹³C and ¹H–¹⁵N HSQC spectra, three-dimensional HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(CCCO)NH, (H)CC(CO)NH, HCCH-TOCSY, HCCH-COSY, CCH-TOCSY, and NOESY spectra (Cavanagh et al. 2018; Clore and Gronenborn 1998) were used to assign all carbon, nitrogen, and hydrogen atoms of the ¹³C/¹⁵N-labeled sample.
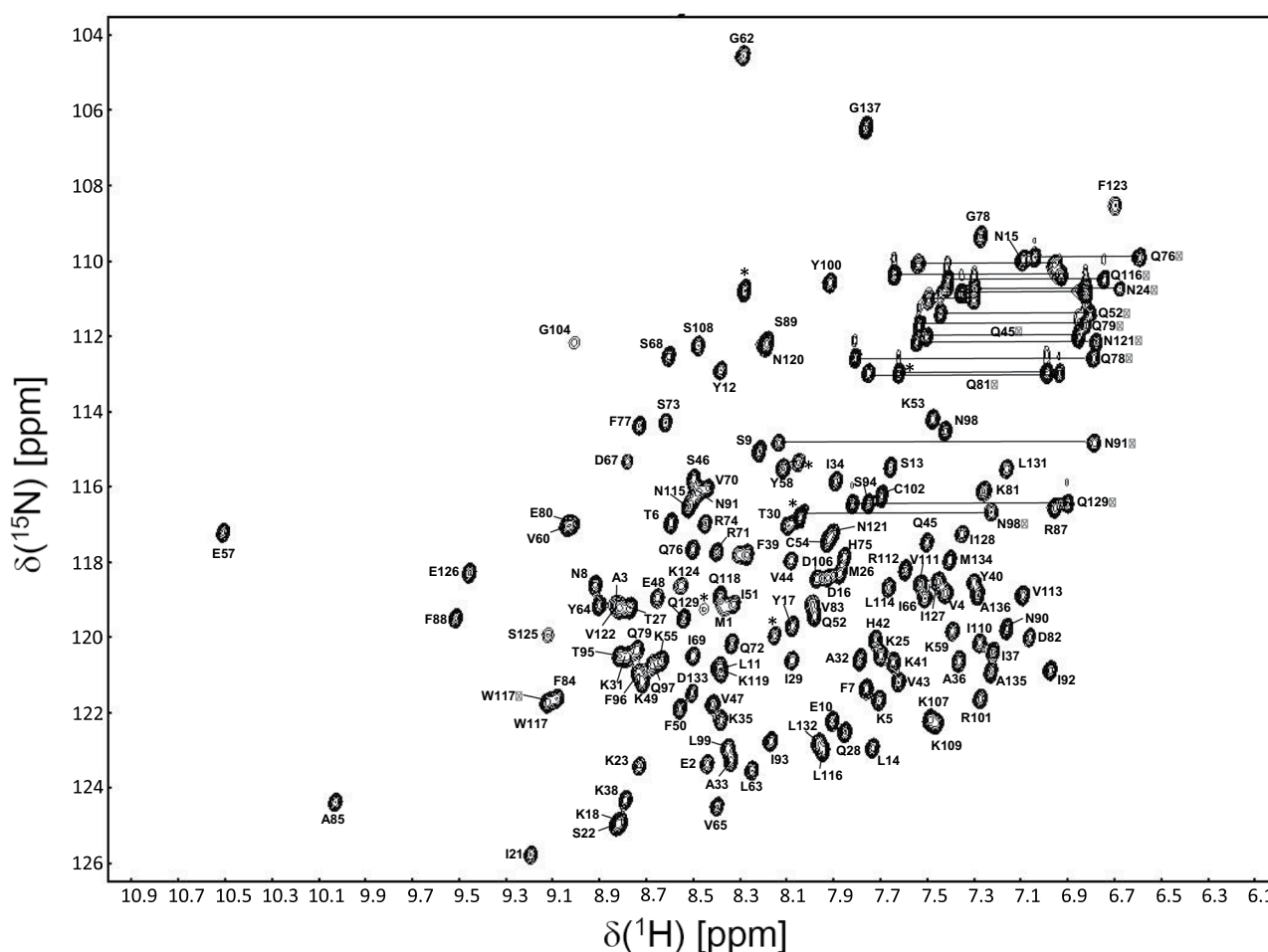
Nuclear Overhauser effect (NOE) peaks from the ¹⁵N- and ¹³C-edited 3D NOE spectroscopy (NOESY) spectra with a mixing time of 80 ms were converted to distance restraints for structural calculations of the CID domain of human SCAF8. The three-dimensional structure of the labeled protein was determined by combining automated NOESY cross-peak assignment and structural calculations with torsion angle dynamics (Herrmann et al. 2002) implemented in the program CYANA 2.1 (Güntert et al. 1997). The dihedral angle restraints for ϕ and ψ were obtained from the main-chain and ¹³Cᵝ chemical shift values using the TALOS program (Cornilescu et al. 1999) and by analyzing the NOESY spectra. Stereospecific assignments for isopropyl methyl and methylene groups were determined based on the patterns of inter- and intra-residual NOE intensities (Powers et al. 1993). The structure calculation started with 400 randomized conformers using a standard CYANA-simulated annealing schedule with 40,000 torsion angle dynamics steps per conformer (Güntert and Buchner 2015). The atomic coordinates of the 20 structures with the lowest CYANA target function values were deposited in the Protein Data Bank under the PDB accession code 2DIW.

