Structural basis for CDK7 activation by MAT1 and Cyclin H

Stefan Peissert[a], Andreas Schlosser[a], Rafaela Kendel[a], Jochen Kuper[a,1], and Caroline Kisker[a,b,1]

[a]Rudolf Virchow Center for Integrative and Translational Bioimaging, Institute for Structural Biology, University of Würzburg, 97080 Würzburg, Germany; and [b]Comprehensive Cancer Center Mainfranken, University Hospital Würzburg, 97080 Würzburg, Germany

Edited by Seth A. Darst, Rockefeller University, New York, NY, and approved September 15, 2020 (received for review May 28, 2020)

Cyclin-dependent kinase 7 (CDK7), Cyclin H, and the RING-finger protein MAT1 form the heterotrimeric CDK-activating kinase (CAK) complex which is vital for transcription and cell-cycle control. When associated with the general transcription factor II H (TFIIH) it activates RNA polymerase II by hyperphosphorylation of its C-terminal domain (CTD). In the absence of TFIIH the trimeric complex phosphorylates the T-loop of CDKs that control cell-cycle progression. CAK holds a special position among the CDK branch due to this dual activity and the dependence on two proteins for activation. We solved the structure of the CAK complex from the model organism *Chaetomium thermophilum* at 2.6-Å resolution. Our structure reveals an intricate network of interactions between CDK7 and its two binding partners MAT1 and Cyclin H, providing a structural basis for the mechanism of CDK7 activation and CAK activity regulation. In vitro activity measurements and functional mutagenesis show that CDK7 activation can occur independent of T-loop phosphorylation and is thus exclusively MAT1-dependent by positioning the CDK7 T-loop in its active conformation.

CDK-activating kinase | transcription | cell-cycle control | kinase

Transcription and cell-cycle control are two essential processes which are tightly regulated by cyclin-dependent kinases (CDKs). The activity of CDKs is controlled by two events: 1) phosphorylation of the T-loop or activation loop and 2) interaction with their respective cyclins. Structural studies have shed light on the activation and regulation of heterodimeric CDK–cyclin complexes (reviewed in ref. 1). Interestingly, the cyclins share little sequence homology among each other but all contain either one or two versions of the cyclin box (2, 3). All CDKs exhibit a central kinase domain and variable N- and C-terminal domains. The vital activation loop is located in between the DFG and APE motifs within the kinase domain (1). As exemplified for human CDK2, activation is usually achieved sequentially through phosphorylation of Thr160 in the activation loop by a CDK-activating kinase (CAK) and subsequent binding to Cyclin A (4).

Among the CDK family, CDK7 assumes a unique and highly important position. CDK7 functions not only at the direct interface between the important processes of cell-cycle regulation and transcription but also, in contrast to other CDKs, requires two proteins for its activation, Cyclin H and the RING-finger protein MAT1. The heterotrimeric complex formed by these proteins activates the cell-cycle CDKs 1, 2, 4, and 6 via T-loop phosphorylation (5) and in humans is thus referred to as the CAK complex. Compared with the other CDKs, however, formation of the trimeric CAK complex does not necessitate phosphorylation of the threonine in the activation loop to activate the kinase (6–9).

In transcription, CAK is tethered to the core of the general transcription factor II H (TFIIH), ensuring the correct positioning of the active complex directly at the target site (10, 11), and phosphorylates the C-terminal domain (CTD) of the RNA polymerase II (RNA Pol II) subunit Rpb1 (12–15) at Ser5 and Ser7 of the CTD heptad repeat YSPTSPS (16–19). CTD phosphorylation disrupts the interaction between the Mediator and the RNA polymerase (20, 21) and thus triggers the release of RNA Pol II and enables promoter clearance. Even though the phosphorylation of Thr170 is not a requirement for CDK7 activity in the CAK complex, CTD phosphorylation was found to be enhanced when the activation loop threonine was phosphorylated in vitro by the CDK2–Cyclin A heterodimer (22), suggesting a regulatory role toward substrate specificity through this phosphorylation event. CDK7 was also shown to be involved in promoter-proximal pausing, cotranscriptional chromatin modification, and termination (18, 23), emphasizing the essential role during the entire process of RNA Pol II-based transcription. Finally, recent studies have implicated CAK in the resolution of phase-separated RNA Pol II droplets to enable transcription initiation (24).

Sustained proliferative signaling and the evasion of growth suppressors are both hallmarks of cancer (25), and it is therefore not surprising that CDKs and cyclins are often overexpressed and CDK inhibitory proteins are dysregulated in cancer cells (26). The potential of CDKs for therapeutic intervention in cancer treatment (27) has been exploited for CDK7, resulting in the selective CDK7 inhibitor THZ1 (28) and its derivate SY-1365 (29). Both inhibitors have been shown to specifically eliminate cancer cells that heavily rely on high levels of transcription such as triple-negative breast cancer (30, 31).

Despite its vital function in cellular processes and implications for cancer, structural information on the CAK complex has been limited to the single subunits Cyclin H and inactive CDK7 (32–34), impeding a detailed understanding of CDK7 activation and regulation by its binding partners. Here, we report the crystal structures of the apo and ATPγS-bound CAK complex from *Chaetomium thermophilum* which disclose the extensive interaction network within the heterotrimeric complex leading to the activated form of CDK7. While CDK7 and Cyclin H are engaged in conserved interactions between the N-terminal lobes, MAT1 is embedded in a gap between the C-terminal lobes of CDK7–Cyclin

Significance

The fundamental processes of cell-cycle regulation and transcription are linked by the heterotrimeric CDK-activating kinase (CAK) complex. We solved the crystal structure of the active CAK complex and provide a molecular rationale for CAK activation, regulation, and substrate recognition. Our data thus highly advance our understanding of this essential factor which is also a proven target for cancer therapy.

Author contributions: S.P., J.K., and C.K. designed research; S.P., A.S., and R.K. performed research; S.P., A.S., J.K., and C.K. analyzed data; and S.P., J.K., and C.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jochen.kuper@virchow.uni-wuerzburg.de or caroline.kisker@virchow.uni-wuerzburg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2010885117/-/DCSupplemental.

First published October 14, 2020.

H, thereby extensively interacting with both proteins and significantly stabilizing the heterotrimeric complex. The activation loop of CDK7 is tethered in a catalysis-competent conformation by Cyclin H and especially MAT1. Structure-based mutagenesis and subsequent functional analysis confirm the unique mechanism by which MAT1 activates CDK7.

