



OPEN

Structural assessment of the full-length wild-type tumor suppressor protein p53 by mass spectrometry-guided computational modeling

Alessio Di Ianni^{1,2,5}, Christian Tüting^{3,5}, Marc Kipping^{1,2}, Christian H. Ihling^{1,2}, Janett Köppen^{1,2}, Claudio Iacobucci^{1,2,4}, Christian Arlt^{1,2}✉, Panagiotis L. Kastiris³ & Andrea Sinz^{1,2}✉

The tetrameric tumor suppressor p53 represents a great challenge for 3D-structural analysis due to its high degree of intrinsic disorder (ca. 40%). We aim to shed light on the structural and functional roles of p53's C-terminal region in full-length, wild-type human p53 tetramer and their importance for DNA binding. For this, we employed complementary techniques of structural mass spectrometry (MS) in an integrated approach with computational modeling. Our results show no major conformational differences in p53 between DNA-bound and DNA-free states, but reveal a substantial compaction of p53's C-terminal region. This supports the proposed mechanism of unspecific DNA binding to the C-terminal region of p53 prior to transcription initiation by specific DNA binding to the core domain of p53. The synergies between complementary structural MS techniques and computational modeling as pursued in our integrative approach is envisioned to serve as general strategy for studying intrinsically disordered proteins (IDPs) and intrinsically disordered region (IDRs).

The tumor suppressor p53, which is often referred to as the “guardian of the genome”, plays an important role in DNA repair, cell cycle control, and apoptosis in human cells^{1–3}. Human p53 is biologically active as a homotetramer formed by association of initial dimers⁴. It is one of the most studied and prominent representatives of intrinsically disordered proteins (IDPs)⁵. The structural and functional characterization of IDPs has been recognized as a crucial task in structural biology, considering the abundance of proteins containing intrinsically disordered regions (IDRs) in all domains of life (archaea, bacteria, eukaryotes) and in viruses^{6,7}. More than 40 years after the discovery of the prominent tumor suppressor protein the structure and spatial arrangement of the tetrameric p53 complex, remain elusive. The protein has a modular domain organization (Fig. 1) with two ordered domains, the DNA binding domain (DBD) and tetramerization domain (TET), and contains approximately 40% of IDRs comprising the N-terminal transactivation domain (TAD) and proline-rich region (PRR) as well as the C-terminal domain (CTD)⁸. The p53 tetramer cooperatively binds to its target duplex DNA in a sequence-specific manner via its DBD^{9,10}. The target binding sites consist of two palindromic decameric motifs (half-sites) of the general sequence 5'-RRRCWWGYYY-3' (R = A, G; W = A, T; Y = C, T)¹¹. P53 specifically recognizes genes which contain these motifs and regulate their expression levels. Structures of the DBD^{12,13} and the TET^{14,15} are available, as well as combinations of the two^{16,17}, but detailed structural information on the N- and C-terminal regions in context of full-length, wild-type human p53 is still missing. Recently, a cryo-electron microscopy (cryo-EM) structure of p53 with the nucleosome, H2A, H2B, H3.1, and H4 plus DNA fragment containing the p53 binding site was proposed, but the low resolution obtained for p53 did not allow

¹Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3, 01620 Halle (Saale), Germany. ²Center for Structural Mass Spectrometry, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3, 01620 Halle (Saale), Germany. ³ZIK HALOmern and Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3, 06120 Halle (Saale), Germany. ⁴Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, Coppito, 67100 L'Aquila, Italy. ⁵These authors contributed equally: Alessio Di Ianni and Christian Tüting. ✉email: Christian.arlt@pharmazie.uni-halle.de; andrea.sinz@pharmazie.uni-halle.de

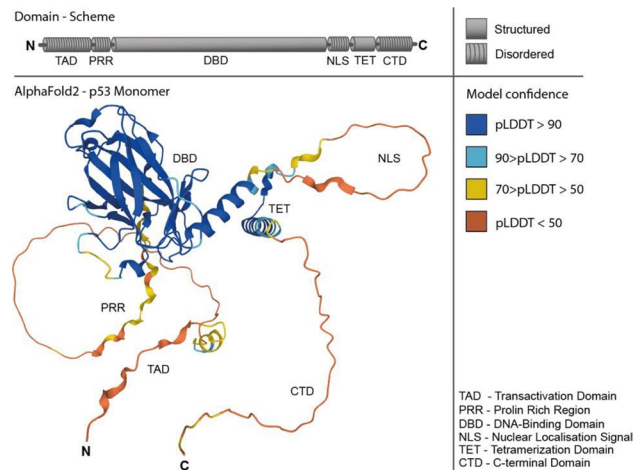


Figure 1. Domain scheme and AlphaFold2 prediction of full-length, wild-type human p53 monomer (cellular tumor antigen p53, <https://alphafold.ebi.ac.uk>)^{27,28}. Upper panel: structured and disordered regions of p53 are indicated in the domain scheme. Lower panel: AlphaFold2 prediction gives low confidence scores (pLDDT) for the IDRs of p53.

resolving its structure in the cryo-EM map¹⁸. Different models have been proposed for the topology and overall conformation of the p53 tetramer in the absence and presence of DNA^{19–26}. Those models were generated using both experimental as well as computational approaches and differ substantially in their domain organization (Supporting Information, Figure S1).

Currently, the method of choice for 3D-structure analysis of IDPs is NMR spectroscopy despite its limitations for larger proteins, but complementary structural biology techniques are under continuous improvement²⁹. These include cryo-EM in combination with structural mass spectrometry (MS) techniques, such as native MS³⁰, ion-mobility MS (IM-MS)³¹, hydrogen/deuterium exchange MS (HDX-MS)³² and cross-linking MS (XL-MS)³³. While a computational prediction of disordered regions within protein sequences is possible, determining the 3D-structures of IDRs remains a significant challenge due to their inherently dynamic nature and shallow energy landscape. Recent advances of AI-based prediction tools, such as AlphaFold2²⁷, have enabled the generation of structures for well-folded protein domains, while predicting the structures of IDRs/IDPs is still a daunting task (Fig. 1). To further investigate the structural properties of p53, with a special focus on the understudied disordered C-terminal region, we adopted an integrative structural biology approach. Specifically, we investigate the topology of p53 in the absence and presence of DNA in solution. In contrast to the majority of structural data published for p53 where IDRs are missing completely or where IDR-derived peptides are used as substitutes^{34–36}, we employ full-length, wild-type human p53³⁷. Our approach combines complementary structural MS methods, AI-based structure prediction with AlphaFold2^{27,38}, data-driven protein modeling, and in-depth scoring and validation of derived models. By this, we gained novel insights into the C-terminal IDR of p53 and evaluate potential structural changes thereof upon DNA binding.

Experimental

Chemicals. All chemicals used in this study were of the highest purity available and were obtained from Sigma. Bis(sulfosuccinimidyl)glutarate (BS²G) cross-linkers (D₀ and D₄) were obtained from Thermo Fisher Scientific; trypsin (porcine, sequencing grade) was obtained from Promega. AspN (purified from *Flavobacterium meningosepticum*, sequencing grade) was obtained from New England Biolabs. DNA response element (top strand: 5'-CGCGGACATGTCCGGACATGTCCCGC-3' and complementary bottom strand: 5'-GCGGGACATGTCCGGACATGTCCGCG-3') was purchased from Microsynth. Nano-HPLC solvents were LC/MS grade (HiPerSolv CHROMANORM, VWR).