Further refinements by restrained molecular dynamics simulation followed by restrained energy minimization were performed for the 80 conformers with the lowest final CYANA target function values, using the Amber12 program with the Amber 2012 force field and a generalized Born model (Case et al. 2005), as described previously (Tsuda et al. 2011). Finally, 20 conformers with the lowest Amber energy values were selected as the final structures and deposited in the Protein Data Bank under the PDB accession code

9U9X. The PROCHECK-NMR (Laskowski et al. 1996) and MOLMOL (Koradi et al. 1996) programs were used to validate and visualize deposited structures, respectively.

Extent of assignments and data deposition

The protein used in NMR analysis was a truncated and tagged version of the CID domain of human SCAF8. It is comprised of 137 residues, including nine and six tag residues at the N- and C-termini, respectively. The assigned ^{1}H-^{15}N HSQC spectrum of the human SCAF8 CID is shown in Fig. 1. All backbone amide hydrogen and nitrogen atoms were assigned except for the amide protons of Ala24 and Asp114 (Fig. 1). In total, all C^{α}, C^{β} chemical shifts, and C' chemical shifts except for Pro28 were determined. Furthermore, the chemical shifts of all non-exchangeable side-chain proton resonances and their related carbons were assigned. On the basis of the difference in the C^{β}-C^{γ} chemical shift of Pro residues (Schubert et al. 2002), we predicted that

the X-Pro peptide bond for Pro29 was in the *cis* conformation, whereas that of the other Pro residues was in the *trans* conformation. As confirmation of the *cis*-conformation of Pro29, we also identified a strong H^{α}- H^{α} sequential NOE between Pro28 and Pro29 (Wüthrich 1986). The chemical shift assignments for the human SCAF8 CID domain were deposited in the BMRB database under accession number 36749.

Solution structure of the CID domain of human SCAF8 (RBM16)

The quality of the NOESY spectra of the CID domain of human SCAF8 was appropriate for straightforward structural calculations. In the ^{15}N- and ^{13}C-edited 3D NOESY spectra, 4750 non-redundant distance restraints were identified, including 1407 long-range distance restraints. The backbone torsion angle restraints calculated using the TALOS program (Cornilescu et al. 1999) were also used for

