Structural and dynamical insights into the PH domain of p62 in human TFIIH

Masahiko Okuda[1], Toru Ekimoto[1], Jun-ichi Kurita[1], Mitsunori Ikeguchi[1,2] and Yoshifumi Nishimura [1,3,*]

[1]Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan, [2]RIKEN Medical Sciences Innovation Hub Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan and [3]Graduate School of Integrated Sciences for Life, Hiroshima University, 1-4-4 Kagamiyama, Higashi-Hiroshima 739-8258, Japan

ABSTRACT

TFIIH is a crucial transcription and DNA repair factor consisting of the seven-subunit core. The core subunit p62 contains a pleckstrin homology domain (PH-D), which is essential for locating TFIIH at transcription initiation and DNA damage sites, and two BSD (BTF2-like transcription factors, synapse-associated proteins and DOS2-like proteins) domains. A recent cryo-electron microscopy (cryo-EM) structure of human TFIIH visualized most parts of core, except for the PH-D. Here, by nuclear magnetic resonance spectroscopy we have established the solution structure of human p62 PH-D connected to the BSD1 domain by a highly flexible linker, suggesting the flexibility of PH-D in TFIIH. Based on this dynamic character, the PH-D was modeled in the cryo-EM structure to obtain the whole human TFIIH core structure, which indicates that the PH-D moves around the surface of core with a specific but limited spatial distribution; these dynamic structures were refined by molecular dynamics (MD) simulations. Furthermore, we built models, also refined by MD simulations, of TFIIH in complex with five p62-binding partners, including transcription factors TFIIEα, p53 and DP1, and nucleotide excision repair factors XPC and UVSSA. The models explain why the PH-D is crucially targeted by these factors, which use their intrinsically disordered acidic regions for TFIIH recruitment.

INTRODUCTION

The general transcription factor TFIIH is a multifunctional protein complex involved in transcriptional initiation (1–3), activation (4–13) and elongation (14–16), the cell cycle (17–20) and nucleotide excision repair (21–24). TFIIH consists of 10 different subunits (~500 kDa) and can be divided into two sub-complexes: the core TFIIH complex consisting of 7 subunits (XPB, XPD, p62, p52, p44, p34 and p8) and the Cdk-activating kinase (CAK) complex formed by 3 subunits (Cdk7, Cyclin H and MAT1). TFIIH possesses several enzymatic activities: for example, the ATPase/helicase activities of XPB and XPD, both of which are required for promoter melting at transcription initiation sites and damaged DNA opening at nucleotide excision repair sites; the ubiquitin ligase activity of p44 (in yeast), which is thought to mediate the transcriptional response to DNA damage; and the protein kinase activity of Cdk7, which is necessary for phosphorylation of the C-terminal domain of the largest subunit of RNA polymerase II, other transcription factors and nuclear receptors (25,26). In particular, mutations in XPB, XPD and p8 cause three autosomal recessive disorders: xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy (27).

p62 is an essential subunit of TFIIH, not only for formation of the core TFIIH complex but also for recruiting TFIIH to appropriate functional sites. Human p62 has four structural domains: an N-terminal pleckstrin homology (PH) domain (PH-D) (residues 1–103); two BSD (BTF2-like transcription factors, synapse-associated proteins and DOS2-like proteins) domains, BSD1 (residues 109–147) and BSD2 (residues 189–225); and a C-terminal three-helix bundle (residues 454–548). The PH-D is responsible for recruitment of TFIIH through its specific interaction with intrinsically disordered acidic regions of various transcription factors and several nucleotide excision repair factors, such as TFIIEα (28,29), p53 (30,31), EKLF (32), DP1 (33), VP16 (34,35) and EBNA2 (36) during transcription, and XPC (37,38), UVSSA (39) and XPG (40) during nucleotide excision repair. Recent cryo-electron microscopy (cryo-EM) studies of human apo TFIIH (41) and yeast holo TFIIH in complex with the pre-initiation complex (PIC) and Mediator (42) have provided detailed pictures of the core TFIIH complex and how it is connected to the CAK complex. In the cryo-EM structure of human TFIIH, al-

---

*To whom correspondence should be addressed. Tel: +81 45 508 7211; Fax: +81 45 508 7360; Email: nisimura@yokohama-cu.ac.jp

most all parts of the seven subunits in the core can be visualized except for the N-terminal PH-D of p62.

To gain insight into the structure of the p62 PH-D in the core TFIIH complex, here, keeping in mind the fact that the BSD1 domain can be clearly identified in the cryo-EM structure, we have solved the solution structure of the tandem PH and BSD1 domains of human p62 (residues 1–158) by nuclear magnetic resonance (NMR) spectroscopy. These domains are connected by a short linker (residues 104–108) and do not interact with each other. The flexibility of the interdomain region has been evaluated by measuring ^{15}N relaxation parameters and ^{1}H exchange rates with solvent. Based on our dynamic NMR structures of the tandem PH and BSD1 domains coupled with the cryo-EM structure (43), we have built reasonable dynamic structural models of human TFIIH including the flexible p62 PH-D by superposing the BSD1 domains from the solution and cryo-EM structures; we have further refined these dynamic structures by using all-atom molecular dynamics (MD) simulations. In addition, we have modeled the structures, also refined by the MD simulations, of TFIIH in complex with five p62-binding partners that are responsible for recruiting TFIIH to sites of transcription initiation and DNA damage.

MATERIALS AND METHODS

Preparation of human p62$_{1–158}$

The ^{13}C/^{15}N-labeled fragment of human p62$_{1–158}$ (residues 1–158) was expressed as a hexa-histidine-tagged product in a pET15b vector (Merck Millipore) in *Escherichia coli* BL21 (DE3) Gold (Agilent Technologies). The cells were grown at 37°C in M9 minimal medium containing [^{15}N]ammonium chloride and [^{13}C]glucose. The product was expressed by induction with 1 mM isopropyl-β-D-thiogalactopyranoside. After 21 h of growth at 20°C, the cells were collected, resuspended in buffer A [20 mM Tris–HCl (pH 8.0), 10% glycerol, 1 M NaCl], lysed by sonication and then centrifuged. The supernatant was loaded onto a His-Tag Purification Resin (Roche) column equilibrated with buffer A, and the column was washed with buffer A containing 10 mM imidazole–HCl. The sample was eluted by 500 mM imidazole–HCl, peak fractions were pooled and the buffer was changed to thrombin cleavage buffer (10 mM Na$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$, 500 mM NaCl, 2.7 mM KCl, pH 7.3) by using an Amicon Ultra device (Merck Millipore). To remove the histidine tag, the sample was digested with thrombin for 63 h at 20°C. After concentration via an Amicon Ultra, the sample was applied to a Superdex75 (GE Healthcare) column equilibrated with 20 mM potassium phosphate (pH 6.8) and 500 mM NaCl.

NMR structure determination

For structure determination, we used 0.9–1.2 mM ^{13}C/^{15}N-labeled p62$_{1–158}$ in 20 mM potassium phosphate (pH 6.8) and 5 mM deuterated DTT prepared in either 90% H$_2$O/10% D$_2$O or 99.9% D$_2$O. NMR experiments were performed at 20°C on Bruker AVANCE III HD 600-MHz and 950-MHz spectrometers, each equipped with a cryo-TCI probe. Backbone and side-chain resonances were assigned by using standard triple-resonance NMR experiments (44).

Stereospecific assignments were obtained from a combination of HNHB, HN(CO)HB, HNCG, HN(CO)CG, and ^{13}C-edited and ^{15}N-edited NOESY-HSQC spectra. Distance restraints were obtained from ^{15}N-edited NOESY-HSQC (τ_m = 150 ms) and ^{13}C-edited NOESY-HSQC (τ_m = 50 and 100 ms) spectra. Side-chain torsion angles, χ_1 and χ_2, were obtained from a combination of HNHB, HN(CO)HB, HNCG, HN(CO)CG, and ^{13}C-edited and ^{15}N-edited NOESY-HSQC spectra. Hydrogen bond restraints were obtained by backbone amide H/D-exchange experiments. Spectra were processed by using NMRPipe (45) and analyzed by using NMRView (46).

Structure calculation

Interproton distance restraints derived from NOE (nuclear Overhauser effect) intensities were grouped into four distance ranges, 1.8–2.7 Å (1.8–2.9 Å for NOEs involving HN protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving HN protons), 1.8–5.0 and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak NOEs, respectively. The upper limit was corrected for constraints involving methyl groups, aromatic ring protons and non-stereospecifically assigned methylene protons. Dihedral angle restraints for φ and ψ were obtained from analysis of the backbone chemical shifts with TALOS+ (47). χ_1 and χ_2 angles were restrained ±30° for three side-chain rotamers. Structure calculations were performed by distance geometry and simulated annealing using the program Xplor-NIH (48,49). In total, 100 structures were calculated. All structures were then subjected to water refinement (50), in which the structures were immersed in a 7.0-Å layer of water molecules. After minimization with 120 steps, a heating stage from 100 to 500 K with 200 steps of MD for every 100-K increment, a refinement stage with 2500 steps at 500 K and a cooling stage from 500 to 25 K with 200 steps for every 25-K decrement were carried out. The refinement protocol was finished with 200 steps of minimization. Structural statistics for the 20 best structures are summarized in Table 1. Structures were analyzed and displayed by using PROCHECK-NMR (51), MOLMOL (52) and PyMol (http://www.pymol.org).