Expression and Purification of p53. The HLT-p53 fusion protein was expressed in *Escherichia coli* BL21 (DE3) cells as previously described^{37,39}. Cells were harvested after incubation (16 h, 18 °C) and lysed by ultrasonication in 50 mM 2-[4-(2-hydroxyethyl)piperazine-1-yl]ethanesulfonic acid (HEPES), 300 mM NaCl, 2.5 mM tris(2-carboxyethyl)phosphine (TCEP), 20 mM imidazole, pH 8.0. The HLT-p53 fusion protein was purified by affinity chromatography on a FPLC system (ÄKTA Pure, GE Healthcare) using a 1-mL HisTrap FF column (GE Healthcare), followed by size-exclusion chromatography (SEC, Superdex 200 column, GE Healthcare) in 50 mM HEPES, 300 mM NaCl, 2.5 mM TCEP, 10% (v/v) glycerol, pH 7.2. After TEV cleavage *o/n* at 4 °C the size exclusion purification step was repeated. The functionality of purified p53 was verified by native MS showing tetramerization of p53 and by a DNA binding assay as shown previously^{37,39}.

Cross-linking. Solutions containing 10 μM wild-type p53 in 50 mM HEPES, 300 mM NaCl, and 2.5 mM TCEP, 10% (v/v) glycerol, pH 7.2, were incubated overnight with or without DNA response element (RE) and cross-linked at 4 °C for 1 h with 0.5 mM BS²G-D₀ and BS²G-D₄, respectively. Cross-linkers were dissolved in neat

dimethyl sulfoxide (DMSO) immediately before adding it to the protein solutions. The cross-linking reactions were performed in triplicates for each condition and repeated with reversed deuterium labeling. The reactions were quenched by adding ammonium bicarbonate to a final concentration of 20 mM. For XL-MS, samples with and without RE were mixed (1:1) and separated by SDS-PAGE (10% resolving/5% stacking gel). The monomeric p53 band was excised from the gel (Supporting Information, Figure S2) and *in-gel* digested with AspN (37 °C, overnight) and trypsin (37 °C, 4 h). Peptides were recovered using an extraction buffer (5% (v/v) TFA/acetonitrile; 1:2) and the solution was subjected to LC-MS/MS analysis.

Mass spectrometry. Peptide mixtures were analyzed by LC-MS/MS on an UltiMate 3000 RSLC nano-HPLC system (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer equipped with an EASYSpray ion source (Thermo Fisher Scientific). Peptides were trapped on a C18 precolumn (Acclaim PepMap 100, 300 $\mu\text{m} \times 5 \text{ mm}$, 5 μm , 100 Å, Thermo Fisher Scientific) and separated on a μPAC C18, 50-cm column (PharmaFluidics). After trapping, peptides were eluted by a concave 90-min water-acetonitrile gradient from 3 to 30% B (solvent B: 80% (v/v) acetonitrile, 0.1% (v/v) formic acid) at a flow rate of 300 nL/min.

Data were acquired in data-dependent MS/MS mode using stepped higher-energy collision-induced dissociation (HCD) at normalized collision energies of 27%, 30%, and 33%. High-resolution full scans (m/z 300–1700, $R = 120,000$ at m/z 200) were followed by high-resolution product ion scans ($R = 15,000$) for 5 s, starting with the most intense signal in the full-scan mass spectrum. Precursor ions with charge states $> 2+$ and $< 8+$ were selected for fragmentation; the isolation window was set to 2 Th. Dynamic exclusion of 60 s (window 2 ppm) was enabled to allow the detection of less abundant ions. Data acquisition was controlled via XCalibur 4.3 (Thermo Fisher Scientific).

Data analysis. Cross-links were manually quantified with XCalibur 4.0. Extracted ion chromatograms (XICs) of the precursor masses for BS²G-D₀ and BS²G-D₄ cross-linked peptides were extracted and BS²G-D₄/D₀ ratios were calculated using the relative intensities of the precursor ions' isotope distributions. In case of identifying a cross-linked peptide at different charge states or in case it contained an oxidized methionine, ratios were calculated for all possible peptide states. Identification of cross-links (Supporting Information, Figures S3–S16) and “dead-end” cross-links (Supporting Information, Figures S17–S31) was performed with MeroX 2.0.1.7⁴⁰. The following settings were applied: proteolytic cleavage C-terminally at Lys and Arg (up to 3 missed cleavages were allowed) and N-terminally at Asp and Glu (up to 3 missed cleavages were allowed); peptide length: 5–30 amino acids; modifications: alkylation of Cys by iodoacetamide (fixed), oxidation of Met (variable); cross-linker specificity: Lys, Ser, Thr, Tyr, N-terminus; search algorithm: quadratic mode; precursor mass accuracy: 10 ppm; fragment ion mass accuracy: 20 ppm; signal-to-noise ratio > 1.5 ; precursor mass correction enabled; false discovery rate (FDR) cut-off: 1%, and minimum score cut-off: 20.

“Dead-end” cross-links were analyzed to extract quantitative footprinting data using Mass Spec Studio 2.0 (<https://www.msstudio.ca>)⁴¹. The following settings were applied: protein states, plusDNA, minusDNA; reagents, BS²G-OH-D₀ (composition C₅H₆O₃, monoisotopic mass 114.03169), BS²G-OH-D₄ (composition C₅D₄H₂O₃, monoisotopic mass 118.05680); amino acid modification, methionine oxidation (variable), cysteine carbamidomethylation (fixed); acquisition mode, DDA; fragmentation, HCD; precursor tolerance, 10 ppm; number of tolerable termini, 2; peptide charge, 2–6; peptide length, 5–40; max modifications per peptide, 3; enzyme, trypsin; XIC m/z selector tolerance, 6 ppm; XIC peak to background ratio, 2; fragment mass selector, 20 ppm. Absolute residue labeling yields were normalized to the unbound state. MS data are available via ProteomeXchange with identifier PXD037030.

HDX-MS. H/D exchange experiments of p53 were performed at room temperature. 1 μL of 10 μM p53 in 50 mM HEPES, 300 mM NaCl, 2.5 mM TCEP, 10% (v/v) glycerol, pH 7.2 was diluted with 9 μL of 10% (v/v) glycerol in 90% (v/v) D₂O for 30 s, 1 min or 2 min. Samples were injected to a home-built HDX apparatus, consisting of an HP1200 (Agilent, Santa Clara, USA) HPLC system, cooled to 4 °C in a DB951 showcase (Polar Refrigeration, Bristol, UK), and a column setup comprising an Enzymate Protein Pepsin Column, XBridge BEH C18 with VanGuard cartridge 5 mm \times 2.1 mm, 2.5 μm , 300 Å and an XBridge Peptide BEH C18 column 100 mm \times 1 mm, 3.5 μm , 300 Å (Waters, Milford, USA). Digestion time was 10 min at 25 $\mu\text{L}/\text{min}$ 0.1% (v/v) formic acid in water; separation was performed within 6 min at a flow rate of 80 $\mu\text{L}/\text{min}$ using a gradient of 3–50% (v/v) acetonitrile in water with 0.1% (v/v) formic acid. Peptic peptides were analyzed with a timsTOF Pro mass spectrometer (Bruker Daltonik, Bremen, Germany) with electrospray ionization source with a source temperature of 50 °C, nebulizer gas pressure of 3 bar, dry gas flow of 8 L/min, and capillary voltage of 3500 V.