Fig. 1 Backbone resonance assignment of the CID domain of human SCAF8. Assigned 2D ^{1}H-^{15}N HSQC spectrum of the CID domain of human SCAF8. Data were acquired on a Bruker 700 MHz spectrometer using the States-TPPI method with the water flip-back pulse sequence.
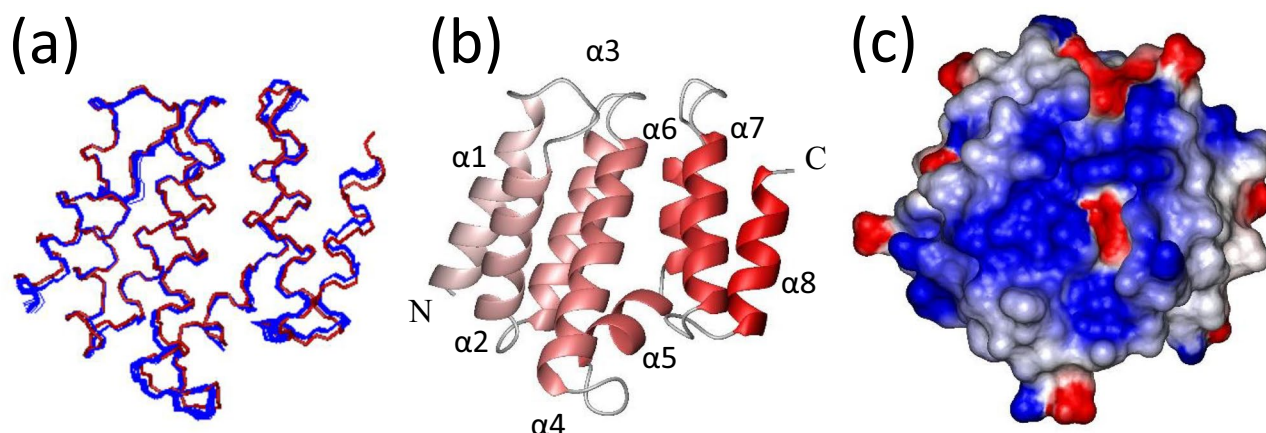
Fig. 2 Solution structures of the CID domain of human SCAF8. (**a**) The best-fit superposition of the backbone atoms from the 20 conformers of the CID domain of human SCAF8 with the lowest energy (blue lines) was calculated using CYANA2.1 and refined using Amber12. Red lines show the crystal structure determined previously (3D9O). (**b**) Ribbon presentation of the lowest energy structure of the SCAF8 CID domain. The eight main helices (α1–α8) are shown in coral with increasing density. (**c**) Electrostatic surface presentation of the SCAF8 CID domain in the same view as (**b**). Blue and red represent positive and negative electrostatic surface potentials, respectively.

structure calculations with the CYANA 2.1 (Güntert 2004; Güntert et al. 1997; Herrmann et al. 2002) and Amber12 programs (Case et al. 2005) (Supplementary Table 1). Among the 137 residues of the SCAF8 CID, the first Met1–Ala136 region was converged well. Thus, a bundle of 20 conformers representing the solution structures of the CID domain of human SCAF8, spanning residues Met1–Ala136, is shown in Fig. 2(a). Eight long helices (α1: 2–15, α2: 23–36, α3: 40–53, α4: 59–77, α5: 83–89, α6: 92–99, α7: 107–119, and α8: 125–135) were identified. The structural precision is characterized by RMSD values to the mean coordinates of 0.24 ± 0.05 Å for the backbone atoms and 0.82 ± 0.07 Å for all heavy atoms of residues Glu2–Ala136 of the SCAF8 CID domain. The structural quality of the SCAF8 CID domain was reflected by 100.0% of the (φ, ψ) backbone torsion angle pairs in the most favored and additionally allowed regions of the Ramachandran plot, according to the PROCHECK-NMR program (Laskowski et al. 1996). Statistics regarding the quality and precision of the final 20 best conformers representing the solution structure of the SCAF8 CID domain are provided in Supplementary Table 1.

The present NMR solution structure of the human SCAF8 CID domain is almost identical to that obtained from X-ray crystallographic studies (Becker et al. 2008). The crystal structure (3d9k) and the lowest energy NMR structure can be superimposed with a backbone RMSD of 0.60 Å for residues 2–136 (Fig. 2a).

The amino-acid sequence of the SCAF8 CID domain is 80% identical to that of SCAF4 (Supplementary Fig. 1). Several hydrophobic amino-acid residues in SCAF4 were replaced with hydrophilic residues in SCAF8. Therefore, the hydrophobic patches observed on the surface composed of helices α1 and α2 in SCAF4 (Supplementary Fig. 2a) were missing from the corresponding surface of SCAF8 (Supplementary Fig. 2b and c). It is probable that the difference in the hydrophobic patch patterns on the surfaces of α1 and α2 alters the molecular interactions mediated by the CID domains between SCAF8 and SCAF4.

We expect that the present NMR chemical shift data and structural study will provide a basis for functional studies of the CID domain of human SCAF8, leading to further understanding of its function in the production of accurate transcripts.

Amber12 refinement under accession code 9U9X.

Declarations

Competing interests The authors declare no competing interests.