NMR relaxation analysis

The backbone ^{15}N relaxation parameters were measured by using a ^{13}C/^{15}N-labeled sample (0.60 mM) at 20°C on Bruker AVANCE III HD 600-MHz and 950-MHz spectrometers equipped with a cryo-TCI probe. Relaxation durations of 24, 49, 64, 128, 256, 384, 512, 768, 1024, 1280, 1536, 1792 and 2048 ms for the longitudinal relaxation rate (R_1) and 16.98, 33.95, 50.93, 67.90, 84.88, 101.86, 135.81, 203.71, 271.62, 339.52, 441.38 and 543.23 ms for the transverse relaxation rate (R_2) were used. Heteronuclear ^{15}N–{^{1}H} NOE experiments were recorded in the presence and absence of proton saturation, which was achieved with a 5.0-s duration consisting of 120° ^{1}H pulses applied every 5.0 ms. R_1 and R_2 were obtained by fitting peak intensities at a series of relaxation durations to an exponential decay curve using CurveFit (53). The uncertainties of the peak intensities were estimated by using duplicated data from the shortest relaxation delay. The uncertainties of the relaxation

Table 1. Statistics for the 20 best structures of the PH and BSD1 domains of TFIIH p62

Experimental restraints	
Total NOE	3338
Intra-residue	465
Sequential ($i − j = 1$)	799
Medium range ($1 < i − j < 5$)	709
Long range ($i − j \geq 5$)	1365
Hydrogen bond	48×2
Number of dihedral restraints	
φ	147
ψ	146
χ_1	75
χ_2	13
Statistics for structure calculations	
Root-mean-square deviation (RMSD) from experimental restraints[a]	
Distance (Å)	0.038 ± 0.001
Dihedral (°)	0.384 ± 0.070
RMSD from idealized covalent geometry	
Bonds (Å)	0.00513 ± 0.00009
Angles (°)	0.642 ± 0.011
Improper (°)	0.649 ± 0.021
Coordinate precision	
Average pairwise RMSD from the mean structure	
Backbone atoms (Å)	0.46 ± 0.10,[b] 0.76 ± 0.34,[c] 0.45 ± 0.12[d]
Heavy atoms (Å)	1.22 ± 0.18,[b] 1.49 ± 0.35,[c] 1.12 ± 0.19[d]
Ramachandran plot statistics	
Residues in most favored regions (%)	84.0[e]
Residues in additional allowed regions (%)	15.8[e]
Residues in generously allowed regions (%)	0.2[e]
Residues in disallowed regions (%)	0.0[e]
Clash score	26

[a]None of the structures exhibited distance violations >0.5 Å and dihedral angle violations >5°.
[b]The value was calculated over residues 7–103.
[c]The value was calculated over residues 109–147.
[d]The value was calculated over residues 114–147.
[e]The value was calculated over residues 5–154.

rates were determined by CurveFit using a Monte Carlo simulation. The steady-state ^{15}N–{^{1}H} NOE values were determined from peak intensity ratios obtained from spectra acquired with and without proton saturation. The uncertainties were determined from the standard deviation in background noise levels by using NMRView (46).

HET$_{ex}$-BEST-TROSY experiment

HET$_{ex}$-BEST-TROSY (54) experiments were performed at 20°C on a Bruker AVANCE III HD 950-MHz spectrometer equipped with a cryo-TCI probe. A series of spectra was collected with relaxation delays of 339, 709, 1080, 1450 and 2006 ms with and without water saturation. The water signal intensity was measured at each relaxation delay to determine the polarization of water using the ^{1}H–^{15}N BEST-TROSY pulse sequence with an additional small flip angle reading pulse (0.1 μs). The hydrogen exchange rate k_{ex} was determined by simple fitting using Equations (1) and (2) for the amide proton polarizations:

$$A_z^{\text{ref}}(d_{\text{relax}}) = A_z^{\text{eq}} \frac{R_1^A + k_{ex} W_z^{\text{ref}}(d_{\text{relax}})}{R_1^A + k_{ex}} \left[1 - \exp\left\{ -\left(R_1^A + k_{ex} \right) d_{\text{relax}} \right\} \right], \quad (1)$$

$$A_z^{\text{sat}}(d_{\text{relax}}) = A_z^{\text{eq}} \frac{R_1^A}{R_1^A + k_{ex}} \left[1 - \exp\left\{ -\left(R_1^A + k_{ex} \right) d_{\text{relax}} \right\} \right], \quad (2)$$

where R_1^A is the longitudinal relaxation rate of amide protons, A_z^{eq} is the equilibrium amide proton polarization and

W_z^{ref} is the water polarization. All NMR spectra were processed by using the program NMRPipe (45) and analyzed by using the program NMRView (46). Least-squares fitting of the data was performed by using Gnuplot software (http://www.gnuplot.info).

CLEANEX-PM experiment

CLEANEX-PM (phase-modulated CLEAN chemical exchange spectroscopy) (55) experiments were performed at 20°C on a Bruker AVANCE III HD 950-MHz spectrometer equipped with a cryo-TCI probe using a standard Bruker pulse program. A series of spectra was collected with mixing times for the CLEANEX spin lock of 7.5, 15, 30, 50 and 80 ms. The hydrogen exchange rate k_{ex} was determined by using established methods. All NMR spectra were processed by using the program NMRPipe (45) and analyzed by using the program NMRView (46). Least-squares fitting of the data was performed by using Gnuplot software (http://www.gnuplot.info).

Structural modeling of human TFIIH

Initially, 1000 structures of p62$_{1–158}$ were calculated as described in the 'Structure calculation' section, and then 399 structures were selected on the basis of violation criteria (distance violations <0.5 Å and dihedral angle violations

<5°). These 399 structures were aligned to the BSD1 domain region (residues 107–147) of p62 in the cryo-EM structure of human apo TFIIH (PDB code 6NMI) by using the program PyMol (http://www.pymol.org). The BSD1 domain region (residues 107–147) of p62 in the cryo-EM structure was then replaced with the PH-D and BSD1 domain (residues 1–147) of p62 in the 399 structures of $p62_{1–158}$. Finally, undesirable models in which the PH-D of p62 clashed against other parts of TFIIH were removed, resulting in 275 model structures.

Docking structural modeling of human TFIIH and its p62-binding partners

To build structural models of TFIIH docking with its p62-binding partners, TFIIEα (29), p53 (31), DP1 (33), XPC (38) and UVSSA (39), five models in which the PH-D of p62 was located in free space were selected from the 275 structural models of TFIIH (Figure 5A). Structure calculation was done by the program Xplor-NIH (48,49) using the scripts 'rigid_min.inp' and 'sa_cross_tor.inp'. We employed experimentally obtained intermolecular distance restraints in each complex of the $p62_{1–108}$ domain with TFIIEα (PDB code 2RNR), p53 (PDB code 2RUK), DP1 (PDB code 5GOW), XPC (PDB code 2RVB) and UVSSA (PDB code 5XV8).

MD simulations of human TFIIH and its complexes with p62-binding partners

The initial structure used for MD simulations was prepared from the whole apo human TFIIH core complex, in which the NMR structure was used for the PH-D and BSD1 domain (residues 1–147) of p62 and the cryo-EM structure (PDB code 6NMI) was used for the other parts of TFIIH. The 275 model structures described above were classified in terms of the position of the PH-D, and 6 representative model structures (Figure 5B, D, F, H, J and L) were subjected to MD simulations. First, in a typical structure (Figure 5L), the missing regions and side chains in the cryo-EM structure, as well as the region joining the NMR and cryo-EM structures of p62 (residues 145–147), were built by MODELLER (56). The long loops at the termini (residues 1–31 and 731–782 of XPB, residues 1–14 of p44, residues 294–308 of p34, residues 69–71 of p8 and residues 213–309 of MAT1) were not modeled. Next, using the model structure in Figure 5L, the other five whole TFIIH structures (Figure 5B, D, F, H and J) and five TFIIH complexes (Figure 6) were built: the whole TFIIH structures were aligned to the model structure by using a helix of the three-helix bundle of p62 (residues 487–511), and the PH-D and BSD1 domain (residues 1–143) of each structure were replaced with the same domains of the model structure.

The systems for MD simulations of the model TFIIH structures were prepared by using the psfgen, solvate and autoionize plugins implemented in VMD (57). The protonation states of His, Asp, Glu and Lys were determined by using pK_a values calculated with PROPKA (58,59) implemented in PDB2PQR (60) under the pH 7 condition. Missing hydrogen atoms were inserted in accordance with topology in the CHARMM36m force field (61–63). The

N- and C-termini were set to NH_3^+ and COO^-, respectively. The coordinate bonds among anionic cysteine or histidine residues and zinc ions were incorporated, as defined in the CHARMM36m force field, and the oxidized iron–sulfur cluster and the coordinate bonds of anionic cysteine residues with the iron–sulfur cluster were also incorporated by using the topologies and parameters determined by Chang and Kim (64). Here, both phosphorylated serine and threonine in the p62-binding partner p53 were set to the dianionic state. The MD unit cell was set to a rectangular cell, and the cell size was determined such that the distance between the edge of the cell and the centered model structure was at least 15 Å. Using the TIP3P water model (65), the cell was filled with water molecules, and counterions (Na^+) and 150 mM NaCl were added. The final systems consisted of ~504 000 atoms.

The all-atom MD simulations were performed by using the MD program package GROMACS version 2019.6 (66–68) with the CHARMM36m force field under periodic boundary conditions. The topology and parameter files were converted into the GROMACS format by using ParmEd (https://github.com/ParmEd/ParmEd). Electrostatic interactions were handled by the smooth particle mesh Ewald method (69), and the van der Waals interactions were truncated by the switching function with a range of 10–12 Å. Bond lengths involving hydrogen atoms were constrained by the P-LINKS algorithm (70). In all simulations, position constraints at the Cα atoms with a 60 kJ mol^{-2} nm^{-2} force constant were applied to the TFIIH structure, except for the PH-D, the interdomain linker and the N-terminal part of the BSD1 domain (residues 1–114) of p62, the modeled loops that were originally missing in the cryo-EM structure, and the p62-binding partners. The temperature and pressure were 298 K and 1 atm, respectively. Before the production run, energy minimization by the steepest descent method and three equilibration runs were performed sequentially: the first 125-ps run with *NVT* ensemble and a 1-fs timestep, and the second and third 100-ps runs with *NPT* ensemble and a 2-fs timestep. The thermostat in the first and third equilibration runs and the production run was the Nosé–Hoover scheme (71,72), and that in the second equilibration run was the Berendsen algorithm (73). The barostat in the second equilibration run was the Berendsen algorithm (73), and that in the third equilibration run and the production run was the Parrinello–Rahman approach (74,75). After the equilibration runs, a 100-ns production run was carried out with *NPT* ensemble and a 2-fs timestep. These MD simulations were performed for each initial model structure of the six whole TFIIH models and the five complex models with p62-binding partners (1.1 μs in aggregation of the production runs).