Mass spectra were recorded in the m/z range 100–1700 using the standard proteomics settings of the timsTOF Pro in the COMPASS software. The HDExaminer 3.3.0 software (Sierra Analytics, Modesto, USA) was used to extract the deuterium content of peptic peptides. Data were corrected for back-exchange using data from six independent 10- μL injections of p53 stored in buffer containing 80% (v/v) D₂O, 10% (v/v) glycerol, and 10% (v/v) H₂O.

AlphaFold2 structure prediction. A locally installed version of AlphaFold2 (v2.2) was used^{27,38} with the database accession date April 1, 2022. Tetrameric p53 was predicted with the pre-set ‘multimer’ feature (Supporting Information, Figure S32). The amino acid sequence of full-length, wild-type human p53 was used as input.

Distance restraints definition. Experimental distance restraints were defined on the basis of the experimentally obtained BS²G cross-linking data. Distance restraints were defined in MODELLER⁴² as distance

between the Ca atoms of both cross-linked residues within a monomeric p53 chain as *UpperBound* restraint with a mean distance of 15 Å and a standard deviation of 5 Å. For ‘all restraints’, all distance restraints were applied to all four p53 chains symmetrically. For ‘reduced restraints’, 6 out of 12 experimentally determined cross-links were randomly selected for each p53 monomer, resulting in an asymmetric selection of tetrameric p53 structures. For ‘random restraints’, two residues within the amino acid sequence of p53 (residues 301–393) were randomly selected, 12 random restraints were defined and applied to all monomeric p53 chains symmetrically (Supporting Information, Figure S33).

Model refinement. The initial AlphaFold2 model of tetrameric full-length p53 was truncated to the region comprising residues 301–393 and used as initial template for the MODELLER refinement. MODELLER was run by default with activated variable target function method (VTFM) and molecular dynamics (MD) optimization. Distance restraints were redefined for each unique refinement as described above. For the ‘all restraints’ condition, 12,000 unique models were predicted. For ‘reduced restraints’ and ‘random restraints’ conditions, 100 unique restraints were defined and 20 unique models were predicted for each condition. In total, 16,000 refined models were predicted (Supporting Information, Figure S33).

Model validation and scoring. All predicted models were analyzed regarding cross-link distance as well as labeling satisfaction. BS²G cross-linking data generated in this study as well as published disuccinimidyl-dibutyric urea (DSBU) cross-linking data³⁹ were used, Ca–Ca distances were determined, and a scoring matrix was applied (Supporting Information, Figure S34; Supporting Information, Table S1). For satisfaction of the footprinting data, the solvent-accessible surface area (SASA) was calculated with the MSMS-toolkit⁴³, the pK_a values were calculated with PropKa⁴⁴. For SASA scoring, a scoring matrix was applied (Supporting Information, Figure S35; Supporting Information, Table S2). The radius of gyration was calculated with Python code, adapted from <https://github.com/sarisabban/Rg>.

Results

The tumor suppressor p53 poses immense challenges for deriving structure–function relationships due to its high content of intrinsic disorder⁴⁵. The majority of p53 structural studies has been carried out by using truncated p53 variants where the IDRs have been partially or completely removed³⁴ or by using peptide representatives derived from p53’s IDRs^{35,36}. In this work, we use full-length, wild-type human p53 without any truncations or modifications to gain more detailed mechanistic insights into the structure–function relationships of p53, with a special focus on the C-terminal region. The C-terminal region of p53 plays a crucial role in target recognition and is highly modified by acetylation, ubiquitination, phosphorylation, etc.⁴⁶, but detailed molecular information on this region is still limited. It has been hypothesized that the lysine-rich C-terminal region of p53 assists in DNA-binding, mainly mediated via electrostatic interactions. Then, p53 slides along the DNA until the response element is recognized by the DBD of p53 for high-affinity binding⁴⁷.

Structural investigation of full-length p53. One cornerstone of our structural investigation is XL-MS as it is one of the few techniques in structure biology that can handle IDPs/IDRs. In this study, we employ the homobifunctional cross-linker BS²G that targets primary amine groups in proteins, i.e., lysines and N-termini⁴⁸. Its reactivity makes BS²G an ideal reagent to specifically target the CTD of p53 with its high lysine content. Additionally, two BS²G isotopologs, BS²G-D₀ and BS²G-D₄, are commercially available (Fig. 2A). These reagents generate distinct MS patterns, simplifying cross-link identification and quantification (Fig. 2A). We identified several unique intramolecular cross-links in p53, mainly covering the disordered NLS and CTD (Fig. 2B; Supporting Information, Table S3). BS²G cross-links were quantified in the mass spectra to deduce eventual structural changes in the DNA-free versus DNA-bound states of p53 (Fig. 2A). Cross-links identified in at least two out of six replicates were used for quantification. Strikingly, the cross-link patterns did not show major differences between DNA-bound and DNA-free states, indicating similar conformational ensembles of p53 (Fig. 2C).

The same p53 samples were used to extract protein footprinting MS data by quantifying “dead-end” cross-links (Fig. 2B)⁴⁹. “Dead-end” cross-links are peptides that are modified by a partially hydrolyzed cross-linker. These species do not provide direct distance information, but yield important insights into the overall topology of proteins based on solvent accessibilities and local pK_a values of specific amino acid side chains. The combined information from interpeptide and “dead-end” cross-links is a highly valuable input for 3D-protein modeling. The mean labeling rates range between ~2 and 30%, and residues were classified as buried (<5% labeling rate), partially accessible (<10% labeling rate), and accessible (>10% labeling rate) (Supporting Information, Table S4). Conclusively, the footprinting results indicate that the accessibility of p53 CTD remains unaltered upon DNA binding (Fig. 2B,C). This finding is in agreement with XL-MS data and suggests that the C terminus of p53 retains its overall topology upon DNA binding.