Results

Biochemical Characterization of the CAK Complex. Toward the functional and structural characterization of the CAK complex, we recombinantly expressed and purified the homologs of CDK7 (ctCDK7), Cyclin H (ctCyclin H), and MAT1 (ctMAT1) from the model organism *C. thermophilum*. All proteins share a high degree of sequence conservation with their human and yeast homologs (*SI Appendix*, Fig. S1), indicating functional conservation. Full complex formation of the heterotrimeric complex was achieved using a heterologous approach. CtCyclin H and ctMAT1 were coexpressed in *Escherichia coli*, whereas ctCDK7 was expressed in insect cells and purified separately (*Materials and Methods*). CtMAT1 was expressed either as the full-length protein or as the CAK interaction module (10) containing residues 250 to 338 (ctMAT1short). Heterotrimeric complex formation was achieved by mixing the purified proteins and subsequent size-exclusion chromatography resulting in the complexes ctCAK or ctCAKshort depending on which MAT1 was used (Fig. 1 *A* and *B*). Both ctCAK complexes were stable during size-exclusion chromatography, indicating tight complex formation. The complexes were then analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), revealing an equimolar and homogeneous composition. To assess the kinase activity properties of the ctCAK complexes, we analyzed the phosphorylation activity of ctCAK against the complete CTD of the RNA Pol II subunit Rpb1 from *Saccharomyces cerevisiae* (scCTD) containing 26 heptad repeats coupled to the maltose binding protein (MBP-scCTD) using a fluorescence-based in-gel phosphoprotein-staining assay after SDS-PAGE (Fig. 1*C*). CtCAK efficiently phosphorylates the MBP-scCTD substrate as observed in the fluorescence scan (Fig. 1 *C*, *Inset*). Specific phosphorylation of the scCTD within the MBP-scCTD substrate was confirmed by mass spectrometry revealing that phosphorylation mainly occurred on serine 5 in the YSPTSPS motif of the analyzed heptad repeats (*SI Appendix*, Fig. S2*A*); minor phosphorylation was also observed on serine 2. In contrast, the activity toward the scCTD was drastically decreased to 2 and 25% for ctCDK7 alone or in the ctCDK7–ctCyclin H heterodimer, respectively (Fig. 1*C*, bar graph), indicating that both ctCyclin H and ctMAT1 are required for full activation of ctCDK7. Notably, the observed activity did not require T253 phosphorylation in the activation loop. CtCDK7 was obtained almost exclusively unphosphorylated after purification and was not autophosphorylated at T253 during the experiments as confirmed by mass spectrometry (*SI Appendix*, Fig. S2*B*). Subsequently, we analyzed the CDK7 complexes utilizing a coupled ATPase assay (Fig. 1*D*) by incubation with saturating amounts of adenosine triphosphate (ATP) supplemented with the MBP-scCTD substrate as indicated in Fig. 1*D*. Despite the difference in detection between the two assay systems, the relative activation pattern did not change. CtCDK7 in isolation is barely active with 5% (2.7 μmol ATP·L^{-1}·min^{-1}) and ctCDK7–ctCyclin H retains 29% (15.0 μmol ATP·L^{-1}·min^{-1}) of the ctCAK activity (52.4 μmol ATP·L^{-1}·min^{-1}). Importantly, the ATPase activity of the ctCAK complex corresponds to a k_{cat} of 4.4 (1/s) and is thus within the range of values reported for CDK–cyclin complexes. Phosphorylated CDK2–Cyclin A (k_{cat} ranges from 4.5 to 42.7 depending on the study) (35, 36), phosphorylated CTD kinase CDK9–Cyclin T (k_{cat} of 3.0) (37), and phosphorylated human CAK complex (k_{cat} of 3.7) (22) show comparable activities. Nevertheless, our results do not rule out a beneficial effect through the phosphorylation of the activation loop threonine on the CTD kinase activity of the ctCAK complex. Both experimental setups, however, clearly showed that significant activity can only be obtained if all three CAK components are present, in agreement with previous reports (6–8). Notably, the activity of ctCAKshort is highly comparable to full-length ctCAK (Fig. 1*D*), confirming that ctMAT1short is sufficient to activate CDK7 which is well in line with previous data from the Egly group (10).

Overall Structure of the CAK Complex. The ctCAKshort complex was amenable to crystallization and led to the apo structure of the heterotrimeric ctCAKshort complex (Fig. 2*A*). The structure was solved by molecular replacement using the structures of human Cyclin H and CDK7 as search models (32–34). After initial refinement, clear additional difference density permitted de novo building of ctMAT1. The crystal form belongs to space group $P2_1$ and contains two almost identical complexes (rmsd of 0.3 Å) in the asymmetric unit (*SI Appendix*, Fig. S3*A*). In the two heterotrimeric complexes, residues 273 to 338 or 272 to 338 of ctMAT1 and residues 78 to 399 or 77 to 399 of ctCDK7 are resolved in complex A and B, respectively. For ctCyclin H, residues 1 to 4, 50 to 74, 263 to 311, and 391 to 425 or 1 to 3, 50 to 74, and 391 to 425 are disordered in complex A and B, respectively. The final ctCAKshort model was refined to a maximum resolution of 2.6 Å and *R* factors of 20.3/24.9% (*R* and R_{free},

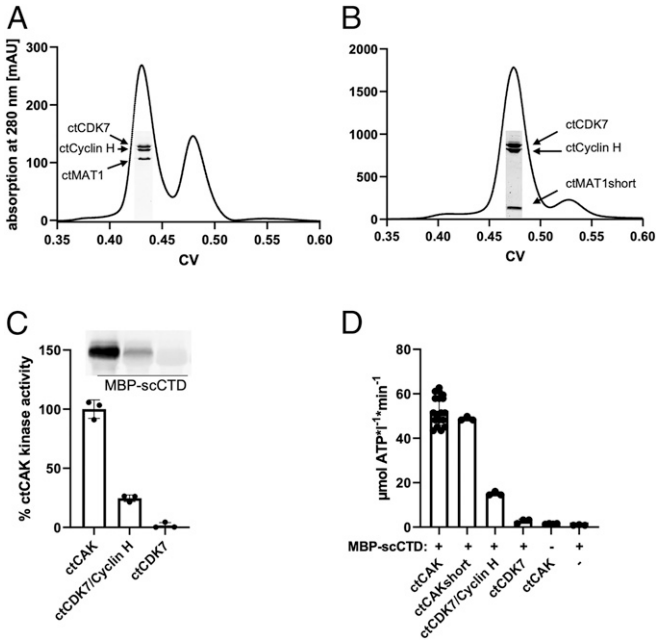

Fig. 1. Biochemical characterization of ctCAK complexes. (*A*) Size-exclusion chromatography profile of the ctCAK complex. (*A*, *Inset*) The main peak fraction on the SDS-PAGE. Proteins are indicated at their respective size. The *x* axis displays the column volume (CV). (*B*) As in *A* but in this case the formation of ctCAKshort was analyzed. (*C*) Kinase activity measurements using a fluorophore that interacts with the phosphorylated protein. Kinase activity was assessed in the presence of either ctCAK, ctCDK7–ctCyclin H, or ctCDK7 and an MBP-scCTD substrate containing 26 heptad repeats. A representative fluorescence scan is depicted in the gel (*Inset*) and data derived from three technical replicates were quantified (see *Materials and Methods* for details). (*D*) Kinase activity measurements monitoring the ATPase function in a coupled assay. The indicated protein complexes were incubated with saturating amounts of ATP and supplemented with the MBP-scCTD substrate containing 26 heptad repeats, as indicated. The highest velocity toward the end of the reaction was fitted and used to determine the ATP consumption over time. The error bars indicate the standard deviation.