The root-mean-square fluctuation (RMSF) of Cα atoms, the solvent-accessible surface area (SASA) of the amide hydrogen and the protein secondary structures were calculated by using each 100-ns trajectory. Snapshots were saved every 1 ns (100 snapshots in a trajectory). As an estimate of the RMSF, an averaged structure was obtained by structural alignment of the snapshots using Cα atoms in the PH-D except for the N-terminus (residues 9–100). The RMSF values of the six apo TFIIH simulations or the five complex simulations were then averaged. The SASA of the amide hydro-

gen was calculated by the GROMACS module. The average SASA of each residue was obtained from each trajectory, and then the averaged SASA was calculated from the simulations for each apo TFIIH or each complex. Protein secondary structures were determined by using the DSSP program (76) implemented in CPPTRJ (77). The population of snapshots forming the helix or 3_{10} helix structure was obtained for each trajectory, and its averaged value was calculated.

NMR chemical shift perturbation experiments

Samples of p62-binding partners, including TFIIEα (residues 378–439) (29), Ser46- and Thr55-phosphorylated p53 (residues 41–62) (31), DP1 (residues 392–410) (33), XPC (residues 124–142) (38) and UVSSA (residues 399–419) (39), were prepared as previously described. Each of the unlabeled samples was added to 0.1 mM ^{13}C/^{15}N-labeled p62$_{1–158}$ at a molar ratio of 1:1 except for DP1 (molar ratio 1:4) in 20 mM potassium phosphate (pH 6.8), 100 mM sodium chloride, 5 mM deuterated DTT and 10% D$_2$O. ^{1}H–^{15}N HSQC spectra were acquired before and after the addition of unlabeled samples at 20°C on a Bruker AVANCE III HD 600-MHz spectrometer equipped with a cryo-TCI probe. Spectra were processed by NMRPipe (45) and analyzed by NMRView (46).

RESULTS AND DISCUSSION

Invisible PH-D of p62 in the cryo-EM structure of human TFIIH

Understanding the structure and dynamics of the PH-D of p62 in TFIIH is indispensable for elucidating the mechanism of recruitment of human TFIIH. At present, high-resolution cryo-EM structures of human and yeast TFIIH are available. The structure of human TFIIH has been resolved for an apo form (41,43), while that of yeast TFIIH has been resolved for a holo form in the transcription PIC (42). The corresponding structures of each subunit and the overall architecture of the core TFIIH are relatively similar, except for the corresponding PH-Ds of the yeast homolog Tfb1 (*Saccharomyces cerevisiae*) and human p62: the Tfb1 PH-D is clearly visible, whereas the human p62 PH-D is invisible (Supplementary Figure S1). Recently, an interesting structural study of human PIC has been reported in which all available cryo-EM data have been integrated to build a structural model of PIC in which the PH-D of p62 is visible (78).

Here, we initially compared the amino acid sequence of human p62 with that of its well-studied homolog Tfb1 (Figure 1A and Supplementary Figure S2). The sequence alignment identified a marked difference in the region between the PH and BSD1 domains. In human p62, the PH-D is immediately followed by the BSD1 domain; in yeast Tfb1, in contrast, a linker of ∼50 residues is present between these domains. This substantial sequence dissimilarity is reflected in their cryo-EM structures (Figure 1B). In both p62 and Tfb1, the BSD1 and BSD2 domains are aligned at nearly the same position with a beads-on-a-string-like topology, and contribute to stabilizing the core architecture by interacting with other four subunits: XPD, p52, p44 and p34 for p62; and Rad3, Tfb2, Ssl1 and Tfb4 for Tfb1. The PH-D of Tfb1

is also visualized in the yeast holo TFIIH structure (Figure 1B, right). Its PH-D lies apart from the BSD1 domain owing to its long linker, and interacts with another general transcription factor, Tfa1/TFIIEα (not shown in Figure 1B). In contrast, the PH-D of p62 is disordered and invisible in the human apo TFIIH structure (Figure 1B, left), suggesting that, despite the short interdomain linker, this domain is surface exposed and behaves flexibly in the absence of proteins other than TFIIH. The possible state of p62 in human holo TFIIH is discussed in detail below.

To explore the predicted flexible behavior of the PH-D of p62 in human apo TFIIH, we applied solution NMR because it is a powerful tool for investigating both the dynamics and structure of proteins in solution. Although NMR is not advantageous for investigating large proteins like TFIIH, the PH-D of p62 is likely to be free in apo TFIIH; we considered therefore that we would not need to target the whole TFIIH protein for our purpose. In the cryo-EM structure, the p62 subunit is visualized from Ala107 onward, just after the PH-D and before the BSD1 domain (Figure 1B, left). Considering this, we focused on the region including both the PH and BSD1 domains (residues 1–158, hereafter designated p62$_{1–158}$) (Figure 2A), and analyzed the secondary structure of p62$_{1–158}$ by NMR.

The p62$_{1–158}$ construct exhibited well-dispersed signals on a ^{1}H–^{15}N-HSQC spectrum (Figure 2B). Almost all backbone and side-chain signals were assigned without using any deuterated samples. The chemical shift data (CSI) (79) of the backbone atoms indicated that there were seven β-strands and one α-helix in the PH-D and three α-helices in the BSD1 domain, consistent with the structures of their isolated forms (40). No secondary structure elements were found in the interdomain linker region between the α1 and α2 helices (Figure 2C). Based on the chemical shift data, it is predicted (RCI) (80) that the interdomain linker region is flexible (Figure 2D).

Structure of the tandem PH and BSD1 domains of human p62

Next, we determined the tertiary structure of p62$_{1–158}$ by using NMR (Figure 3). In total, 3338 NOE-derived distance restraints, 96 hydrogen bond restraints and 381 dihedral angle restraints were collected and used to define the structure (Figure 3A–D and Table 1). Although the RMSD of the backbone atoms for residues 7–147 in the 20 structures was as high as 8.81 ± 2.55 Å, each domain structure was well determined with RMSD values of 0.46 ± 0.10 Å for residues 7–103 (PH-D) and 0.76 ± 0.34 Å for residues 109–147 (BSD1), suggesting that the interdomain linker (104–108) has high mobility (Table 1). In the PH-D, the seven β-strands (β1–β7) form a β-sandwich, in which strands β1–β4 and β5–β7 form the first and second antiparallel β-sheets, respectively. These two β-sheets are linked by a contact between strands β1 and β7, and by a 3_{10} helix connecting strands β4 and β5 (Figure 3E). The following α-helix (α1) is inserted between the two β-sheets and extensively interacts with residues in both, thereby stabilizing the β-sandwich. Pro101 is likely to participate in the α-helix because Phe103 stabilizes the helix by inserting into the hydrophobic core. The structure of the PH-D in p62$_{1–158}$ is essentially identical to that in the isolated form (40). Likewise, the structure
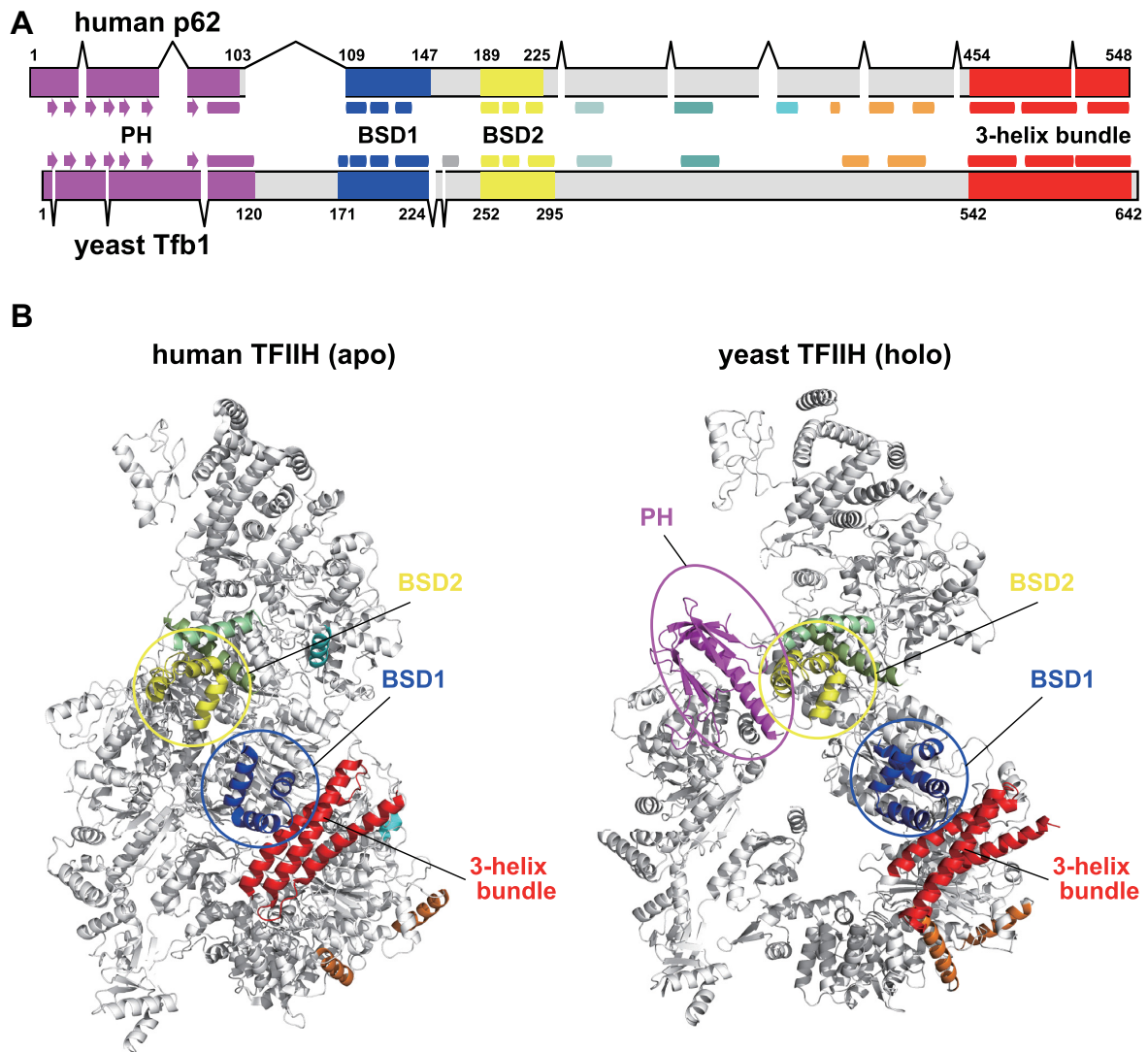
Figure 1. Structure comparison between human and yeast TFIIH p62/Tfb1. (**A**) Domain organization of TFIIH p62/Tfb1. The figure is based on sequence alignment of the two proteins (Supplementary Figure S2). Arrow, β-strand; rectangle, α-helix. (**B**) Ribbon representation of cryo-EM TFIIH structures. Left, human TFIIH, apo form (PDB code 6NMI); right, yeast TFIIH, holo form (PDB code 5OQM). For clarity, only the p62/Tfb1 domain is colored.

of the BSD1 domain in $p62_{1–158}$ is essentially the same as that in the isolated form, which has been determined by another research group (PDB code 2DII) although the study has not been published.