References

Becker R, Loll B, Meinhart A (2008) Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. J Biol Chem 283:22659–22669. https://doi.org/10.1074/jbc.M803540200

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26:1668–1688. https://doi.org/10.1002/jcc.20290

Cavanagh J, Fairbrother WJ, Palmer AG III, Skelton NJ (2018) Protein NMR spectroscopy: principles and practice. Academic Press, San Diego

Chaves-Arquero B, Pérez-Cañadillas JM (2023) The Nrd1–Nab3–Sen1 transcription termination complex from a structural perspective. Biochem Soc Trans 51:1257–1269. https://doi.org/10.1042/BST20221418

Clore GM, Gronenborn AM (1998) New methods of structure refinement for macromolecular structure determination by NMR. Proc Natl Acad Sci U S A 95(11):5891–5898. https://doi.org/10.1073/pnas.95.11.5891

Corden JL (2013) RNA polymerase II c-terminal domain: tethering transcription to transcript and template. Chem Rev 113:8423–8455. https://doi.org/10.1021/cr400158h

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302. https://doi.org/10.1023/a:1008392405740

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. https://doi.org/10.1007/BF00197809

Eick D, Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev 113:8456–8490. https://doi.org/10.1021/cr400071f

Gregersen LH, Mitter R, Ugalde AP, Nojima T, Proudfoot NJ, Agami R, Stewart A, Svejstrup JQ (2019) SCAF4 and SCAF8, mRNA anti-terminator proteins. Cell 177:1797-1813.e18. https://doi.org/10.1016/j.cell.2019.04.038

Gregersen LH, Mitter R, Svejstrup JQ (2022) Elongation factor-specific capture of RNA polymerase II complexes. Cell Rep Methods 2:100368. https://doi.org/10.1016/j.crmeth.2022.100368

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378. https://doi.org/10.1385/1-59259-809-9:353

Güntert P, Buchner L (2015) Combined automated NOE assignment and structure calculation with CYANA. J Biomol NMR 62:453–471. https://doi.org/10.1007/s10858-015-9924-9

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298. https://doi.org/10.1006/jmbi.1997.1284

Han Z, Jasnovidova O, Haidara N, Tudek A, Kubicek K, Libri D, Stefl R, Porrua O (2020) Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. EMBO J 39:e101548. https://doi.org/10.15252/embj.2019101548

Herrmann T, Güntert P, Wüthrich 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. https://doi.org/10.1016/s0022-2836(02)00241-3

Hollingworth D, Noble CG, Taylor IA, Ramos A (2006) RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11 RNA 12:555–60. https://doi.org/10.1261/rna.2304506

Jasnovidova O, Krejcikova M, Kubicek K, Stefl R (2017a) Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. EMBO Rep 18:906–913. https://doi.org/10.15252/embr.201643723

Jasnovidova O, Klumpler T, Kubicek K, Kalynych S, Plevka P, Stefl R (2017b) Structure and dynamics of the RNAPII CTDsome with Rtt103. Proc Natl Acad Sci U S A 114:11133–11138. https://doi.org/10.1073/pnas.1712450114

Jeronimo C, Collin P, Robert F (2016) The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. J Mol Biol 428:2607–2622. https://doi.org/10.1016/j.jmb.2016.02.006

Kigawa T, Yabuki T, Matsuda N, Matsuda T, Nakajima R, Tanaka A, Yokoyama S (2004) Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. J Struct Funct Genomics 5:63–68. https://doi.org/10.1023/B:JSFG.0000029204.57846.7d

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55. https://doi.org/10.1016/0263-7855(96)00009-4

Kubíček K, Pasulka J, Černá H, Löhr F, Štefl R (2011) 1H, 13C, and 15N resonance assignments for the CTD-interacting domain of Nrd1 bound to Ser5-phosphorylated CTD of RNA polymerase II. Biomol NMR Assign 5:203–205. https://doi.org/10.1007/s12104-011-9300-y

Kubíček K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R (2012) Serine phosphorylation and proline isomerization in RNAPII CTD control recruitment of Nrd1. Genes Dev 26:1891–1896. https://doi.org/10.1101/gad.192781.112

Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) Aqua and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486. https://doi.org/10.1007/BF00228148

LeBlanc BM, Moreno RY, Escobar EE, Mukesh Ramani MKV, Brodbelt JS, Zhang Y (2021) What's all the phos about? insights into the phosphorylation state of the RNA polymerase II C-terminal domain via mass spectrometry. RSC Chem Biol 2:1084–1095. https://doi.org/10.1039/d1cb00083g

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol 17:1195–1201. https://doi.org/10.1038/nsmb.1893

Lyons DE, McMahon S, Ott M (2020) A combinatorial view of old and new RNA polymerase II modifications. Transcription 11:66–82. https://doi.org/10.1080/21541264.2020.1762468

Martinez DL, Svejstrup JQ (2025) Mechanisms of RNA polymerase II termination at the 3 '-end of genes. J Mol Biol 437:168735. https://doi.org/10.1016/j.jmb.2024.168735

Matsuda T, Koshiba S, Tochio N, Seki E, Iwasaki N, Yabuki T, Inoue M, Yokoyama S, Kigawa T (2007) Improving cell-free protein synthesis for stable-isotope labeling. J Biomol NMR 37:225–229. https://doi.org/10.1007/s10858-006-9127-5