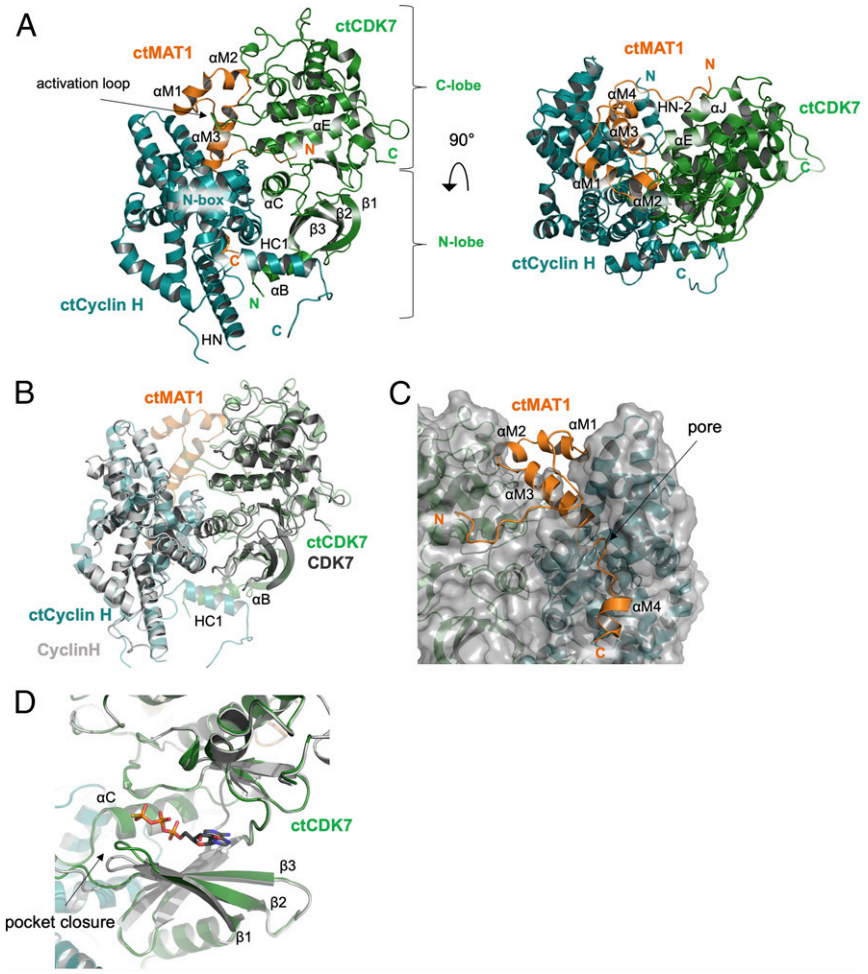

Fig. 2. Structure of the ctCAKshort complex from *C. thermophilum*. (*A*) Cartoon representation with two views rotated by 90° of the heterotrimeric ctCAKshort complex. Important secondary structure elements are indicated. CtCDK7 is depicted in green, ctCyclin H in deep teal, and MAT1 in orange. (*B*) Superposition of the human CDK7 (PDB ID 1UA2; dark gray) and Cyclin H (PDB ID 1KXU; light gray) structures onto the ctCAKshort structure. The color coding for ctCAKshort was chosen as in *A*. (*C*) The interface of ctMAT1 with ctCDK7 and ctCyclin H within ctCAKshort. Color coding is as in *A*. CtMAT1 is shown in cartoon representation whereas ctCDK7 and ctCyclin H are shown in transparent surface representation with underlying cartoon representation. (*D*) Superposition of the ctCAKapo (gray) and ctCAK ATPγS (green) structures; ATPγS is shown in stick representation. The arrow indicates the pocket closure.

respectively; Table 1). For all subsequent analyses, we have used complex A.

The structure of the heterotrimeric complex reveals a tight interaction of all three subunits with each other. Most contacts between ctCDK7 and ctCyclin H are formed via the N-terminal lobe of ctCDK7 and the N-terminal cyclin box fold of ctCyclin H (Fig. 2*A*) with an interface area of ∼1,450 Å^2. CtMAT1short adopts a loop-rich, extended fold featuring four α-helices in total (Fig. 2*A* and *SI Appendix*, Fig. S1*A*). All resolved parts of ctMAT1short are essentially involved in protein–protein interactions. Intriguingly, ctMAT1short predominantly seals a large cleft between the C-lobes of ctCyclin H and ctCDK7, thereby contributing a total interface area of 2,840 Å^2, thus forming a highly stable complex with extensive interactions. The interface area between ctCyclin H and ctMAT1 (∼1,730 Å^2) exceeds the one buried by ctCDK7 and ctMAT1 (∼1,110 Å^2). However, both ctMAT1 interfaces together bury more surface area than through the interaction between ctCDK7 and ctCyclin H, indicating the importance of MAT1 for the stability of the CAK complex as previously described (6–8).

The superposition of the individual subunits of the ctCAK complex with human Cyclin H and CDK7 reveals that not only their sequences but also their folds are highly comparable, yielding rmsd values of 1.7 and 1.6 Å, respectively, suggesting high functional conservation (Fig. 2*B*). A closer look at the CDK7 superposition shows that an additional N-terminal α-helix (αB) is resolved in ctCDK7. This helix interacts with HN and HC1 of ctCyclin H, of which HC1 was not resolved in the apo structure of human Cyclin H. Another noticeable difference is the loop between β3 and the C-helix of ctCDK7 which is now resolved and could not be observed in the human CDK7 structure. Additionally, an N-terminal extension of ctCyclin H (amino acids 5 to 20) featuring helices HN-2 and HN-1 can be seen in our ctCAKshort structure which form extensive contacts with both ctCDK7 and ctMAT1short.

Interestingly, the general architecture of the ctCDK7–ctCyclin H binary complex within the heterotrimeric ctCAKshort structure closely resembles that of other dimeric CDK–cyclin complexes such as CDK2–Cyclin A, CDK8–Cyclin C, and CDK9–Cyclin T (38–40). One of MAT1's functions in the trimeric complex is thus to act as a stabilizing factor and tighten the binary complex through its interactions with both CDK7 and Cyclin H.

The N-terminal loop (270 to 293) of ctMAT1 mostly forms contacts with αE and αJ of ctCDK7 and then traverses the cleft toward ctCyclin H to interact with HN-2. The loop folds back on

Table 1. Data collection and refinement statistics

	ctCAKshort, PDB ID 6Z3U	ctCAKshort ATPγS, PDB ID 6Z4X
Data collection		
Wavelength, Å	1.0723	0.9184
Space group	$P2_1$	$P2_1$
Unit-cell parameters		
a, b, c, Å	81.4, 85.2, 160.4	90.0, 84.9, 154.5
α, β, γ, °	90, 96.9, 90	90, 96.2, 90
Resolution range, Å	75.9–2.6 (2.82–2.6)	46.7–3.0 (3.2–3.0)
Unique reflections	33,808 (1,690)	29,053 (1,703)
R_{merge}, %	33.6 (186.5)	27.7 (98.3)
R_{pim}, %	13.9 (81.2)	17.9 (63.4)
$I/\sigma I$	6.6 (1.3)	5.2 (1.3)
CC(1/2)	0.980 (0.249)	0.952 (0.409)
Multiplicity	6.8 (6.2)	3.3 (3.3)
Completeness, %		
Spherical	50.4 (12.1)	61.3 (17.5)
Ellipsoidal	83.4 (64.4)	89.3 (67.1)
Refinement		
Resolution range, Å	75.9–2.6	46.4–3.0
Unique reflections	33,774	29,048
Number of atoms	11,398	11,442
R_{work}, %	20.3	20.3
R_{free}, %	24.9	24.7
Mean *B* factor, Å^2	45.6	53.2
Rms deviations		
Bond lengths, Å	0.003	0.002
Bond angles, °	0.507	0.47
Ramachandran statistics, %		
Favored	98.05	97.33
Allowed	1.73	2.53
Outliers	0.22	0.1

Values in parentheses correspond to the highest resolution shell. Data were scaled using the STARANISO server.

α-helix αM3 (312 to 324) of ctMAT1 which is buried in the center of the gap between CDK7 and Cyclin H, thereby extensively interacting with both proteins (Fig. 2*C*). Residues 294 to 311 of ctMAT1 form αM1 and αM2 that continue back toward ctCDK7 and, importantly, form contacts with the activation loop of ctCDK7. αM3 extends toward Cyclin H and the following loop (324 to 330) closes a pore formed by helices HN-2, HN-1, HN, H2″, and H5′ of ctCyclin H as well as ctMAT1 residues 283 to 286. This pore is formed by the previously unresolved α-helices from Cyclin H that are likely stabilized due to the presence of MAT1 being tightly embedded in the CDK7–Cyclin H interface. This network further emphasizes the special role of MAT1 in the heterotrimeric CAK complex toward stabilization and activation. Interestingly, the linker in Cyclin H between the two helices HN-1 and HN-2 harbors the conserved S15 (S5 in human Cyclin H) which is the primary target of CDK8–Cyclin C–mediated inhibition of CAK activity and concomitant blocking of transcription initiation by TFIIH (41).

We also determined the crystal structure of the ctCAKshort complex bound to ATPγS at 3-Å resolution and refined it to *R* factors of 20.3/24.7% (*R* and R_{free}, respectively; Table 1). The conformation of the ATPγS-bound ctCAKshort complex is highly comparable to that of the apo structure with an rmsd value of 0.5 Å despite the conformational change of the loop between β-strands β1 and β2 leading to the closure of the ATP-binding pocket (Fig. 2*D*). The nucleotide-binding mode of ctCAKshort is similar to that observed for CDK7 and CDK2 (34, 42).