Here, we describe the structure in a little more detail. The first helix (α2) of the BSD1 domain bends nearly at a right angle to the second helix (α3) at Pro122 (Figure 3E). After the turn (Val133–Ser138), the third helix (α4) runs antiparallel to helix α3 and perpendicularly to helix α2. These three helices interact with one another, rendering the BSD1 domain stable and compact. Lys114, Met117 and Leu118 in α2, Val123, Leu124, Phe125, Leu127, Tyr128, Leu131 and Val132 in α3, and Ala139, Phe142, Trp143, Arg146 and Leu147 in α4 form the hydrophobic core. The turn between helices α3 and α4, where Val136 and Ile137 closely interact with Leu131, looks rigid. Ile137 also contacts Val132 and Phe142, participating in the formation of the hydrophobic core.

Dynamics of the tandem PH and BSD1 domains of human p62

To investigate the suggested high mobility of the interdomain linker quantitatively, we measured the ^{15}N relaxation rates R_1 and R_2 (Figure 4A–C and Supplementary Figure S3) and $^{15}N–\{^1H\}$ NOE for $p62_{1–158}$ by NMR (Figure 4D and Supplementary Figure S3). These relaxation data define how much each residue in the interdomain linker contributes to the mobility. Lys104 and Arg105 clearly displayed higher mobility as compared with the residues before the linker. Lys104 and Arg105 exhibited, respectively, R_2/R_1 values of 24.1 and 27.5, and $^{15}N–\{^1H\}$ NOE values of 0.556 and 0.502. These values were close to those of 21.7 and 25.6 (R_2/R_1), and 0.562 and 0.605 ($^{15}N–\{^1H\}$ NOE) observed for Glu110 and Leu111, the N-terminal residues in the first helix (α2) of the BSD1 domain. The following four residues, Lys106, Ala107, Asn108 and Lys109, displayed even higher mobility with R_2/R_1 values of 17.9, 12.9,
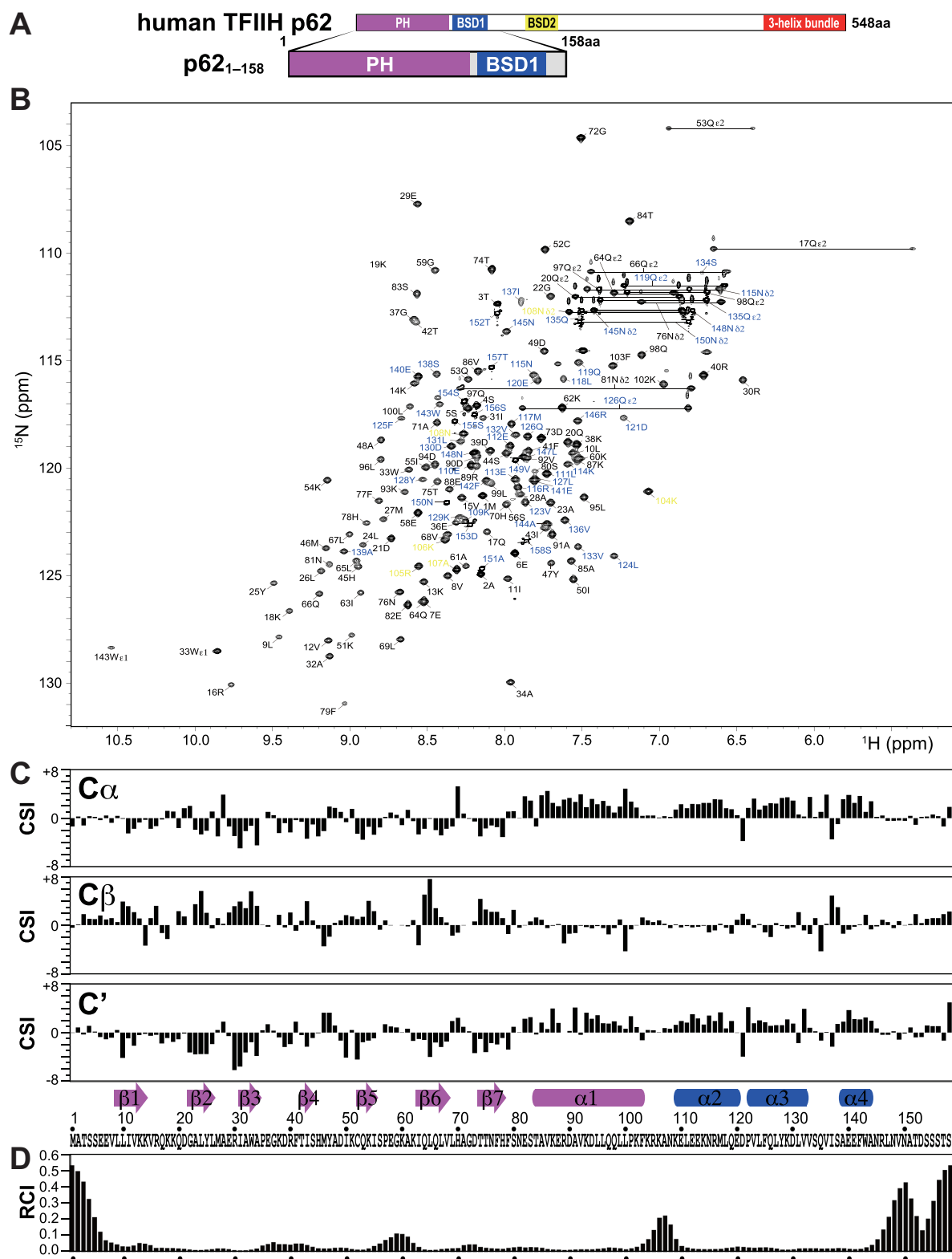
Figure 2. NMR signal assignment and secondary structure of the PH and BSD1 domains of human TFIIH p62. (**A**) p62_{1-158} construct used in the NMR experiments. (**B**) ^{1}H–^{15}N-HSQC spectrum. Residues 1–103, 104–108 and 109–158 are labeled in black, yellow and blue, respectively. (**C**) Chemical shift index (CSI) of ^{13}Cα, ^{13}Cβ and ^{13}C′, and identified secondary structure elements. Arrow: β-strand; rectangle; α-helix. The secondary structure identification was performed using the CSI 2.0 (http://csi.wishartlab.com) web server (79). (**D**) Random coil index (RCI), calculated by using the difference between observed and reference random coil shifts of ^{13}Cα, ^{13}Cβ, ^{13}C′, ^{15}N and ^{1}H. RCI should be proportional to the flexibility of the protein backbone (80).
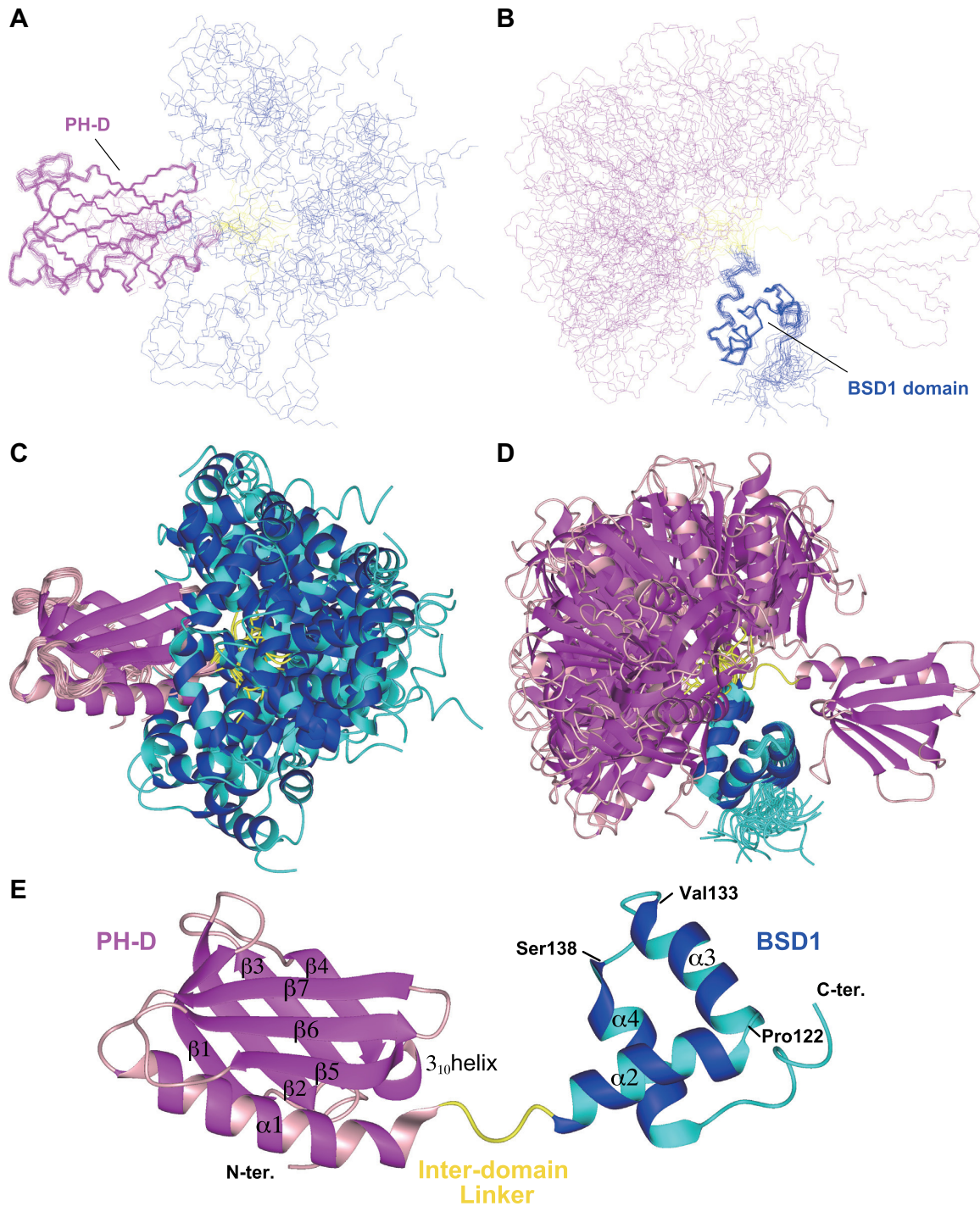
Figure 3. Solution structure of the PH and BSD1 domains of human TFIIH p62. Overlay of the backbone structures (**A**, **B**) and ribbon representation (**C**, **D**) of the 20 best structures, superimposed on the PH (**A**, **C**) and BSD1 (**B**, **D**) domains. (**E**) Ribbon representation of a single structure. The PH-D is shown in magenta/pink, the BSD1 domain in blue/cyan and the interdomain linker in yellow.
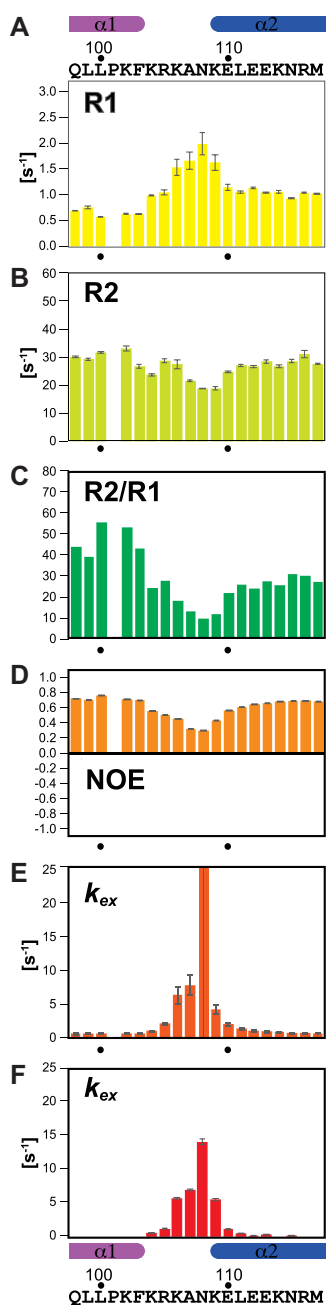
Figure 4. Dynamics of the linker between the PH and BSD1 domains of human TFIIH p62. (**A**) ^{15}N longitudinal relaxation rate (R_1). (**B**) Transverse relaxation rate (R_2). (**C**) Ratio of R_2 to R_1. (**D**) ^{15}N–{^{1}H} NOE. Water–amide proton exchange rate (k_{ex}) measured by HET$_{ex}$-BEST-TROSY (54) (**E**) and CLEANEX-PM (55) (**F**).