Structural modeling of p53. The cross-links identified in this study as well as in previous work³⁹ are predominantly located in the C-terminal region of p53, specifically in the TET with flanking IDRs. The data-driven modeling of p53 is therefore focused on the C-terminal region (residues 301 to 393; comprising NLS, TET and CTD). The initial tetrameric full-length p53 model was generated with AlphaFold2^{27,38} and the C-terminal regions were isolated. The ordered TET shows high confidence and is in agreement with previously solved X-ray and NMR structures^{14,15}. However, AlphaFold2 gives only low-confidence models for the disordered regions of p53 (NLS, CTD; Fig. 1; Supporting Information, Figure S32). To optimize the structures by satisfying the distance restraints, the initial AlphaFold2 models were subjected to MODELLER. The experimentally derived distance information as well as a reduced and a randomized data set (see “Model refinement” for details) were

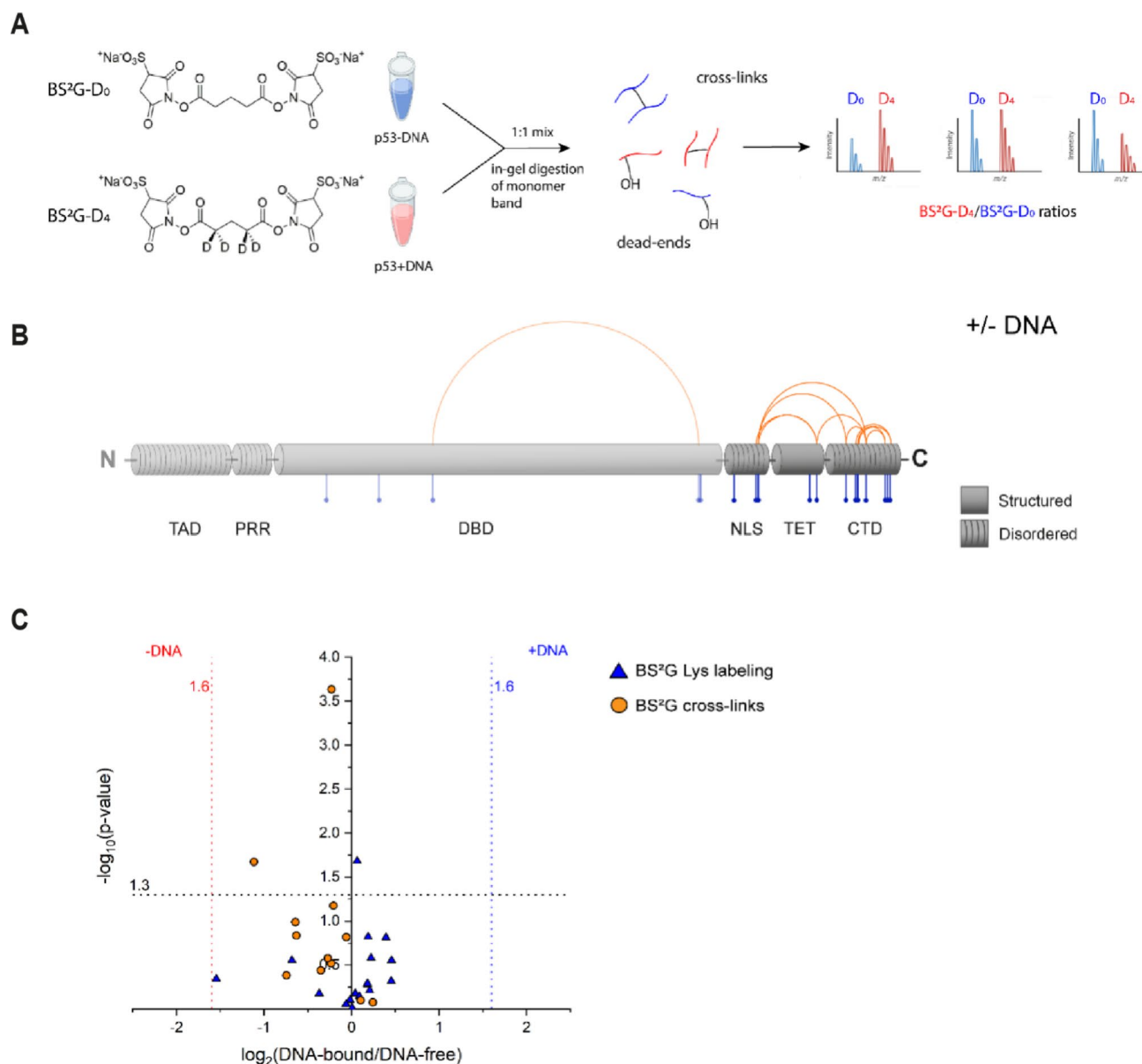


Figure 2. (A) Quantitative cross-linking of full-length p53 was performed with the lysine-reactive, isotope-labeled cross-linker pair BS²G-D₀/D₄ to reveal differences between DNA-free and DNA-bound p53. (B) Schematic overview of BS²G cross-linking (orange) and footprinting data (blue). Structured and disordered regions of p53 are indicated. (C) Comparison of cross-linking and footprinting data for DNA-free (left hand side) and DNA-bound (right hand side) p53. Vertical asymptotes represent the fold change thresholds for determining if a cross-linked or labeled peptide is enriched in the DNA-free or DNA-bound state of p53. The horizontal asymptote defines the statistical significance limit of the data points (p value < 0.05).

used to define distance restraints and 16,000 p53 tetrameric models were generated. The refinement strategy is shown in Fig. 3A.

Validation of the refined p53 structure. The generated models were validated with the experimentally determined cross-links (BS²G, 12 cross-links) in addition to previously published cross-links (DSBU; 21 cross-links)³⁹, and BS²G footprinting data. A stepwise scoring matrix was applied with a Ca-Ca distance threshold of 25 Å for BS²G and 30 Å for DSBU. Distances of 10 Å above these values were included to account for the high flexibility in the IDRs, based on the overall distance distribution (Supporting Information, Figure S34). In addition, footprinting data yield information on the labeling efficiency of specific lysine residues, depending on their reactivity and accessibility. The reactivity is encoded by the pK_a values of lysines and was calculated for each model with PropKa⁴⁴, while the accessibility can be described by the SASA and was calculated with the MSMS-tool⁴³. To access the distribution of pK_a and SASA values of each lysine residue, all models were plotted against the experimentally determined accessibilities of lysines (Supporting Information, Figure S35). Notably,