Structural Basis of CDK7 Activation. Our structure provides an initial glimpse of CDK7 activation through the formation of the heterotrimeric complex. Two events lead to the active form of the kinase: First, α-helix H3 of Cyclin H pushes the C-helix of CDK7 toward the active site. Subsequently, the salt bridge between the active-site lysine K122 (K41 in human CDK7) involved in the phosphotransfer reaction and the conserved glutamate E141 (E62 in human CDK7) from the C-helix can be formed. The conformational change also fosters the required orientation of D238 (D155 in human CDK7) in the DFG motif ("DFG in") to be involved in magnesium binding (Fig. 3*A*), transforming the kinase to a state capable of ATP hydrolysis. Then, final activation is achieved by one of the most prominent features revealed in the heterotrimeric ctCAKshort complex: the extensive interaction pattern of the activation loop from ctCDK7 with ctMATshort and partly ctCyclin H leading to a conformation of the activation segment that is highly comparable to the active form of CDK2 (Fig. 3*B*). The position of the activation loop is stabilized by hydrophobic interactions including the tip of the activation loop around the highly conserved P248 that points into a hydrophobic pocket consisting of residues L296, Y299, Y310, F312, and Y315 from ctMAT1. Two hydrogen bonds between Y310 of ctMAT1 (Y287 in human MAT1) and the carbonyl of A246 of ctCDK7 (G173 in human CDK7) as well as between R190 of ctCyclin H (R165 in human Cyclin H) and D247 of ctCDK7 (S165 in human CDK7) further stabilize the position of the activation loop (Fig. 3*C*). This intricate network is likely holding the loop in a catalysis-competent position, thereby activating ctCDK7 (Fig. 3 *B* and *C*). Importantly, our structure represents the unphosphorylated state of the activation loop, and thus supports previous findings (6–8) and is well in line with our own biochemical data that this activation mechanism does not

require the additional phosphorylation of the activation loop threonine (Fig. 1 *C* and *D*). However, at the position of T253, we observe that the hydroxyl group points toward a positively charged pocket. This pocket is formed by R140 from the repositioned C-helix, R219 from the HRD motif, and R243 within the activation loop (R61, R136, and K160 in human CDK7, respectively) that seems to be well-suited to accommodate the potentially phosphorylated T253 (Fig. 3*D*). In our structure, this position is occupied by a glutamate from the C terminus of a symmetry-related ctCyclin H subunit (*SI Appendix*, Fig. S3*B*), indicating that the complex might be stabilized further once phosphorylation has occurred.

To validate our structural observations that the stabilization of the activation loop is MAT1-dependent, we performed structure-based mutagenesis on ctMAT1. The variants were designed to specifically weaken the interaction with the activation loop and not to interfere with the overall interaction pattern. We chose L296 and Y299 (L272 and V275 in human MAT1), with the first residue being strictly conserved and the second being type conserved (*SI Appendix*, Fig. S1) and generated the L296R and Y299A variants. CtCAK complexes were assembled as described above in the presence of the MAT1 variants and subsequently analyzed for their activity by monitoring the phosphorylation of the MBP-scCTD substrate. Both the L296R and the Y299A ctCAK variants showed a significant decrease in kinase activity with 20 and 23 μmol ATP·L^{-1}·min^{-1}, respectively, compared with the wild-type complex with 52 μmol ATP·L^{-1}·min^{-1} (Fig. 3*E*). The reduced activity of the ctCAK variants is similar to that observed for the binary ctCDK7–ctCyclin H complex with 15 μmol ATP·L^{-1}·min^{-1}, indicating a substantial loss of MAT1

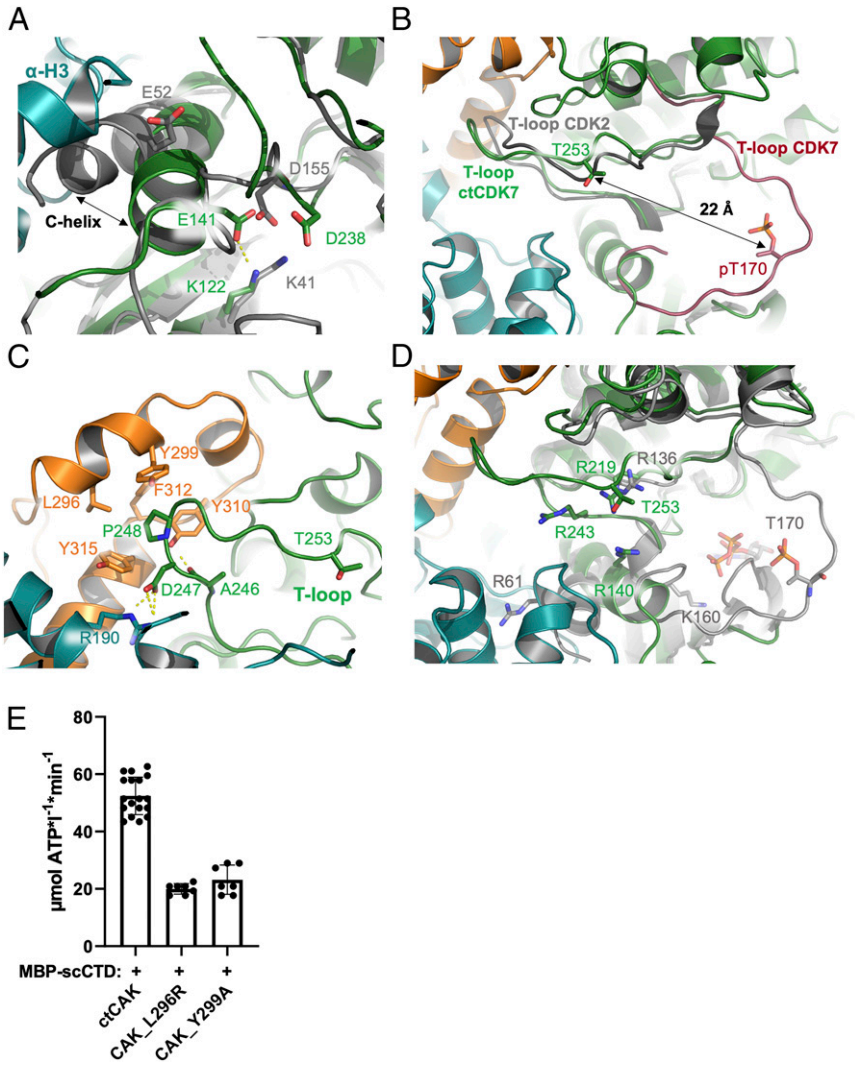

Fig. 3. Structural basis of ctCDK7 activation. (*A*) Close-up of a superposition containing human CDK7 (gray; PDB ID 1UA2) and ctCAKshort (CDK7 in green and Cyclin H in deep teal). The C-helix of CDK7 and relevant active-site residues are labeled. (*B*) T-loop movement between ctCAKshort (colored as in *A* and MAT1 in orange), CDK2 (PDB ID 1QMZ; gray), and human CDK7 (PDB ID 1UA2; raspberry). The structures were superimposed based on ctCKD7. For clarity, only the T-loops of CDK2 and human CDK7 are shown. The T-loop threonine is highlighted. (*C*) ctCDK7 T-loop stabilization via ctMAT1 and ctCyclin H in ctCAKshort. Coloring is as in *B*. Relevant residues that contribute to the stabilization are shown in stick mode and hydrogen bonds are shown as dotted lines. (*D*) The T-loop phosphothreonine binding site in ctCAK. Close-up of a superposition between ctCAK (based on ctCDK7) and human CDK7 (PDB ID 1UA2; in gray). Coloring is as in *B*. Residues are shown in stick mode, labeled, and colored according to their respective protein (green CDK7 from the ctCAK complex, and gray human CDK7). (*E*) ctCAK ATPase activity experiments comparing wild-type ctCAK with ctCAK MAT1 variants using the MBP-scCTD as a substrate (for reference, see Fig. 1 and *Materials and Methods*). Data for wild-type ctCAK were taken from Fig. 1*D* for comparison. The error bars indicate the standard deviation.

activation (compare Fig. 1*D* and Fig. 3*E*). To investigate the effect of ctMAT1 and the ctMAT1 variants on the stability of the ctCAK complex, thermal shift assays were performed with ctCDK7–Cyclin H, ctCAK, and ctCAK containing the MAT1 variants. Both mutations in MAT1 led to a decrease in melting temperature, suggesting reduced complex stability due to the missing contacts with the activation loop (*SI Appendix*, Fig. S4). Importantly, the melting curve is still highly cooperative, indicating a single unfolding event of the heterotrimeric complex. In contrast, the melting curve for the ctCDK7–Cyclin H heterodimer reveals two separate unfolding events at lower temperatures than the wild type and the mutated complexes. Our data thus suggest that there are fundamental differences between the stability of the heterodimeric complex and the heterotrimeric complexes, highlighting the important role of ctMAT1 for complex integrity and function. Both L296 and Y299 from ctMAT1 seem to play a significant role toward trimeric complex formation and correct positioning of the CDK7 activation loop.