9.5 and 11.6, and ^{15}N–{^{1}H} NOE values of 0.451, 0.319, 0.297 and 0.431, respectively.

The high mobility of the interdomain linker residues was confirmed in both the HET$_{ex}$-BEST-TROSY experiment, where the amide proton solvent exchange rates were measured as 1.05 ± 0.11 s^{-1} (Lys104), 2.08 ± 0.20 s^{-1} (Arg105), 6.32 ± 1.22 s^{-1} (Lys106), 7.73 ± 1.51 s^{-1} (Ala107), >10 s^{-1} (Asn108), 4.21 ± 0.64 s^{-1} (Lys109), 2.05 ± 0.18 s^{-1} (Glu110) and 1.40 ± 0.14 s^{-1} (Leu111) (Figure 4E and

Supplementary Figure S3), and the CLEANEX-PM experiment, in which the rates were 0.52 ± 0.05 s^{-1} (Lys104), 1.17 ± 0.08 s^{-1} (Arg105), 5.63 ± 0.09 s^{-1} (Lys106), 6.86 ± 0.15 s^{-1} (Ala107), 13.9 ± 0.40 s^{-1} (Asn108), 5.42 ± 0.18 s^{-1} (Lys109), 1.10 ± 0.06 s^{-1} (Glu110) and 0.45 ± 0.05 s^{-1} (Leu111) (Figure 4F and Supplementary Figure S3). In all experiments, Asn108 displayed the highest mobility among the interdomain linker residues. Thus, this highly mobile interdomain linker causes the dynamic behavior of the PH and BSD1 domains in solution.

Structural model of human TFIIH

Our results enabled us to complete a structural model of the whole human apo TFIIH core complex by first combining the cryo-EM structure with our NMR structure of p62$_{1–158}$ and then considering the dynamics of p62. To build the structural model, initially 399 NMR structures of p62$_{1–158}$ were chosen from the total 1000 calculated structures on the basis of violation criteria. The structure of the BSD1 domain in p62$_{1–158}$ is essentially identical to that of the isolated form (PDB code 2DII), which was used to build p62 in the cryo-EM structure of human apo TFIIH, enabling the BSD1 domain of p62$_{1–158}$ to be placed at the same position in the cryo-EM structure. Hence, the BSD1 domain in the cryo-EM structure could be replaced with the PH and BSD1 domains of our 399 NMR structures of p62$_{1–158}$. Ultimately, we selected 275 models in which the PH-D of p62 did not clash against other parts of TFIIH. Of note, nearly 70% (275/399) of the models had no steric clashes.

The structural models allow us to visualize the dynamic behavior of the PH-D of p62 in apo TFIIH (Figure 5A). The PH-D can move close to both the BSD2 domain (Figure 5B) and the three-helix bundle (Figure 5D), as well as the BSD1 domain (Figure 5F). In addition to its own p62 subunit, it can move close to the C-terminal RecA-like (RecA2) domain of XPB (Figure 5H) and the von Willebrand factor A domain of p44 (Figure 5J). This spatially proximity is in accordance with the results of a chemical cross-linking/mass spectrometry study that detected cross-links between Lys38 of p62 and Lys549 of XPB, and between Lys38 of p62 and Lys126 of p44 (81). Notably, in the majority of our models, the PH-D occupies free space (Figure 5L).

To refine the docking structural models, we performed all-atom MD simulations of the apo whole TFIIH structure. The MD simulations for six representative structural models (Figure 5B, D, F, H, J and L) were carried out in solution, and the snapshot structures obtained at 100 ns are illustrated in Figure 5C, E, G, I, K and M. In the MD simulations, the PH-D, the interdomain linker and the N-terminal part of BSD1 domain of p62 (residues 1–114) were freely moving, whereas the Cα atoms in the other part were restrained to the cryo-EM structure (see the 'Materials and Methods' section). The structures of the PH-D itself were retained (Supplementary Figure S4A), but their positions and orientations changed from the initial structure. In MD simulations of the model structures in Figure 5B, D and F, the PH-D fluctuated around its initial position because it interacted with the BSD2 domain (Figure 5C), the three-helix bundle (Figure 5E) or both the BSD1 domain and the three-helix bundle (Figure 5G) of p62. In the MD simulations
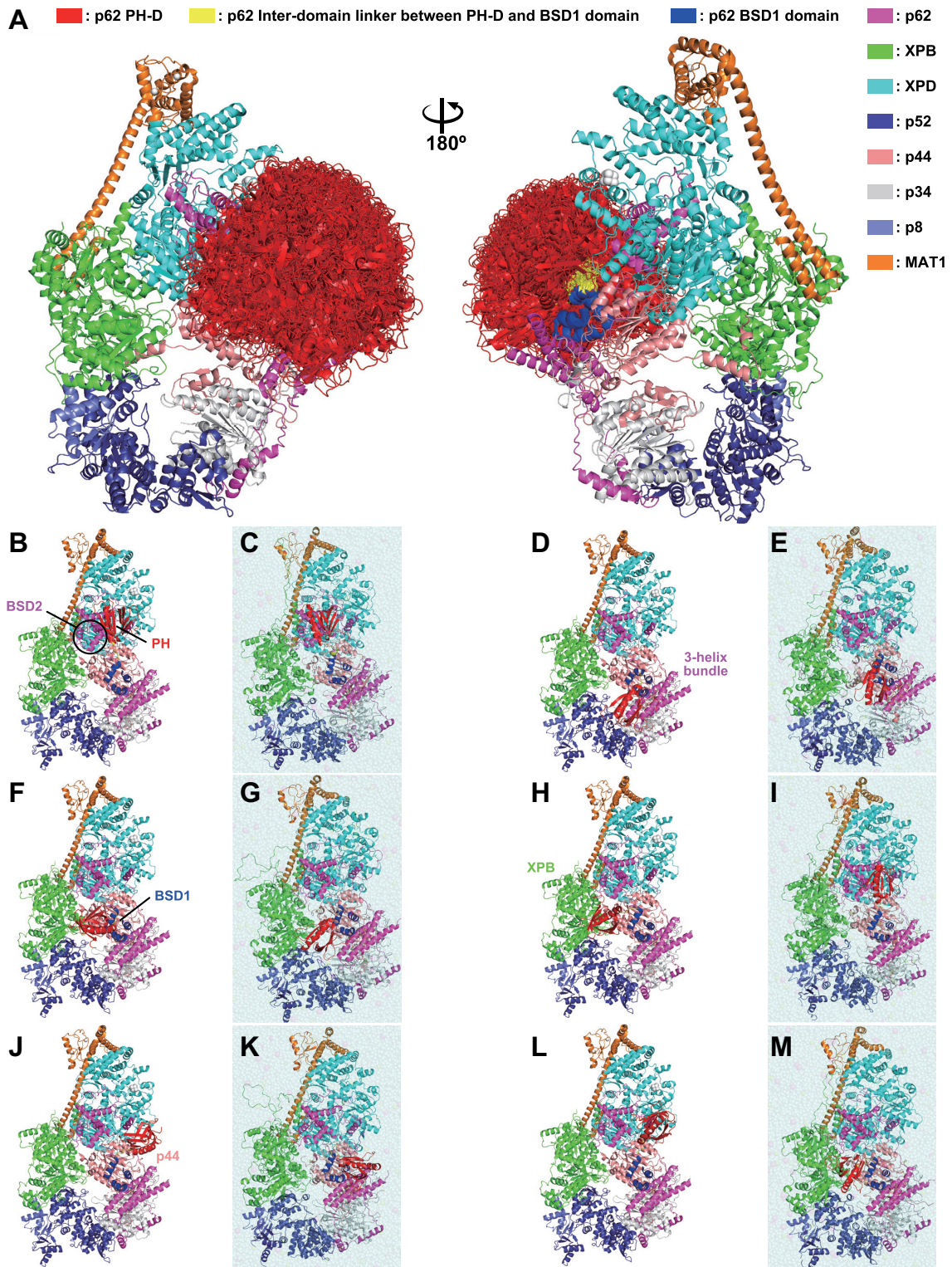
Figure 5. Structural model of human TFIIH. Structural model of human TFIIH built by a combination of the cryo-EM (PDB code 6NMI) and NMR (PDB code 7BUL) structures. The 275 models (**A**), single representative models (**B**, **D**, **F**, **H**, **J**, **L**) and their corresponding MD simulation models (**C**, **E**, **G**, **I**, **K**, **M**) are shown.