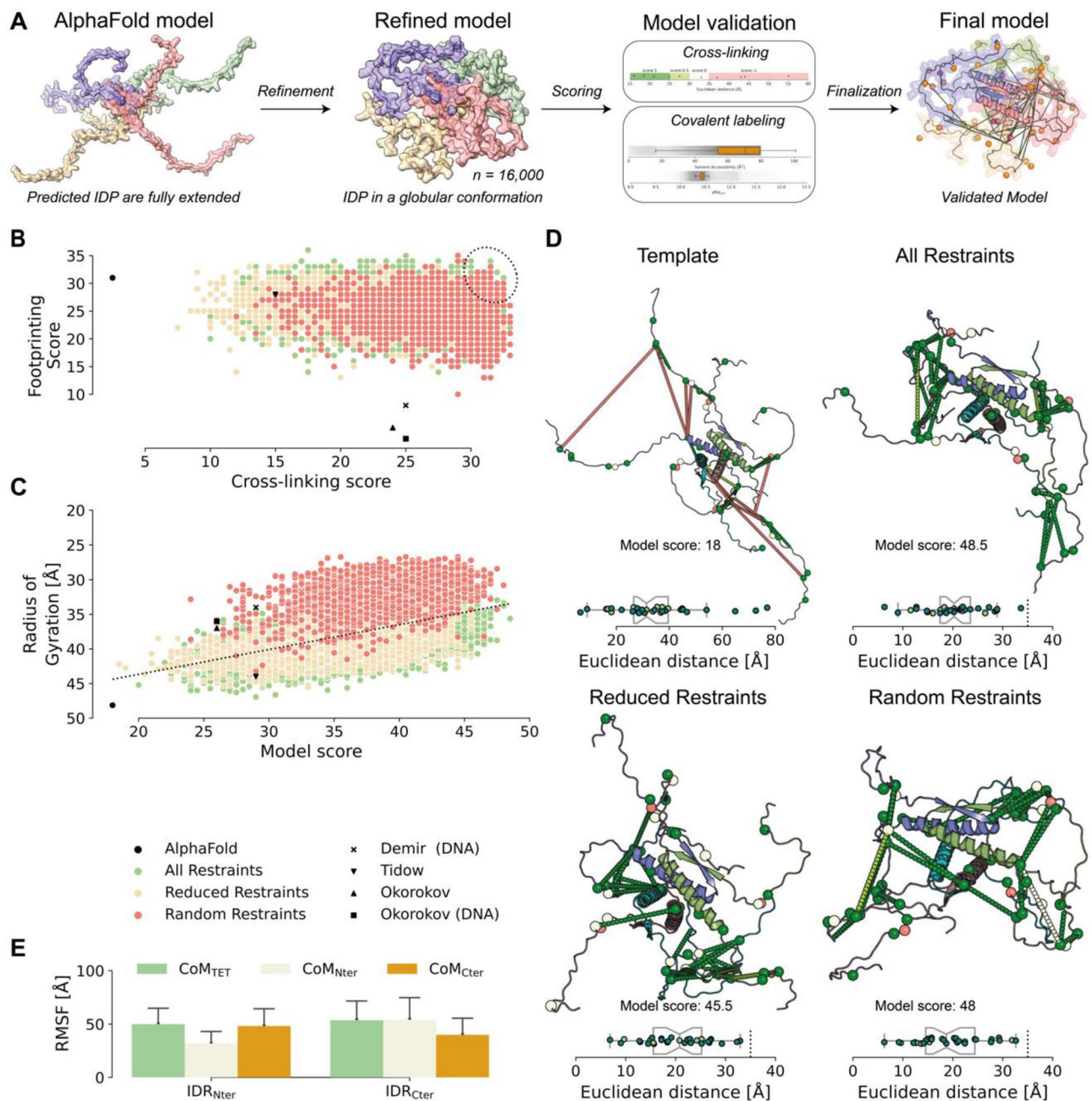


Figure 3. AI-based and knowledge-based remodeling of the disordered C-terminal region of p53. **(A)** Refinement workflow: the initial model was generated using AlphaFold2. The initial model was refined using experimentally and randomly generated restraints, and the 16,000 models generated were classified and validated using cross-linking and footprinting data. **(B)** Distribution of all models generated. The dotted ellipsoid marks the highest ranked models. **(C)** Variation of the radius of gyration (R_g) as a function of the model score. High compaction (low R_g) correlates with high scores of the models. **(D)** Presentation of input model (template) and top-ranked refined models (all restraints, reduced restraints, and random restraints). For details see “Model refinement” part. The dotted lines indicate cross-links, the spheres present labeled residues. The box-plots indicate Euclidian distances between the $C\alpha$ -atoms of cross-linked residues; BS²G cross-links (light green) and DSBUs cross-links (teal). **(E)** Structural variation of disordered p53 termini. The center of mass (CoM) of the ordered tetramerization domain (TET), as well as the disordered NLS (Nter)- and CTD (Cter) in each refined model were used to calculate the variations between the refined structures.

the pK_a value does not allow differentiating the results, while the SASA shows a population of lysine residues with reduced accessibility, correlating with footprinting data (Supporting Information, Table S4). For scoring, a stepwise rating matrix was applied that penalizes clear outliers (Supporting Information, Figure S35).

For model selection, XL-MS and footprinting data were plotted against each other, and the highest ranked solution of each refinement strategy was picked (Fig. 3B). A clear improvement of scoring, both in cross-link distance and footprinting data satisfaction, is visible when compared to the AlphaFold-generated input model and already published tetrameric p53 models (Fig. 3B), which clearly indicates an experimentally-driven optimization of computationally predicted AlphaFold models.

One striking observation is that the refined p53 models appear to be more compact than already existing models. To underpin this, we calculated the radius of gyration (R_g) for each model. There is a significant correlation between the scoring and the R_g of p53 (Fig. 3C; Supporting Information, Figure S36). Selecting and aligning the top-ranked models reveal an ordered and structurally preserved TET, but the neighboring IDRs exhibit various conformational states (Fig. 3D,E).

An identical or similar kind of compaction of the p53 tetramer has already been observed previously^{25,39,50}, but is not reflected in alternative models (Supporting Information, Figure S37)²⁰. Using our experimental data and our scoring, all previously proposed models of the C-terminal region perform worse in the overall scoring than our compact models (Fig. 3C).

We additionally validate our p53 tetrameric models with HDX-MS data. HDX-MS displays protein dynamics and changes in a time-dependent manner and complements XL-MS and footprinting data^{32,51,52}. In HDX, backbone amide protons of a polypeptide chain are limited in their H/D exchange if they are involved in hydrogen bonding networks. We mapped the deuterium incorporation kinetics of full-length p53 tetramer for the first time, yielding fast exchange rates for the IDRs, TAD, NLS, and CTD, and slow exchange rates for regions with known secondary structures, DBD and TET (Fig. 4A). Plotting the accessibilities of the backbone amide protons in the AlphaFold2 template as well as in the previously proposed models (Supporting Information, Figure S1, Demir and Tidow models) and top-ranked refined models (Fig. 3D), the expected correlation with deuterium incorporation in HDX was confirmed (Fig. 4B; Supporting Information, Figure S37).

Discussion

The majority of published p53 structures^{12,15,16} as well as proposed models^{19–21,25,26} either lack information about the CTD (residues 363–393) or are inconsistent among each other. This leads to inherent difficulties in interpreting experimental data regarding the CTD, especially in the functional and cellular context of tetrameric p53. By combining complementary techniques of structural MS with computational modeling we were able to create a conformational ensemble for C-terminal region of tetrameric human wild-type p53 (TET, NLS and CTD), satisfying the experimental data.