Discussion

The human CAK complex assumes a special position among the CDK family due to the dual activity in both transcription and cell-cycle progression. This is further emphasized by the requirement for a third protein acting together with the cyclin to activate CDK7. To obtain insights into the underlying activation mechanism, we solved the crystal structure of the CAK complex from *C. thermophilum* in its active form that sheds light on how MAT1 achieves CDK7 activation in cooperation with Cyclin H. The most prominent feature of the ctCAK complex is the intricate interaction pattern of ctMAT1 with its two binding partners resulting in a tightly packed heterotrimeric complex (Fig. 2*A*). These extensive interactions explain the observed stabilization effect of ctMAT1 on the ctCDK7–Cyclin H heterodimer (6–8). Our structure reveals the two key events which lead to the activation of ctCDK7: 1) α-Helix H3 of Cyclin H pushes the C-helix of ctCDK7 toward the active site and 2) the activation loop is stabilized by a hydrophobic pocket formed by residues from ctMAT1 without prior phosphorylation (Fig. 3 *A*–*D*). Site-directed mutagenesis of specific residues within this hydrophobic pocket confirmed its importance for CDK7 activation (Fig. 3*E*). To our knowledge, the specific ctCDK7 activation mechanism has not been observed for any other CDK–cyclin pair (reviewed in ref. 1) so far. Notably, the other CDK–cyclin pairs require phosphorylation for activation, whereas ctCAK displays significant activity without this modification. The PHO85–PHO80 complex from *S. cerevisiae* (43) and the complex of the Ringo activator Spy1–CDK2 (44) can bypass the necessity for phosphorylation like the CAK complex. However, in both cases the binding partner of the kinase mimics the phosphorylation of the activation loop by providing negatively charged residues, thus suggesting an entirely different mechanism.

The structures of the CDK7–Cyclin H pair within the CAK complex are highly comparable to the overall conformation of other known CDK–cyclin structures with the conserved contacts between the N-terminal lobes of both proteins (Fig. 4*A*). Notably, there is some variance in the nature and extent of the interfaces resulting in markedly different rotational angles of the cyclins with respect to the CDKs, especially when the cell-cycle CDKs are compared with the transcriptional CDKs (Fig. 4*A*). The CDK7–Cyclin H conformation can be positioned in between the "closed" form of CDK2–Cyclin A and the "open" form of CDK9–Cyclin T. Another key difference between cell-cycle and transcriptional CDKs is the α-helix HN of the cyclins that forms extensive interactions with cell-cycle CDKs but not with transcriptional CDKs. Importantly, α-helix HN in CDK2–Cyclin A and CDK1–Cyclin B (45) is located in a similar position as α-helix M3 of MAT1, suggesting a conserved regulatory mechanism (Fig. 4*A*). The additionally resolved N-terminal α-helix (αB) of ctCDK7 is in a comparable position as the one observed for CDK8 which was described as a unique Cyclin C specificity helix (40). Despite the low sequence conservation between the N-termini of CDK8 and CDK7, helix αB of ctCDK7 also interacts with its specific cyclin (Fig. 4*A*). Residues D79 and E89 from the αB-helix of ctCDK7 form multiple hydrogen bonds with R42, R365, and R375 of ctCyclin H, thereby contributing to the CDK7–Cyclin H interface. This interaction of the N terminus of the CDK with the corresponding cyclin may be shared among the transcriptional CDKs since they contain N-terminal extensions compared with the cell-cycle CDKs. It is tempting to speculate that the additional interaction compensates for the open CDK–cyclin form of transcriptional kinases.

Another interesting feature of the CAK complex is the loop between β3 and the C-helix of CDK7 that was not resolved in the inactive human CDK7 structure. Superposition of three exemplary CDK–cyclin structures indicates that this loop could contribute to the active-site cleft formation (Fig. 4*B*). The sequence conservation between different CDKs is rather low in this area (Fig. 4*B*) and it is thus tempting to speculate that the loop between the conserved AXK motif and the C-helix may contribute to substrate binding or specificity.

Our biochemical studies of the ctCAK complex have shown that it specifically phosphorylates the MBP-scCTD substrate containing 26 heptad repeats of the RNA Pol II subunit Rpb1 at position 5 (S5), which is well in line with previous data (16–18). Furthermore, activation of ctCDK7 requires the presence of ctCyclin H and ctMAT1 (Fig. 1 *C* and *D*) but importantly does not require T-loop phosphorylation (*SI Appendix*, Fig. S2). These results are supported by previous in vivo studies which showed that mutation of the activation loop threonine of the CDK7 homolog Kin28 exhibited no growth phenotype whereas catalytic-site mutants led to a clear growth inhibition or were not viable (9). Within this analysis it was also observed that the activation loop threonine is not required for CTD phosphorylation in vivo. However, we cannot rule out that phosphorylation of the activation loop threonine has additional beneficial effects on CTD kinase activity of the ctCAK complex. Our data thus implicate that the active site of CDK7, as observed in our structure, resembles an architecture of a catalytically competent CDK7.

A superposition of our ATPγS-bound CAK structure with the active CDK2–Cyclin A structure containing a nucleotide analog and a substrate peptide shows that the overall architecture of both active sites is well-conserved (Fig. 4*C*, side-by-side view). This similarity permits a closer look at how CDK7 might recognize the heptad-repeat substrate and phosphorylates the S5 position. Based on the CDK2–Cyclin A structure, we built a substrate model of the heptad-repeat CTD and placed it in our ATPγS CAKshort structure (Fig. 4 *C*, *Left*). Intriguingly, the S4 of the CDK2 substrate would assume the role of S5 in the CAK heptad-repeat model and would be ideally positioned for the phosphotransfer reaction. The +1 site would be occupied by the vital proline as in the CDK2–Cyclin A peptide, preserving the special proline geometry at this position. The next residue in the +2 site would be S7 at the end of the heptad repeat. In the CDK2–Cyclin A/substrate structure the following +3 position is a lysine residue that interacts with the phosphothreonine of the T-loop which is additionally stabilized by a positive patch generated by three conserved arginines (Fig. 4 *C*, *Right*). Strikingly, this position would be occupied by Y1 of the next heptad repeat in our substrate model of the ctCDK7 active site. This next repeat tyrosine would be stabilized by a hydrophobic site created by Y130 and P136 of ctCDK7 and F150 of ctCyclin H. The hydrophobic nature of the site is conserved in the human and yeast proteins (*SI Appendix*, Fig. S1), thus supporting our model. Notably, phosphorylation of the T-loop threonine would be well in line with this model since the tyrosine could readily engage in hydrogen bonding to the phosphothreonine as the corresponding