of the model structures in Figure 5H, J and L, the PH-D moved from its initial position largely because of the weak interactions that held it in its initial place (Figure 5I, K and M).

This dynamic behavior of the PH-D originates from the high mobility of the interdomain linker (residues 104–108) between the PH-D and BSD1 domain. To estimate the interdomain linker mobility, the RMSF of Cα atoms was calculated from the MD trajectories (Supplementary Figure S5A). The fluctuation in the PH-D except for the N-terminus was up to ∼2 Å. In contrast, the RMSF of the interdomain linker region was markedly higher, indicating that the conformation of the linker is flexible. We found that the flexibility of the interdomain linker region is related to the SASA of the amide hydrogen in backbones (Supplementary Figure S6A). The amide hydrogens in the linker region are exposed to solvent. The dynamic behavior of the PH-D and the mobility of the interdomain linker region observed in MD simulations are in good agreement with our NMR measurements.

Thus, the PH-D of p62 is able to move dynamically around the surface of the core TFIIH complex, transiently interacting with some regions; however, the spatial distribution is limited to some extent owing to the short interdomain linker. As mentioned above, the structural model of human PIC has been reported (78). In that model, the PH-D of p62 is unambiguously positioned between XPB and XPD (Supplementary Figure S7A, left). In our structural model of apo TFIIH (Supplementary Figure S7A, right), it is impossible for the PH-D to sit in the same position because it is too far away from the BSD1 domain. Structural changes of either the PH-D or BSD1 domain, or both, would be required to achieve that positioning. Notably, the C-terminal part of the α1 helix of the PH-D (residues 101–103) and the N-terminal part of the α2 helix of the BSD1 domain (residues 109–111) were unfolded in the published model of holo TFIIH (78) (Supplementary Figure S7B, left; compare with our apo TFIIH model in Figure S7B, right). During formation of the PIC at a transcription start site, the PH and BSD1 domains extensively interact with another transcription factor, TFIIEα (28,29,78). It is not clear whether the partial helix unfolding of the PH-D and BSD1 domain is a result of binding to TFIIEα. Further studies are needed to explore the possibility.

Complex structural models of human TFIIH and p62-binding partners

We have previously determined the solution structures of the PH-D of p62 in complex with five binding partners: the acidic domain of TFIIEα (29); the transactivation domains of the tumor suppressor p53 (31) and the cell cycle regulatory transcription factor DP1 (33); and the acidic regions of the nucleotide excision repair factors XPC (38) and UVSSA (39). All of these proteins bind specifically to the PH-D of p62 via an intrinsically disordered acidic region. Upon binding, this region forms an extended string-like conformation and interacts with a basic surface of the PH-D via extensive electrostatic contacts. Simultaneously, it inserts a phenylalanine or tryptophan residue surrounded by acidic residues into a pocket in the PH-D, enhancing bind-

ing specificity. In those studies, we used the isolated form of the PH-D, consisting of residues 1–108 ($p62_{1-108}$). The current study now demonstrates that this construct design is appropriate because Asn108 is located at the C-terminus of the interdomain linker. Nevertheless, it has been suggested that some binding partners interact with several residues in the C-terminal region of the $p62_{1-108}$ construct (29,38,39). Here, based on the complete structural model of the human apo TFIIH core complex (Figure 5), we have explored whether the p62-binding partners would be able to bind to the PH-D of p62 in TFIIH in the same manner as they bind to $p62_{1-108}$.

To build structural models of TFIIH docked to its p62-binding partners, we selected five structures with sufficient variety from the 275 structural models of TFIIH (Figure 5A). In almost all calculated models, TFIIEα, p53, DP1, XPC and UVSSA could bind to the PH-D of p62 in TFIIH without steric clashes except for the TFIIEα–TFIIH complex, in which the C-terminus of the acidic domain of TFIIEα collided with the N-terminus of the BSD1 domain of p62 (data not shown).

To refine the docking models of the TFIIH with its p62-binding partners, MD simulations were performed for the complex structures. The interactions between the PH-D and the p62-binding partner were maintained for all complexes, and none of the partners dissociated during the simulations (Figure 6). The relative positions and orientations of the PH-D and the binding partner changed with respect to the whole TFIIH, owing to the flexibility of the linker region of p62. As observed for apo TFIIH, the RMSF (Supplementary Figure S5B) and SASA of the amide hydrogens (Supplementary Figure S6B) indicated that the linker region is flexible, even in complex with binding partners.

It should be noted that TFIIEα differs from the other p62-binding partners in possessing some structure in the binding site, although the N-terminal intrinsically disordered acidic region is mainly responsible for binding to the PH-D (29); however, our MD simulation indicates that a stable complex forms between the TFIIEα acidic domain and the p62 PH-D in human TFIIH without any helix unfolding of the PH-D or BSD1 domain (Supplementary Figure S4B). The present complex model of the TFIIEα acidic domain bound to p62 PH-D is likely to correspond to the initiation step in formation of the final PIC. As mentioned above, yeast Tfb1 PH-D is followed by a long (∼50-residue) linker, with the result that the PH-D sits at a different location in PIC as compared with the position of the human p62 PH-D. Even if the p62 PH-D could be located at the same position as the Tfb1 PH-D, it seems unlikely that TFIIEα could bind to the p62 PH-D due to steric hindrance from the structured part of the acidic domain; both the structure in the binding site of TFIIEα and the length of the linker of p62 are likely to determine their positions in PIC.

These results verify the importance of an intrinsically disordered region in recognition of the PH-D of p62 in TFIIH. Namely, the PH-D in TFIIH is sufficiently free to allow access to intrinsically disordered partners, but not completely free to allow access by a structured protein because of the short interdomain linker of p62, which restricts the range of motion of the PH-D in TFIIH as mentioned above. Although the dynamic behavior of the PH-D is essential for
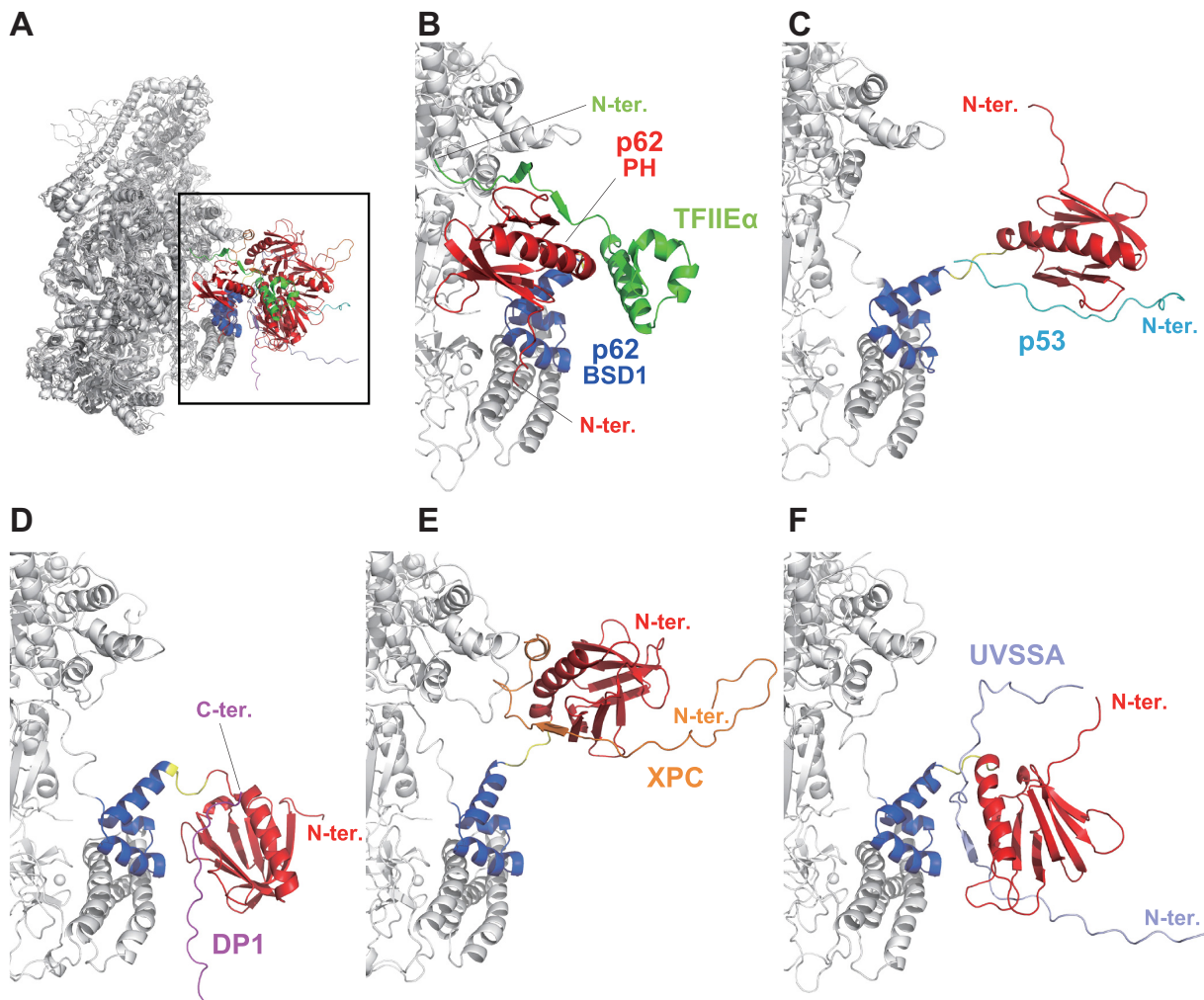
Figure 6. MD simulation models of the complexes of human TFIIH with p62-binding partners. (**A**) Overall view. Extended view of the area indicated by the square in (A): (**B**) complex of TFIIH and TFIIEα; (**C**) complex of TFIIH and p53; (**D**) complex of TFIIH and DP1; (**E**) complex of TFIIH and XPC; and (**F**) complex of TFIIH and UVSSA.