Computational modeling. The default AlphaFold2 model generation includes template identification, structure prediction, and AMBER relaxation⁵³. The AlphaFold2 relaxation does not have a fixed time scale, but a scoring scale. If the model fulfils certain criteria, MD is stopped. Although this approach works very well in the context of well-structured, globular proteins we encountered challenges by generating models for IDPs

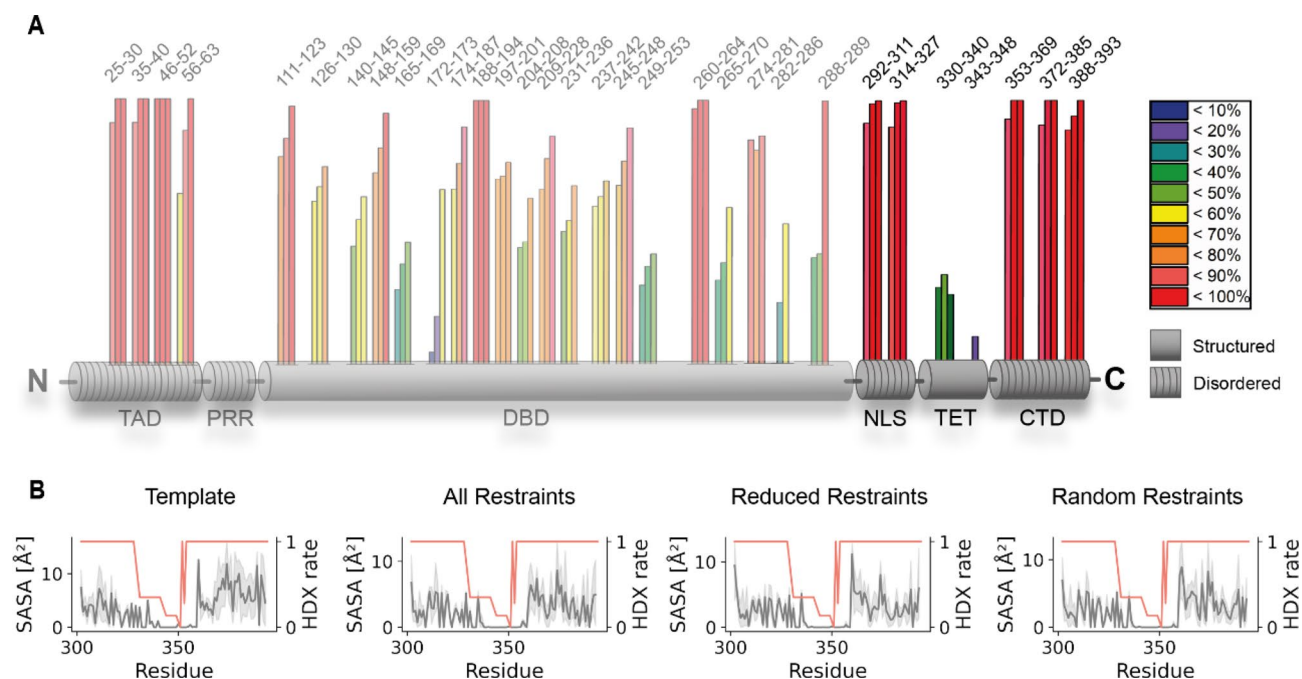


Figure 4. (A) HDX-MS of backbone amide protons in DNA-free p53. Three bars (HDX after 30 s, 1 min, and 2 min) are presented for each peptide analyzed; amino acid numbers are shown above. Data from NLS, TET, and CTD confirm the top-ranked refined models. (B) Models (Fig. 3D) show a high correlation between SASA (grey) and HDX (orange). The disordered and accessible nature of p53's IDRs is reflected by high SASA and HDX rates, while the structured TET (residues 326–356) exhibits low SASA and HDX rates.

with three major key factors. These include (a) lack of template structures due to their disordered nature, (b) the potential impact of the choice of force field and the type of water solvation, although these limitations are negligible within AlphaFold2's refinement timescale, and (c) AlphaFold2's training data predominantly consisting of ordered structures from the PDB, leading to a bias towards ordered structures in its predictions. In addition, another important aspect is the hypervariability of IDRs that lead to low-confidence AlphaFold2 models⁵⁴. Therefore, a refinement of IDP/IDR models generated with AlphaFold2 that are based on experimental distance restraints is inevitable. In addition to generating distance restraints for computational modeling, protein footprinting adds another layer of validation as only solvent-exposed residues will be labeled efficiently. Notably, local pK_a values alone are not suitable for explaining different labeling efficiencies of lysine residues in p53, which is most likely due to the solvent-exposed nature of lysines in the disordered regions. Although these lysines are buried in a subset of derived models they are apparently not involved in ionic interactions or extended hydrogen bonding networks and have therefore similar pK_a values. This is reasonable for IDRs as lysine residues that form before-mentioned interactions will likely be part of an ordered structure, while buried lysines with unchanged pK_a values will most likely be located in an IDR. The final models show a compact shape of p53, which is in agreement with all input data. The disordered character of p53 is recapitulated in the assembly of selected models as no clear conformation is visible and each model has a unique localization of the IDRs (Fig. 3D,E). Interestingly, even without similar conformations, the radii of gyration are similar among the highest-ranked models. This could have an implication on how IDRs behave in solution, at least those of p53. As the IDRs are apparently part of the same conformational ensemble, independently of RE-DNA binding, no secondary structure or domain folding was assumed. This scenario might be completely different for IDRs that undergo conformational changes upon binding or during substrate channeling, e.g. in pyruvate dehydrogenase complex⁵⁵.

Compaction of p53 tetramer. A major finding of our study is the compaction of the whole NLS-TET-CTD arrangement in tetrameric full-length, wild-type human p53, which was initially indicated by our experimental data and is also mirrored in our generated models. It has to be noted that cross-links will capture conformational states depending on the cross-linker's spacer length and its reactivity. We cannot rule out that only a fraction of conformations is captured, while others will remain elusive. To extend the coverage of the entire conformational space, additional cross-linking principles with other chemistries and spacer lengths should be employed.

The first published structural model by Okorokov et al.¹⁹ was generated by cryo-EM utilizing murine p53, while all other studies were performed with human p53^{56,57} (Supporting Information, Figure S1B,D). The first human-based, tetrameric p53 structural model derived by the Fersht group²⁰ differed dramatically in the DNA-free (extended cross-shape, Supporting Information, Figure S1A) and DNA-bound form (compact basket shape, Supporting Information Figure S1C). The DNA-bound tetrameric p53 model was further refined by a follow-up study utilizing negative stain EM²¹. Here, different conformational ensembles (classes) were observed, which already gave hints of inter-domain proximities of CTD and NLS, as well as DBD and the DNA itself. Single-molecule FRET (smFRET) experiments from the same study also showed a spatial proximity between the very C terminus of p53 and the L1-loop (residues 112–124) within the DBD in presence of RE-DNA. However, in the absence of DNA, an equilibrium might exist between the formerly published, extended cross-shaped p53 tetramer²⁰ and a more compact arrangement as observed for the DNA-bound p53 tetramer²¹. The existence of an equilibrium state is controversially discussed due to the limited distance information of smFRET data as only one single labeled construct was used. Also, a spatial rearrangement of the flexible CTD was not discussed, but only a complete rearrangement of the whole p53 tetramer²¹. In contrast, our data show that in both cases, DNA-free and DNA-bound states, a more compact p53 structure is present. Another publication agreeing with our findings of a more compact p53 tetramer, both in the DNA-bound and DNA-free states, is based on ion mobility coupled to MS⁵⁰. Here, experimentally derived collisional cross sections for DNA-free, full-length p53 and DNA-bound p53 are smaller than the values calculated for the extended model structures^{20,50}. Also, collisional cross sections were found to be more similar between both states (\pm DNA) than expected from the models. At that time, the observed compaction was described as a gas phase collapse, which can be excluded in our findings as XL-MS, footprinting, and HDX-MS data were exclusively collected in solution at near physiological buffer conditions. In the most recent MD study of a reconstructed p53 tetramer in complex with RE-DNA a compaction and even a contact of the CTD to the bound DNA molecule is described, supporting our findings (Supporting Information, Figure S1E)²⁵. Together, all abovementioned p53 structural studies show similar results to our MS-based data and agree with our p53 models.