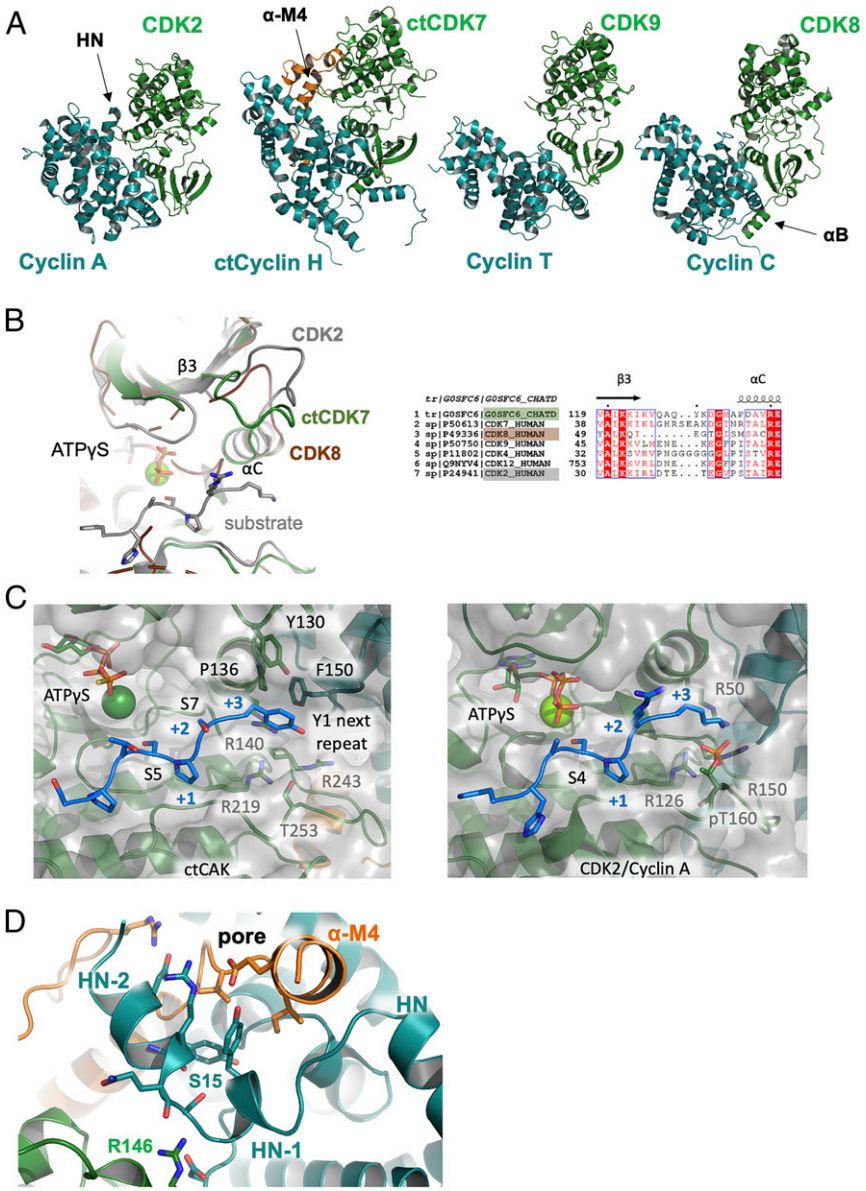

Fig. 4. Comparison of ctCAK with other CDK–cyclin complexes and substrate binding. (*A*) Side-by-side view of ctCAK with structures of CDK2–Cyclin A (PDB ID 1QMZ), CDK9–Cyclin T (PDB ID 3BLH), and CDK8–Cyclin C (PDB ID 4F7S) based on their superposition. The cyclins are colored in deep teal, and the kinases in green. MAT1 is shown in orange. (*B*) Orientation of the loop between β3 and αC as observed in ctCDK7 (green), CDK2 (PDB ID 1QMZ; gray), and CDK8 (PDB ID 3RGF; brown). The peptide substrate (gray) and ATPγS are derived from CDK2 and additionally depicted in stick mode. The sequence alignment (*Right*) highlights the relevant loop section of ctCDK7 (G0SFC6_CHATD) with the other CDKs indicated by their UniProt entry. The superimposed CDKs are highlighted. (*C*) Side-by-side view of the substrate-binding area of ctCAK (*Left*) and CDK2–Cyclin A (*Right*). CDKs are colored green and the cyclins in deep teal. Structures are shown in cartoon mode with a transparent surface and specific residues are shown in stick mode. The peptide model (*Left*) and the peptide (*Right*) are colored in blue. Relevant positions and residues are indicated. (*D*) Close-up view of a regulatory phosphorylation site in Cyclin H from ctCAK. The color scheme is similar to *A*.

lysine does in the CDK2–Cyclin A structure. The surrounding arginine pocket is also well-conserved in human CDK7 (Fig. 3*D* and *SI Appendix*, Fig. S1*A*). The model thus provides a structural rationale as to why the functional unit of the heptad substrate requires at least four additional residues (reviewed in ref. 46). Our model of substrate processing is also highly compatible with the observation that CAK activity could be enhanced or the complex could be further stabilized by T-loop phosphorylation at position 253 (T170 in human CDK7) (6–8, 22).

Finally, the CAK complex is not only activated but can also be inactivated through phosphorylation of S5 and S304 in Cyclin H by the kinase module CDK8–Cyclin C of the Mediator complex (41). S5 phosphorylation displayed a more severe phenotype than S304 phosphorylation and was linked to a substantial reduction of CAK activity toward the CTD of RNA Pol II and, consequently, deficiencies in transcription initiation. The corresponding residue in ctCyclin H is S15 which is located in a three-amino acid-long linker between α-helices HN-2 and HN-1 being involved in the described pore (Fig. 4*D*). Phosphorylation of this residue requires a change in the local conformation, leading to a disruption of the interaction with the C-helix of ctCDK7 and, particularly, the C terminus of ctMAT1short. As a consequence,

the complex will most likely be disrupted, thereby leading to the observed inactivation.

In conclusion, our structure of the fully functional CAK complex provides a comprehensive view of the CDK7 activation mechanism through the tight interaction of CDK7, Cyclin H, and MAT1 revealing the active CAK conformation that is essential for transcription and cell-cycle control.

Materials and Methods

Molecular Biology. For the expression construct of full-length ctCyclin H (UniProtKB G0SH78) and ctMAT1 (UniProtKB G0SF48), a modified pET-22 vector (EMBL) was used in which the pelB leader sequence was replaced by the sequence coding for a Twin-Strep tag (IBA) and a 3C protease cleavage site (pETtwin; provided by Florian Sauer, Rudolf Virchow Center for Integrative and Translational Bioimaging, Institute for Structural Biology, University of Würzburg, Würzburg, Germany). The coding sequences of ctCyclin H followed by a second ribosomal binding site and thioredoxinA-6xHis-3C-ctMAT1 were cloned behind the 3C cleavage site of pETtwin by sequence- and ligation-independent cloning (SLIC) (47). The design of ctMAT1short (250 to 338) was guided by pairwise protein structure prediction (48). For the expression constructs of ctCyclin H and ctMAT1short the corresponding coding sequences were inserted behind the 3C cleavage site of pETtwin (ctMAT1short) or pCOLA-SMT3 (ctCyclin H) which is a modified pCOLA vector (provided by Florian Sauer) containing the expression cassette from pETM-22 (EMBL) except that the thioredoxinA-6xHis tag was replaced by a 6xHis-SMT3 tag. The coding sequence of the RNA Pol II subunit Rpb1 CTD (UniProtKB P04050; 1534 to 1733) was amplified from *S. cerevisiae* S288c genomic DNA and cloned into a pETM-41 vector (EMBL) by restriction cloning using NcoI and KpnI, resulting in an MBP-scCTD construct.

For the expression in insect cells, ctCDK7 was first inserted into the pETtwin vector by SLIC and the resulting Twin-Strep-tag-3C-ctCDK7 construct was then transferred into a pFastBac1 transfer vector (Invitrogen) by restriction cloning using EcoRI and XbaI. The pFastBac1 construct was integrated into the EMBacY baculovirus genome via Tn7 transposition. Mutants were generated according to the QuikChange site-directed mutagenesis protocol (Stratagene).

Protein Expression and Purification. All constructs except for ctCDK7 were expressed in BL21star (DE3) cells (Invitrogen) carrying the pRARE2 plasmid (Novagen) and initially purified using the same protocol. Cells were grown in lysogeny broth (LB) medium supplemented with 100 μg/mL ampicillin (pETtwin) or 50 μg/mL kanamycin (pCOLA-SMT3) and 34 μg/mL chloramphenicol at 37 °C until they reached an OD_{600} of 0.7. The temperature was then reduced to 18 °C and expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside overnight. The cells were harvested and resuspended in lysis buffer A (20 mM Hepes, pH 7.5, 300 mM NaCl, 25 mM imidazole, 1 mM Tris[2-carboxyethyl]phosphine; TCEP) supplemented with 0.5 mM phenylmethanesulfonylfluoride (PMSF) and DNaseI. After cell disruption the cleared lysate was loaded on a 5-mL HisTrap FF column (GE Healthcare) equilibrated in lysis buffer A and the protein was eluted with a gradient of 25 to 250 mM imidazole. CtCyclin H was incubated with 3C protease overnight at 4 °C to remove the Twin-Strep tag. CtCyclin H or MBP-scCTD was concentrated and then loaded on a Superdex 200 16/60 pg column (GE Healthcare) equilibrated in 20 mM Hepes (pH 7.5), 250 mM NaCl, 1 mM TCEP. Fractions containing the target protein were concentrated, flash-frozen in liquid nitrogen, and stored at −80 °C. The coexpressed complex of ctCyclin H and ctMAT1short was loaded on a 5-mL StrepTrap HP column (GE Healthcare) equilibrated in lysis buffer B (20 mM Hepes, pH 7.5, 300 mM NaCl, 1 mM TCEP) as a second affinity purification step and eluted with 2.5 mM desthiobiotin. The complex-containing fractions were concentrated, incubated with 3C protease, and then further processed by size-exclusion chromatography as described above. Purification of the full-length ctCyclin H–ctMAT1 complex was performed as described for ctCyclin H–ctMAT1short except that buffers additionally contained 10% glycerol.