TFIIH recruitment, a completely free state would allow approach by not only true partners but also many other irrelevant proteins, resulting in inefficient recruitment. It is likely that this proper spatial restriction serves to enhance target selectivity.

It has recently been reported that monoubiquitination of Lys414 of UVSSA facilitates transfer of TFIIH to RNA polymerase II stalled at DNA lesions during transcription-coupled nucleotide excision repair ([82]). Our MD simulation shows that Lys414, located at the C-terminus of the p62-binding site, remains accessible even though UVSSA is bound to the PH-D of p62 in TFIIH.

The refined complex models also suggest some electrostatic interactions between acidic residues around the C-terminus of the binding sites of partners and lysine residues in the interdomain linker of p62: namely, Asp436 of TFIIEα and Lys104 of p62; Glu435 of TFIIEα and Lys106 of p62 (Supplementary Figure S8A); Glu62 of p53 and Lys106 of p62 (Supplementary Figure S8B); Glu138 of XPC and Lys104 of p62 (Supplementary Figure S8D); and Glu413 of

UVSSA and Lys104 of p62 (Supplementary Figure S8E). However, there were no electrostatic interactions in the DP1–TFIIH complex, even though DP1 has four consecutive acidic residues at the C-terminus (Supplementary Figure S8C).

Thus, we performed NMR chemical shift perturbation experiments for ^{15}N-labeled p62$_{1-158}$ with the five binding partners and identified significant chemical shift changes of these lysine residues in the interdomain linker of p62 when adding all binding partners except for DP1 (Supplementary Figure S9). The NMR titration experiments were well correlated with the complex models observed in MD simulations. In contrast, Arg105 of p62 displayed a large chemical shift change, but the MD models indicated that this basic residue does not participate in direct binding to the partners except for TFIIEα, where Glu435 or Asp436 might interact with Arg105 of p62; the large chemical shift change of Arg105 in other cases is likely to be caused by interactions of the neighboring lysine residues, Lys104 and/or Lys106, with binding partners (Supplementary Figure S9).

CONCLUSION

Among the 10 subunits of TFIIH, p62 is relatively unique. In addition to its involvement in both formation of the core architecture and recruitment of TFIIH, the cryo-EM structure revealed that an unstructured and extended segment of p62 (residues 266–287) is inserted into the DNA-binding cavity of XPD, suggesting that p62 may function in regulating the helicase activity of XPD (42,43,78). Again, the dynamical behavior of p62 is crucial for TFIIH function. In the present study, we have revealed the dynamic structural model of human apo TFIIH core by using NMR and MD simulations based the cryo-EM structure, and provided the first structural view of the dynamic behavior of the PH-D of p62 in human apo TFIIH, revealing a different range of motion from the PH-D in the previously modeled human holo TFIIH (78) and yeast holo TFIIH (42) structures. We have also presented models of TFIIH complexed with five p62-binding partners refined by MD simulations. These will be useful for discussing interactions between the full-length proteins. For example, a previous study presented a model of a cryo-EM structure of the full-length p53–RNA polymerase II complex docked in the PIC; however, the N-terminal transactivation domain of p53 was invisible owing to its structural flexibility (83). We believe that the complementary approach (NMR, MD and cryo-EM) demonstrated here will help us to better understand the dynamics of such key proteins in large, multi-subunit protein complexes.

DATA AVAILABILITY

Coordinates for the TFIIH p62$_{1-158}$ have been deposited in the RCSB Protein Data Bank (PDB) under ID code 7BUL, and chemical shifts, ^{15}N relaxation parameters ($T_1 = 1/R_1$, $T_2 = 1/R_2$ and heteronuclear NOE) and hydrogen exchange rates (k_{ex}) have been deposited in the BioMagResBank (BMRB) under accession number 36340.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Conaway,R.C. and Conaway,J.W. (1989) An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *Proc. Natl Acad. Sci. U.S.A.*, **86**, 7356–7360.
2. Feaver,W.J., Gileadi,O. and Kornberg,R.D. (1991) Purification and characterization of yeast RNA polymerase II transcription factor b. *J. Biol. Chem.*, **266**, 19000–19005.
3. Gerard,M., Fischer,L., Moncollin,V., Chipoulet,J.M., Chambon,P. and Egly,J.M. (1991) Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.*, **266**, 20940–20945.
4. Xiao,H., Pearson,A., Coulombe,B., Truant,R., Zhang,S., Regier,J.L., Triezenberg,S.J., Reinberg,D., Flores,O., Ingles,C.J. *et al.* (1994) Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol. Cell. Biol.*, **14**, 7013–7024.
5. Tong,X., Drapkin,R., Reinberg,D. and Kieff,E. (1995) The 62- and 80-kDa subunits of transcription factor IIH mediate the interaction with Epstein–Barr virus nuclear protein 2. *Proc. Natl Acad. Sci. U.S.A.*, **92**, 3259–3263.
6. Qadri,I., Conaway,J.W., Conaway,R.C., Schaack,J. and Siddiqui,A. (1996) Hepatitis B virus transactivator protein, HBx, associates with the components of TFIIH and stimulates the DNA helicase activity of TFIIH. *Proc. Natl Acad. Sci. U.S.A.*, **93**, 10578–10583.
7. Lu,H., Fisher,R.P., Bailey,P. and Levine,A.J. (1997) The CDK7–cycH–p36 complex of transcription factor IIH phosphorylates p53, enhancing its sequence-specific DNA binding activity *in vitro*. *Mol. Cell. Biol.*, **17**, 5923–5934.
8. Rochette-Egly,C., Adam,S., Rossignol,M., Egly,J.M. and Chambon,P. (1997) Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. *Cell*, **90**, 97–107.
9. Chen,D., Riedl,T., Washbrook,E., Pace,P.E., Coombes,R.C., Egly,J.M. and Ali,S. (2000) Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Mol. Cell*, **6**, 127–137.
10. Lee,D.K., Duan,H.O. and Chang,C. (2000) From androgen receptor to the general transcription factor TFIIH. Identification of cdk activating kinase (CAK) as an androgen receptor NH$_2$-terminal associated coactivator. *J. Biol. Chem.*, **275**, 9308–9313.
11. Liu,J., He,L., Collins,I., Ge,H., Libutti,D., Li,J., Egly,J.M. and Levens,D. (2000) The FBP interacting repressor targets TFIIH to inhibit activated transcription. *Mol. Cell*, **5**, 331–341.
12. Compe,E., Drané,P., Laurent,C., Diderich,K., Braun,C., Hoeijmakers,J.H. and Egly,J.M. (2005) Dysregulation of the peroxisome proliferator-activated receptor target genes by XPD mutations. *Mol. Cell. Biol.*, **25**, 6065–6076.
13. Compe,E., Malerba,M., Soler,L., Marescaux,J., Borrelli,E. and Egly,J.M. (2007) Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH. *Nat. Neurosci.*, **10**, 1414–1422.
14. Yankulov,K.Y., Pandes,M., McCracken,S., Bouchard,D. and Bentley,D.L. (1996) TFIIH functions in regulating transcriptional elongation by RNA polymerase II in *Xenopus* oocytes. *Mol. Cell. Biol.*, **16**, 3291–3299.
15. Assfalg,R., Lebedev,A., Gonzalez,O.G., Schelling,A., Koch,S. and Iben,S. (2012) TFIIH is an elongation factor of RNA polymerase I. *Nucleic Acids Res.*, **40**, 650–659.
16. Narita,T., Narita,K., Takedachi,A., Saijo,M. and Tanaka,K. (2015) Regulation of transcription elongation by the XPG–TFIIH complex is implicated in Cockayne syndrome. *Mol. Cell. Biol.*, **35**, 3178–3188.
17. Roy,R., Adamczewski,J.P., Seroz,T., Vermeulen,W., Tassan,J.P., Schaeffer,L., Nigg,E.A., Hoeijmakers,J.H. and Egly,J.M. (1994) The MO15 cell cycle kinase is associated with the TFIIH transcription–DNA repair factor. *Cell*, **79**, 1093–1101.
18. Fesquet,D., Morin,N., Doree,M. and Devault,A. (1997) Is Cdk7/cyclin H/MAT1 the genuine cdk activating kinase in cycling *Xenopus* egg extracts? *Oncogene*, **15**, 1303–1307.
19. Larochelle,S., Pandur,J., Fisher,R.P., Salz,H.K. and Suter,B. (1998) Cdk7 is essential for mitosis and for *in vivo* Cdk-activating kinase activity. *Genes Dev.*, **12**, 370–381.

20. Wu,L., Chen,P., Hwang,J.J., Barsky,L.W., Weinberg,K.I., Jong,A. and Starnes,V.A. (1999) RNA antisense abrogation of MAT1 induces G1 phase arrest and triggers apoptosis in aortic smooth muscle cells. *J. Biol. Chem.*, **274**, 5564–5572.