Modes of DNA binding. Initially the CTD was characterized as negative regulator impairing specific DNA binding^{8,58,59}. Later, it was recognized to be a positive regulator being responsible for various functions of the p53 tetramer related to DNA binding^{60,61}. The CTD has been described to facilitate p53 binding to long non-specific DNA molecules, as well as to chromatin⁶², enabling lateral movement along the DNA via electrostatic interactions. This however does not seem to be entirely true as the CTD might be involved in specific DNA binding as well⁶³. A two-mode binding model switching between a sliding- and recognition-mode was introduced, termed S- and R-modes^{47,64}. This two-mode DNA-binding model suggests a hemi-specific binding, with one dimer of tetrameric p53 binding to DNA in a site-specific manner, while the other p53 dimer binds non-specifically to noncognate DNA transitioning into a fully specific, DNA-bound state upon response element recognition⁴⁷. If we reflect our data in the light of that model (S- and R-modes), we obtain a snapshot even for specific DNA binding. We captured a conformational ensemble for one specific 26-bp RE-DNA (one sequence, one length, see also Model refinement section) bound to p53, which might reflect the recognition mode (R-mode). A comparison of that two-mode DNA-binding model proposed by Tafvizi et al.⁴⁷ with the EM-based model of Melero

et al.²¹ revealed the following: in the latter model, EM maps were grouped into classes that depend not only on the sequence of the RE-DNA, but also on the length of the individual DNA fragment^{18-mers/44-mers/60-mer}. Also, the number of p53 conformational states observed differed depending on the nature of the bound DNA-fragment. In our study, we used an intermediate length (26-mer) RE-DNA with high specificity and affinity towards the DBD^{50,65}. This is a strong hint that we indeed captured one of the specific DNA-binding modes, in which the CTD of p53 might be involved to stabilize the R-mode. Our findings are also in line with the reaction kinetics of the cross-linking and labeling reagents used herein that capture dynamics on the minutes to seconds time scale. We are therefore confident that our XL-MS setup captures a conformational ensemble of full-length p53 bound to a specific RE-DNA in the R-mode.

There are however still discrepancies whether CTD and DBD are both binding to DNA, interdependently of each other, or if the CTD controls site-specific binding of the DBD to the DNA by inducing conformational changes within the DBD itself^{26,63}. If there were conformational changes upon DNA-binding on the sub-second time scale, like a channeling of the DNA from the CTD to the DBD, faster reaction kinetics available with photo cross-linkers (diazirines, incorporated photo-amino acids) would be needed^{66,67}. To capture the conformational dynamics of the highly complex binding events between full-length p53 and DNA, integrative approaches are inevitable. One technique alone will not be able to grasp the full picture, but structural MS combined with computational modeling allows getting insights into the high complexity of p53's conformational ensembles.

Conclusion and outlook

As it becomes increasingly obvious that a single study will never be powerful enough to completely unravel the whole mechanistic machinery of p53 tetramerization, DNA-binding, and movement along the bound DNA, integrative approaches are required for a better understanding. We contribute to this tremendous task by delivering a workflow for the in-solution investigation of full-length, wild-type human p53 with a consecutive structural modeling pipeline. Our data show that the structure of the C-terminal region (NLS-TET-CTD) of full-length p53 tetramer is more compact than perceived by previous models. Moreover, even the DNA-free form is more compact and conformationally unchanged upon specific binding of a 26-mer RE-DNA. Our data most likely reflect the recognition mode (R-mode) of p53-DNA binding agreeing with previous data^{25,47}. We are planning to apply our workflow to a variety of different DNA constructs varying in length as well as p53 specificity and affinity to contribute to the understanding of the underlying mechanisms for the transcriptional regulation of p53 by also capturing the sliding-mode (S-mode). The use of photo-reactive cross-linkers as well as HDX-MS will aid in resolving the complex dynamics upon p53 binding to different DNA molecules in addition to the already established distance restraints and labeling approach.

Data availability

All materials described in the manuscript, including all relevant raw data, will be freely available to any researcher wishing to use them for non-commercial purposes, without breaching participant confidentiality. All MS data are available via ProteomeXchange with identifier PXD037030.

Received: 16 February 2023; Accepted: 18 May 2023

Published online: 25 May 2023

References

- Joerger, A. C. & Fersht, A. R. Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.* **77**, 557–582 (2008).
- Joerger, A. C. & Fersht, A. R. The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. *Annu. Rev. Biochem.* **85**, 375–404 (2016).
- Ko, L. J. & Prives, C. p53: Puzzle and paradigm. *Genes Dev.* **10**(9), 1054–1072 (1996).
- Lane, D. P. Cancer p53, guardian of the genome. *Nature* **358**(6381), 15–16 (1992).
- Panca, R. & Tompa, P. Structural disorder in eukaryotes. *PLoS One* **7**(4), e34687 (2012).
- Peng, Z. *et al.* Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol. Life Sci.* **72**(1), 137–151 (2015).
- Piersimoni, L. *et al.* Lighting up Nobel Prize-winning studies with protein intrinsic disorder. *Cell Mol. Life Sci.* **79**(8), 449 (2022).
- Anderson, M. E., Woelker, B., Reed, M., Wang, P. & Tegtmeyer, P. Reciprocal interference between the sequence-specific core and nonspecific C-terminal DNA binding domains of p53: Implications for regulation. *Mol. Cell Biol.* **17**(11), 6255–6264 (1997).
- Weinberg, R. L., Veprintsev, D. B. & Fersht, A. R. Cooperative binding of tetrameric p53 to DNA. *J. Mol. Biol.* **341**(5), 1145–1159 (2004).
- Schlereth, K. *et al.* DNA binding cooperativity of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol. Cell* **38**(3), 356–368 (2010).
- Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.* **9**(5), 402–412 (2008).
- Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **265**(5170), 346–355 (1994).
- Malecka, K. A., Ho, W. C. & Marmorstein, R. Crystal structure of a p53 core tetramer bound to DNA. *Oncogene* **28**(3), 325–333 (2009).
- Jeffrey, P. D., Gorina, S. & Pavletich, N. P. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* **267**(5203), 1498–1502 (1995).
- Clore, G. M. *et al.* Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nat. Struct. Biol.* **2**(4), 321–333 (1995).
- Edamzadah, S., Tropia, L. & Halazonetis, T. D. Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. *Mol. Cancer Res.* **9**(11), 1493–1499 (2011).
- Ethayathulla, A. S., Nguyen, H. T. & Viadiu, H. Crystal structures of the DNA-binding domain tetramer of the p53 tumor suppressor family member p73 bound to different full-site response elements. *J. Biol. Chem.* **288**(7), 4744–4754 (2013).
- Masahiro Nishimura, Y. T., Kayo, N. & Hitoshi, K. Structural basis for p53 binding to its nucleosomal target DNA sequence. *PNAS Nexus* **1**, 1–9 (2022).