For the expression of ctCDK7, Sf21 cells (*Spodoptera frugiperda*) were transfected with bacmid DNA using the SuperFect (Qiagen) transfection reagent according to the manufacturer's instructions. The baculovirus was amplified in Sf21 cells before infecting Hi5 cells (*Trichplusia ni*) seeded to 0.5×10^6 cells per milliliter with 0.1 volume virus titer for expression of ctCDK7 for 72 h. The cells were harvested and resuspended in lysis buffer B supplemented with 0.5 mM protease inhibitor PMSF and DNaseI. The cleared lysate was further processed by StrepTrap affinity chromatography and size-exclusion chromatography as described above.

To assemble the trimeric ctCAK variants, purified proteins were mixed in a stoichiometric ratio and loaded on an SD200 Increase 10/300 GL column (GE Healthcare) equilibrated in 20 mM Hepes (pH 7.5), 200 mM NaCl, 1 mM TCEP. Appropriate fractions were concentrated, flash-frozen in liquid nitrogen, and stored at −80 °C.

Crystallization, Data Collection, and Structure Determination. Crystallization trials were set up with the Honeybee 963 robot using the sitting-drop vapor-diffusion method. The ctCAKshort complex was crystallized at a protein concentration of 4 mg/mL and with a precipitant containing 100 mM Hepes (pH 7.0 to 8.0), 0.2 M sodium formate, 15 to 22% polyethylene glycol 3350. The crystals were flash-frozen in the precipitant solution containing an additional 25% ethylene glycol as cryoprotectant.

For the apo structure, data were collected at 100 K at the European Synchrotron Radiation Facility (ESRF) beamline ID29. Data were integrated and scaled using XDS (49) and then merged using the STARANISO server to 2.6 Å (Table 1). The structure was solved by molecular replacement using Phaser (50) and a homology model of ctCyclin H and ctCDK7 was obtained from the SWISS-MODEL server (51). Two copies of ctCDK7 and ctCyclin H were found in the asymmetric unit. The missing MAT1 was built manually using Coot (52). The structure was improved by alternating cycles of manual model building and automated refinement. The program BUSTER was used for the initial refinement steps (53) and the final rounds of refinement were performed using the PHENIX suite (54). In all refinement steps the anisotropically corrected data were used to the recommended resolution cutoff of 2.6 Å.

For the ATPγS-bound structure, crystals were soaked for 2 h in the precipitant solution supplemented with 1 mM ATPγS, 5 mM $MgCl_2$, and 1 mM phosphorylated peptide (YphoSPTSPS) which was not resolved in the structure. Diffraction data of the ATPγS-bound crystals were collected at 100 K at the BESSY beamline MX-14-1. Data were integrated, scaled, and merged to 3 Å as described above. The structure was solved by molecular replacement using the structure of the apo ctCAKshort complex as search model. The structure was refined using PHENIX and the anisotropically corrected data to 3 Å. Interface areas in the CAK complex were calculated using PDBePISA (55) from the EMBL-EBI platform. Structural figures were generated using PyMOL.

In Vitro Fluorescence-Based In-Gel Kinase Assay. Kinase activity was assessed using a fluorescence-based in-gel phosphoprotein detection assay after SDS-PAGE of the kinase reaction samples. The reactions were carried out in a solution containing 20 mM Hepes (pH 7.5), 200 mM NaCl, 1 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP, 0.2 μM ctCDK7 or ctCDK7-containing complexes, and 20 μM MBP-scCTD substrate with 26 heptad repeats. Samples were incubated at 37 °C for 1 min and stopped by the addition of SDS sample dye, and then 2 μL of the reaction mix was loaded on a 15% SDS gel. After gel electrophoresis, the gel was stained for phosphoprotein species using the Pro-Q Phosphoprotein Gel Stain Kit according to the manufacturer's instructions (Thermo Fisher Scientific). Signals were detected at an excitation wavelength of 532 nm and an emission wavelength of 605 nm using a PharosFX (Bio-Rad) fluorescence scanner. Signal intensities were quantified using ImageJ (56). All measurements were carried out in technical triplicate and mean values were plotted with their associated SD.

In Vitro Substrate-Dependent ATPase Assay. Kinase activity was determined using an in vitro ATPase assay in which ATP consumption is coupled to the oxidation of NADH via pyruvate kinase and lactate dehydrogenase activities. Kinase activities were assessed at 37 °C, containing 1.5 U pyruvate kinase, 1.9 U lactate dehydrogenase, 2 mM phosphoenolpyruvate, 0.3 mM NADH, 20 mM Hepes (pH 7.5), 200 mM NaCl, 1 mM TCEP, and 5 mM $MgCl_2$. The reactions were performed with 200 nM ctCDK7 or ctCDK7-containing complexes and 20 μM MBP-scCTD as the substrate. The mixture including all above-mentioned components was preincubated at 37 °C until a stable baseline was achieved. The reaction was then started by the addition of ATP at a concentration of 1 mM. Activity profiles were recorded at 340 nm using a FluoStar Optima (BMG Labtech) plate reader until NADH was entirely consumed. The velocity increased during the reaction and the highest velocity toward the end of each reaction was fitted with the MARS software package (BMG Labtech). The rate of ATP consumption was calculated using the molar extinction coefficient of NADH (6,220 $M^{-1} \cdot cm^{-1}$). Measurements were carried out in at least triplicate derived from two biological replicates and mean values were plotted with their associated SD.

Thermal Shift Analysis. Protein complexes were analyzed at a concentration of 2.2 to 2.5 mg/mL in a reaction mix containing 20 mM Hepes (pH 7.5), 200 mM

NaCl, and 1 mM TCEP supplemented with 0.1% Sypro orange (Invitrogen). Unfolding was recorded as an increase in fluorescence at an excitation wavelength of 429 nm and an emission wavelength of 610 nm using a qPCR machine (Stratagene; Mx3005p). Melting temperatures represent the mean values derived from two biological replicates combined in eight technical replicates.

Mass Spectrometry Analysis. The kinase reaction prior to the mass spectrometry analysis was carried out as described for the fluorescence-based kinase assay except that the reactions were incubated for 10 min and 5 μL was loaded on the gel. Gel bands were excised and destained with 30% acetonitrile in 0.1 M NH_4HCO_3 (pH 8), shrunk with 100% acetonitrile, and dried in a vacuum concentrator (concentrator 5301; Eppendorf). Digests were performed with 0.1 μg protease (trypsin, elastase, or chymotrypsin) per gel band overnight at 37 °C in 0.1 M NH_4HCO_3 (pH 8). The digest with papain was performed with 0.1 μg protease in 0.1 M NH_4HCO_3, 5 mM cysteine, 1 M Arg amide (pH 7) for 2 h at 37 °C. Papain in the presence of Arg amide generates peptides with a C-terminal arginine. This improves fragmentation behavior and generates intense y-ion series. After removal of the supernatant, peptides were extracted from the gel slices with 5% formic acid, and the extracted peptides were pooled with the supernatant. Papain digests were cleaned using C-18 Stage Tips.

Nano-liquid chromatography-tandem mass spectrometry (nano-LC-MS/MS) analyses were performed on an Orbitrap Fusion (Thermo Fisher Scientific) equipped with a PicoView Ion Source (New Objective) and coupled to an EASY-nLC 1000 (Thermo Fisher Scientific). Peptides were loaded on capillary columns (PicoFrit, 30 cm × 150 μm interior diameter; New Objective), self-packed with ReproSil-Pur 120 C18-AQ, 1.9 μm (Dr. Maisch), and separated with a 30- or 60-min linear gradient, respectively, from 3 to 40% acetonitrile and 0.1% formic acid and a flow rate of 500 nL/min.

Both MS and MS/MS scans were acquired in the Orbitrap analyzer, with a resolution of 60,000 for MS scans and 15,000 for MS/MS scans. A mixed electron-transfer dissociation (ETD)/higher-energy collisional dissociation (HCD) method was used. HCD fragmentation was applied with 35% normalized collision energy. For ETD, a calibrated charge-dependent ETD parameter was applied. A "top-speed" data-dependent MS/MS method with a fixed cycle time of 3 s was used. Dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 10 s; singly charged precursors were excluded from selection. The minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with AGC at a target value of 2e5 for MS scans and 5e4 for MS/MS scans. EASY-IC was used for internal calibration.