21. Schaeffer,L., Roy,R., Humbert,S., Moncollin,V., Vermeulen,W., Hoeijmakers,J.H., Chambon,P. and Egly,J.M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science*, **260**, 58–63.

22. Drapkin,R., Reardon,J.T., Ansari,A., Huang,J.C., Zawel,L., Ahn,K., Sancar,A. and Reinberg,D. (1994) Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature*, **368**, 769–772.

23. Guzder,S.N., Qiu,H., Sommers,C.H., Sung,P., Prakash,L. and Prakash,S. (1994) DNA repair gene RAD3 of *S. cerevisiae* is essential for transcription by RNA polymerase II. *Nature*, **367**, 91–94.

24. Guzder,S.N., Sung,P., Bailly,V., Prakash,L. and Prakash,S. (1994) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature*, **369**, 578–581.

25. Egly,J.M. and Coin,F. (2011) A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. *DNA Repair (Amst.)*, **10**, 714–721.

26. Compe,E. and Egly,J.M. (2012) TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.*, **13**, 343–354.

27. Lehmann,A.R. (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie*, **85**, 1101–1111.

28. Di Lello,P., Jenkins,M.M.L., Mas,C., Langlois,C., Malitskaya,E., Fradet-Turcotte,A., Archambault,J., Legault,P. and Omichinski,J.G. (2008) p53 and TFIIEalpha share a common binding site on the Tfb1/p62 subunit of TFIIH. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 106–111.

29. Okuda,M., Tanaka,A., Satoh,M., Mizuta,S., Takazawa,M., Ohkuma,Y. and Nishimura,Y. (2008) Structural insight into the TFIIE–TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *EMBO J.*, **27**, 1161–1171.

30. Di Lello,P., Jenkins,L.M., Jones,T.N., Nguyen,B.D., Hara,T., Yamaguchi,H., Dikeakos,J.D., Appella,E., Legault,P. and Omichinski,J.G. (2006) Structure of the Tfb1/p53 complex: insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol. Cell*, **22**, 731–740.

31. Okuda,M. and Nishimura,Y. (2014) Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. *J. Am. Chem. Soc.*, **136**, 14143–14152.

32. Mas,C., Lussier-Price,M., Soni,S., Morse,T., Arseneault,G., Di Lello,P., Lafrance-Vanasse,J., Bieker,J.J. and Omichinski,J.G. (2011) Structural and functional characterization of an atypical activation domain in erythroid Kruppel-like factor (EKLF). *Proc. Natl Acad. Sci. U.S.A.*, **108**, 10484–10489.

33. Okuda,M., Araki,K., Ohtani,K. and Nishimura,Y. (2016) The interaction mode of the acidic region of the cell cycle transcription factor DP1 with TFIIH. *J. Mol. Biol.*, **428**, 4993–5006.

34. Di Lello,P., Nguyen,B.D., Jones,T.N., Potempa,K, Kobor,M.S., Legault,P. and Omichinski,J.G. (2005) NMR structure of the amino-terminal domain from the Tfb1 subunit of TFIIH and characterization of its phosphoinositide and VP16 binding sites. *Biochemistry*, **44**, 7678–7686.

35. Langlois,C., Mas,C., Di Lello,P., Jenkins,L.M., Legault,P. and Omichinski,J.G. (2008) NMR structure of the complex between the Tfb1 subunit of TFIIH and the activation domain of VP16: structural similarities between VP16 and p53. *J. Am. Chem. Soc.*, **130**, 10596–10604.

36. Chabot,P.R., Raiola,L., Lussier-Price,M., Morse,T., Arseneault,G., Archambault,J. and Omichinski,J.G. (2014) Structural and functional characterization of a complex between the acidic transactivation domain of EBNA2 and the Tfb1/p62 subunit of TFIIH. *PLoS Pathog.*, **10**, e1004042

37. Lafrance-Vanasse,J., Arseneault,G., Cappadocia,L., Legault,P. and Omichinski,J.G. (2013) Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic Acids Res.*, **41**, 2736–2745.

38. Okuda,M., Kinoshita,M., Kakumu,E., Sugasawa,K. and Nishimura,Y. (2015) Structural insight into the mechanism of TFIIH recognition by the acidic string of the nucleotide excision repair factor XPC. *Structure*, **23**, 1827–1837.

39. Okuda,M., Nakazawa,Y., Guo,C., Ogi,T. and Nishimura,Y. (2017) Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. *Nucleic Acids Res.*, **45**, 13043–13055.

40. Gervais,V., Lamour,V., Jawhari,A., Frindel,F., Wasielewski,E., Dubaele,S., Egly,J.M., Thierry,J.C., Kieffer,B. and Poterszman,A. (2004) TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nat. Struct. Mol. Biol.*, **11**, 616–622.

41. Greber,B.J., Nguyen,T.H.D., Fang,J., Afonine,P.V., Adams,P.D. and Nogales,E. (2017) The cryo-electron microscopy structure of human transcription factor IIH. *Nature*, **549**, 414–417.

42. Schilbach,S., Hantsche,M., Tegunov,D., Dienemann,C., Wigge,C., Urlaub,H. and Cramer,P. (2017) Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature*, **551**, 204–209.

43. Greber,B.J., Toso,D.B., Fang,J. and Nogales,E. (2019) The complete structure of the human TFIIH core complex. *eLife*, **8**, e44771.

44. Cavanagh,J., Fairbroher,W.J., Palmer,A.G. III and Skelton,N.J. (1996) In: *Protein NMR Spectroscopy*. Academic Press, San Diego, CA.

45. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

46. Johnson,B.A. and Blevins,R.A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

47. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.

48. Brünger,A.T. (1993) In: *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, CT.

49. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.

50. Linge,J.P., Williams,M.A., Spronk,C.A., Bonvin,A.M. and Nilges,M. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.

51. Laskowski,R.A., Rullmann,J.A.C., 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.

52. Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

53. Mandel,A.M., Akke,M. and Palmer,A.G. III (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, **246**, 144–163.

54. Rennella,E., Solyom,Z. and Brutscher,B. (2014) Measuring hydrogen exchange in proteins by selective water saturation in ^{1}H–^{15}N SOFAST/BEST-type experiments: advantages and limitations. *J. Biomol. NMR*, **60**, 99–107.

55. Hwang,T.L., van Zijl,P.C. and Mori,S. (1998) Accurate quantitation of water–amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR*, **11**, 221–226.

56. Sali,A. and Blundell,T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

57. Humphrey,A., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

58. Chresten,R.S., Olsson,H.M.M., Rostkowski,M. and Jensen,H.J. (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. *J. Chem. Theory Comput.*, **7**, 2284–2295.

59. Mats,H.M.O., Sondergaard,R.C., Rotkowski,M. and Jensen,H.J. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.*, **7**, 525–537.

60. Dolinsky,T.J., Nielsen,J.E., McCammon,J.A. and Baker,N.A. (2004) PDB2PAR: an automated pipeline for the setup, execution, and analysis of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.

61. Huang,J., Rauscher,S., Nawrocki,G., Ran,T., Feig,M., de Groot,L.B., Grubmükker,H. and Mackerell,D.A. (2017) CHARMM36m: an

improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, **14**, 71–73.

62. MacKerell,D.A., Bashford,D., Bellott,M., Dunbrack,L.R., Evanseck,D.J., Field,J.M., Fischer,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

63. MacKerell,D.A., Feig,M. and Brooks,L.C. (2004) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.*, **126**, 698–699.

64. Chang,H.C. and Kim,K. (2009) Density functional theory calculation of bonding and charge parameters for molecular dynamics studies on [FeFe] hydrogenases. *J. Chem. Theory Comput.*, **5**, 1137–1145.

65. Jorgensen,L.W., Chandrasekhar,J., Madura,D.J., Impey,M.R. and Klein,L.M. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

66. Abraham,J.M., Murtola,T., Schuls,R., Páll,S., Smith,C.J., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19–25.

67. Páll,S., Abraham,J.M., Kutzner,C., Hess,B. and Lindahl,E. (2015) Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In: Markidis,S. and Laure,E. (eds). *Solving Software Challenges for Exascale*, Vol. **8759**. Springer, Cham, Switzerland, pp. 3–27.

68. Pronk,S., Páll,S., Schulz,R., Larsson,P., Bjelkmar,P., Apostolov,R., Shirts,R.M., Smith,C.J., Kasson,M.P., van der Spoel,D. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.

69. Essmann,U., Perera,L., Berkowitz,L.M., Darden,T., Lee,H. and Pedersen,G.L. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.

70. Hess,B. (2008) P-LINKS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, **4**, 116–122.

71. Nosé,S. (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **52**, 255–268.

72. Hoover,G.W. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695–1697.

73. Berendsen,C.J.H., Postma,M.P.J., van Gunsteren,F.W., DiNola,A. and Haak,R.J. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.

74. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.*, **52**, 7182–7190.

75. Nosé,S. and Klein,L.M. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **50**, 1055–1076.

76. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

77. Roe,R.D. and Cheatham,E.T. III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.

78. Yan,C., Dodd,T., He,Y., Tainer,J.A., Tsutakawa,S.E. and Ivanov,I. (2019) Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.*, **26**, 397–406.

79. Hafsa,N.E. and Wishart,D.S. (2014) CSI 2.0: a significantly improved version of the chemical shift index. *J. Biomol. NMR*, **60**, 131–146.

80. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.

81. Luo,J., Cimermancic,P., Viswanath,S., Ebmeier,C.C., Kim,B., Dehecq,M., Raman,V., Greenberg,C.H., Pellarin,R., Sali,A. *et al.* (2015) Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol. Cell*, **59**, 794–806.

82. Nakazawa,Y., Hara,Y., Oka,Y., Komine,O., van den Heuvel,D., Guo,C., Daigaku,Y., Isono,M., He,Y., Shimada,M. *et al.* (2020) Ubiquitination of DNA damage-stalled RNAPII promotes transcription-coupled repair. *Cell*, **180**, 1228–1244.

83. Singh,S.K., Qiao,Z., Song,L., Jani,V., Rice,W., Eng,E., Coleman,R.A. and Liu,W.L. (2016) Structural visualization of the p53/RNA polymerase II assembly. *Genes Dev.*, **30**, 2527–2537.