Mayfield JE, Irani S, Escobar EE, Zhang Z, Burkholder NT, Robinson MR, Mehaffey MR, Sipe SN, Yang W, Prescott NA, Kathuria KR, Liu Z, Brodbelt JS, Zhang Y (2019) Tyr1 phosphorylation

promotes phosphorylation of Ser2 on the C-terminal domain of eukaryotic RNA polymerase II by P-TEFb. Elife 8:e48725. https://doi.org/10.7554/eLife.48725

Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature 430:223–226. https://doi.org/10.1038/nature02679

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. Genes Dev 19:1401–1415. https://doi.org/10.1101/gad.1318105

Nemec CM, Yang F, Gilmore JM, Hintermair C, Ho Y-H, Tseng SC, Heidemann M, Zhang Y, Florens L, Gasch AP, Eick D, Washburn MP, Varani G, Ansari AZ (2017) Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. Proc Natl Acad Sci USA 114:E3944–E3953. https://doi.org/10.1073/pnas.1700128114

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol 12:144–151. https://doi.org/10.1038/nsmb887

Patturajan M, Wei X, Berezney R, Corden JL (1998) A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. Mol Cell Biol 18:2406–2415. https://doi.org/10.1128/MCB.18.4.2406

Pineda G, Shen Z, de Albuquerque CP, Reynoso E, Chen J, Tu CC, Tang W, Briggs S, Zhou H, Wang JYJ (2015) Proteomics studies of the interactome of RNA polymerase II C-terminal repeated domain. BMC Res Notes 29:616. https://doi.org/10.1186/s13104-015-1569-y

Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM (1993) The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 32:6744–6762. https://doi.org/10.1021/bi00077a030

Sanchez AM, Shuman S, Schwer B (2018) RNA polymerase II CTD interactome with 3′ processing and termination factors in fission yeast and its impact on phosphate homeostasis. Proc Natl Acad Sci U S A 115:E10652–E10661. https://doi.org/10.1073/pnas.1810711115

Schmok JC, Jain M, Street LA, Tankka AT, Schafer D, Her HL, Elmsaouri S, Gosztyla ML, Boyle EA, Jagannatha P, Luo EC, Kwon EJ, Jovanovic M, Yeo GW (2024) Large-scale evaluation of the ability of RNA-binding proteins to activate exon inclusion. Nat Biotechnol 42:1429–1441. https://doi.org/10.1038/s41587-023-02014-0

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. J Biomol NMR 24:149–154. https://doi.org/10.1023/a:1020997118364

Soutourina J (2018) Transcription regulation by the mediator complex. Nat Rev Mol Cell Biol 19:262–274. https://doi.org/10.1038/nrm.2017.115

Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, Unzai S, Inoue M, Harada T, Watanabe S, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A, Sugano S, Guntert P, Yokoyama S, Muto Y (2011) Structural basis for the dual RNA-recognition modes of human Tra2-β RRM. Nucleic Acids Res 39:1538–1553. https://doi.org/10.1093/nar/gkq854

Wüthrich K. (1986) NMR of Proteins and Nucleic Acids. John Wiley & Sons; New York ISBN: 978–0–471–82893–8

Xu X, Pérébaskine N, Minvielle-Sébastia L, Fribourg S, Mackereth CD (2015) Chemical shift assignments of a new folded domain from yeast Pcf11. Biomol NMR Assign 9:421–425. https://doi.org/10.1007/s12104-015-9622-2

Yogesha SD, Mayfield JE, Zhang Y (2014) Cross-talk of phosphorylation and prolyl isomerization of the C-terminal domain of RNA polymerase II. Molecules 19:1481–1511. https://doi.org/10.3390/molecules19021481

Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc Natl Acad Sci U S A 93:6975–6980. https://doi.org/10.1073/pnas.93.14.6975

Zhang Y, Chun Y, Buratowski S, Tong L (2019) Identification of three sequence motifs in the transcription termination factor Sen1 that mediate direct interactions with Nrd. Structure 27:1156-1161.e4. https://doi.org/10.1016/j.str.2019.04.005

Zhou M, Ehsan F, Gan L, Dong A, Li Y, Liu K, Min J (2022) Structural basis for the recognition of the S2, S5-phosphorylated RNA polymerase II CTD by the mRNA anti-terminator protein hSCAF4. FEBS Lett 596:249–259. https://doi.org/10.1002/1873-3468.14256