The database search was performed against a custom database (about 500 proteins) containing the protein sequences of interest with the PEAKS X software (BSI; Bioinformatics Solution) with the following parameters: peptide mass tolerance: 10 parts per million; MS/MS mass tolerance: 0.015 Da; enzyme: "none"; variable modifications: acetylation (protein N terminus), oxidation (M), carbamidomethylation (C), pyro-Glu from Q, phosphorylation (STY), Arg amide (C-terminal peptide). Results were filtered to 0.5 or 1% peptide spectrum matches-false discovery rate (PSM-FDR), respectively, by the target-decoy approach. PEAKS search results were exported for Skyline. Skyline 20.1 (https://skyline.ms/project/home/software/Skyline/begin.view) was applied to determine values of the area under the curve from extracted-ion chromatograms of the peptides and phosphopeptides of interest.

Data Availability. The coordinates and structure factors for the apo and ATPγS-bound ctCAK complexes have been deposited in the Protein Data Bank (PDB) under ID codes 6Z3U and 6Z4X, respectively.

All study data are included in the article and *SI Appendix*.

ACKNOWLEDGMENTS. We thank the staff from the beamlines ID29 at the ESRF and MX-14 at BESSY II at the Helmholtz Institute in Berlin for excellent support. We also thank Florian Sauer for providing plasmids used in this study.

1. D. J. Wood, J. A. Endicott, Structural insights into the functional diversity of the CDK-cyclin family. *Open Biol.* **8**, 180112 (2018).
2. T. J. Gibson, J. D. Thompson, A. Blocker, T. Kouzarides, Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107. *Nucleic Acids Res.* **22**, 946–952 (1994).
3. M. Malumbres, Cyclin-dependent kinases. *Genome Biol.* **15**, 122 (2014).
4. K. A. Merrick *et al.*, Distinct activation pathways confer cyclin-binding specificity on Cdk1 and Cdk2 in human cells. *Mol. Cell* **32**, 662–672 (2008).
5. G. Lolli, L. N. Johnson, CAK-cyclin-dependent activating kinase: A key kinase in cell cycle control and a target for drugs? *Cell Cycle* **4**, 572–577 (2005).
6. A. Devault *et al.*, MAT1 ('menage à trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and *Xenopus* CAK. *EMBO J.* **14**, 5027–5036 (1995).
7. R. P. Fisher, P. Jin, H. M. Chamberlin, D. O. Morgan, Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. *Cell* **83**, 47–57 (1995).
8. J. P. Tassan *et al.*, In vitro assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. *EMBO J.* **14**, 5608–5617 (1995).
9. M. C. Keogh, E. J. Cho, V. Podolny, S. Buratowski, Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. *Mol. Cell. Biol.* **22**, 1288–1297 (2002).
10. D. Busso *et al.*, Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. *J. Biol. Chem.* **275**, 22815–22823 (2000).
11. W. Abdulrahman *et al.*, ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E633–E642 (2013).
12. S. Akoulitchev, T. P. Mäkelä, R. A. Weinberg, D. Reinberg, Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature* **377**, 557–560 (1995).
13. W. J. Feaver, J. Q. Svejstrup, N. L. Henry, R. D. Kornberg, Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* **79**, 1103–1109 (1994).
14. R. Shiekhattar *et al.*, Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature* **374**, 283–287 (1995).
15. R. Roy *et al.*, The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* **79**, 1093–1101 (1994).
16. P. Komarnitsky, E. J. Cho, S. Buratowski, Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460 (2000).
17. M. S. Akhtar *et al.*, TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393 (2009).
18. K. Glover-Cutter *et al.*, TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol. Cell. Biol.* **29**, 5455–5464 (2009).
19. M. Kim, H. Suh, E. J. Cho, S. Buratowski, Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J. Biol. Chem.* **284**, 26421–26426 (2009).
20. T. M. Søgaard, J. Q. Svejstrup, Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with Mediator. *J. Biol. Chem.* **282**, 14113–14120 (2007).
21. L. C. Myers *et al.*, The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45–54 (1998).
22. S. Larochelle *et al.*, T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity. *EMBO J.* **20**, 3749–3759 (2001).
23. C. C. Ebmeier *et al.*, Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. *Cell Rep.* **20**, 1173–1186 (2017).
24. M. Boehning *et al.*, RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
25. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
26. J. Bai, Y. Li, G. Zhang, Cell cycle regulation and anticancer drug discovery. *Cancer Biol. Med.* **14**, 348–362 (2017).
27. C. Sánchez-Martínez, M. J. Lallena, S. G. Sanfeliciano, A. de Dios, Cyclin dependent kinase (CDK) inhibitors as anticancer drugs: Recent advances (2015-2019). *Bioorg. Med. Chem. Lett.* **29**, 126637 (2019).
28. N. Kwiatkowski *et al.*, Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).
29. S. Hu *et al.*, Discovery and characterization of SY-1365, a selective, covalent inhibitor of CDK7. *Cancer Res.* **79**, 3479–3491 (2019).
30. Y. Wang *et al.*, CDK7-dependent transcriptional addiction in triple-negative breast cancer. *Cell* **163**, 174–186 (2015).
31. B. Li *et al.*, Therapeutic rationale to target highly expressed CDK7 conferring poor outcomes in triple-negative breast cancer. *Cancer Res.* **77**, 3834–3845 (2017).
32. K. K. Kim, H. M. Chamberlin, D. O. Morgan, S. H. Kim, Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase. *Nat. Struct. Biol.* **3**, 849–855 (1996).
33. G. Andersen *et al.*, The structure of cyclin H: Common mode of kinase activation and specific features. *EMBO J.* **16**, 958–967 (1997).
34. G. Lolli, E. D. Lowe, N. R. Brown, L. N. Johnson, The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12**, 2067–2079 (2004).
35. L. M. Stevenson-Lindert, P. Fowler, J. Lew, Substrate specificity of CDK2-cyclin A. What is optimal? *J. Biol. Chem.* **278**, 50956–50960 (2003).
36. K. Y. Cheng *et al.*, The role of the phospho-CDK2/cyclin A recruitment site in substrate recognition. *J. Biol. Chem.* **281**, 23167–23179 (2006).
37. S. Baumli, A. J. Hole, L. Z. Wang, M. E. Noble, J. A. Endicott, The CDK9 tail determines the reaction pathway of positive transcription elongation factor b. *Structure* **20**, 1788–1795 (2012).
38. A. A. Russo, P. D. Jeffrey, N. P. Pavletich, Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol.* **3**, 696–700 (1996).
39. S. Baumli *et al.*, The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J.* **27**, 1907–1918 (2008).

40. E. V. Schneider *et al*., The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder. *J. Mol. Biol.* **412**, 251–266 (2011).
41. S. Akoulitchev, S. Chuikov, D. Reinberg, TFIIH is negatively regulated by cdk8-containing Mediator complexes. *Nature* **407**, 102–106 (2000).
42. N. R. Brown, M. E. Noble, J. A. Endicott, L. N. Johnson, The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat. Cell Biol.* **1**, 438–443 (1999).
43. K. Huang *et al*., Structure of the Pho85-Pho80 CDK-cyclin complex of the phosphate-responsive signal transduction pathway. *Mol. Cell* **28**, 614–623 (2007).
44. D. A. McGrath *et al*., Structural basis of divergent cyclin-dependent kinase activation by Spy1/RINGO proteins. *EMBO J.* **36**, 2251–2262 (2017).
45. N. R. Brown *et al*., CDK1 structures reveal conserved and unique features of the essential cell cycle CDK. *Nat. Commun.* **6**, 6769 (2015).
46. D. Eick, M. Geyer, The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).
47. M. Z. Li, S. J. Elledge, SLIC: A method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
48. J. Yang, Y. Zhang, I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).
49. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
50. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. A. Waterhouse *et al*., SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
52. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
53. G. B. E. Bricogne *et al*., *BUSTER Version 2.10.3*, (Global Phasing, Cambridge, UK, 2017).
54. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
55. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
56. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).