19. Okorokov, A. L. *et al.* The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *EMBO J.* **25**(21), 5191–5200 (2006).
20. Tidow, H. *et al.* Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. USA* **104**(30), 12324–12329 (2007).
21. Melero, R. *et al.* Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. *Proc. Natl. Acad. Sci. USA* **108**(2), 557–562 (2011).
22. Aramayo, R. *et al.* Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance. *Nucleic Acids Res.* **39**(20), 8960–8971 (2011).
23. Terakawa, T. & Takada, S. p53 dynamics upon response element recognition explored by molecular simulations. *Sci. Rep.* **5**, 17107 (2015).
24. Singh, S. K. *et al.* Structural visualization of the p53/RNA polymerase II assembly. *Genes Dev.* **30**(22), 2527–2537 (2016).
25. Demir, O., Jeong, P. U. & Amaro, R. E. Full-length p53 tetramer bound to DNA and its quaternary dynamics. *Oncogene* **36**(10), 1451–1460 (2017).
26. D'Abramo, M. *et al.* The p53 tetramer shows an induced-fit interaction of the C-terminal domain with the DNA-binding domain. *Oncogene* **35**(25), 3272–3281 (2016).
27. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**(7873), 583–589 (2021).
28. Mihaly Varadi, S. A. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **20**, 21 (2021).
29. Hannah, M., Britt, T. C. & Konstantinos, T. Integration of mass spectrometry data for structural biology. *Chem. Rev.* **20**, 21 (2021).
30. Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E. & Grandori, R. Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. *Proteomics* **19**(6), e1800060 (2019).
31. Beveridge, R., Migas, L. G., Kriwacki, R. W. & Barran, P. E. Ion mobility mass spectrometry measures the conformational landscape of p27 and its domains and how this is modulated upon interaction with Cdk2/cyclin A. *Angew. Chem. Int. Ed. Engl.* **58**(10), 3114–3118 (2019).
32. Balasubramaniam, D. & Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta* **1834**(6), 1202–1209 (2013).
33. Piersimoni, L., Kastritis, P. L., Arlt, C. & Sinz, A. Cross-linking mass spectrometry for investigating protein conformations and protein-protein interactions horizontal line A method for all seasons. *Chem. Rev.* **122**(8), 7500–7531 (2022).
34. Kitayner, M. *et al.* Structural basis of DNA recognition by p53 tetramers. *Mol. Cell.* **22**(6), 741–753 (2006).
35. Hyunjung Kim, K. K., Jongkyu, C., Kyu, H., Hwa, J. B. & Robert, G. Roeder and Woojin An p53 requires an Intact C-terminal domain for DNA binding and transactivation. *J. Mol. Biol.* **20**, 11 (2011).
36. Richard, R. Rust DMBDJW. Structure of the negative regulatory domain of p53 bound to S100B($\beta\beta$). *Nat. Struct. Mol. Biol.* **25**, 5 (2000).
37. Arlt, C., Ihling, C. H. & Sinz, A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. *Proteomics* **15**(16), 2746–2755 (2015).
38. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *BioRxiv* **20**, 21 (2021).
39. Arlt, C. *et al.* An integrated mass spectrometry based approach to probe the structure of the full-length wild-type tetrameric p53 tumor suppressor. *Angew. Chem. Int. Ed. Engl.* **56**(1), 275–279 (2017).
40. Gotze, M., Iacobucci, C., Ihling, C. H. & Sinz, A. A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions. *Anal. Chem.* **91**(15), 10236–10244 (2019).
41. Daniel, S. Z., Vladimir, S., & David, C. S. Quantitative analysis of protein covalent labeling mass spectrometry data in the mass spec studio. 2019.
42. Webb, B. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2199**, 239–255 (2021).
43. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**(3), 305–320 (1996).
44. Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**(2), 525–537 (2011).
45. Uversky, V. N. p53 Proteoforms and intrinsic disorder: An illustration of the protein structure-function continuum concept. *J. Mol. Sci.* **20**, 16 (2016).
46. Meek, D. W. & Anderson, C. W. Posttranslational modification of p53: Cooperative integrators of function. *Cold Spring Harb. Perspect. Biol.* **1**(6), a000950 (2009).
47. Tafvizi, A., Huang, F., Fersht, A. R., Mirny, L. A., & van Oijen, A. M. A single-molecule characterization of p53 search on DNA PNAS. 2011.
48. Madler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NHS esters: A systematic study on amino acid reactivities. *J. Mass Spectrom.* **44**(5), 694–706 (2009).
49. Birgit Schilling, R. H. R., & Bradford W. G. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. 2003.
50. Pagel, K., Natan, E., Hall, Z., Fersht, A. R. & Robinson, C. V. Intrinsically disordered p53 and its complexes populate compact conformations in the gas phase. *Angew. Chem. Int. Ed. Engl.* **52**(1), 361–365 (2013).
51. Ozohanics, O. & Ambrus, A. Hydrogen-deuterium exchange mass spectrometry: A novel structural biology approach to structure, dynamics and interactions of proteins and their complexes. *Life (Basel)* **10**, 11 (2020).
52. Trabjerg, E. *et al.* Conformational analysis of large and highly disulfide-stabilized proteins by integrating online electrochemical reduction into an optimized H/D exchange mass spectrometry workflow. *Anal. Chem.* **87**(17), 8880–8888 (2015).
53. Romelia Salomon-Ferrer, D. A. C. & Ross, C. W. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **20**, 12 (2012).
54. Kiersten, M. & Ruff, R. V. P. AlphaFold and implications for intrinsically disordered proteins. *JMB.* 2021.
55. Kyrilis, F. L. S. D. *et al.* Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. *Cell Rep.* **20**, 21 (2021).
56. Fischer, M. Conservation and divergence of the p53 gene regulatory network between mice and humans. *Oncogene* **20**, 19 (2019).
57. Fischer, M. Mice are not humans: The case of p53. *Trends Cancer* **20**, 20 (2020).
58. Hupp, T. R. & Lane, D. P. Allosteric activation of latent p53 tetramers. *Curr. Biol.* **19**, 94 (1994).
59. Prives, L. J. A. C. Activation of p53 sequence-specific DNA binding by short single strands of DNA requires the p53 C-terminus. *Cell* **19**, 95 (1995).
60. Linghu Nie, M. S. & Carl, G. M. Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination. *J. Biol. Chem.* **20**, 55 (2007).
61. Woojin An, J. K., & Robert, G. R. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell*; 2004.
62. Espinosa, J. M. & Emerson, B. M. Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed cofactor recruitment. *Mol. Cell.* **20**, 1 (2001).

63. Laptenko, O. *et al.* The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain. *Mol. Cell.* **57**(6), 1034–1046 (2015).
64. Anahita Tafvizi, F. H., Jason, S. L., Alan, R. F., Leonid, A. M. & van Oijen, A. M. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **25**, 25 (2008).
65. Fersht, D. B. V. A. A. R. Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA. *Nucleic Acids Res.* **25**, 45 (2008).
66. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein–protein interactions in living cells. *Nat. Methods* **2**(4), 261–267 (2005).
67. Iacobucci, C. *et al.* Carboxyl-photo-reactive MS-cleavable cross-linkers: Unveiling a hidden aspect of diazirine-based reagents. *Anal. Chem.* **90**(4), 2805–2809 (2018).

Acknowledgements

AS acknowledges financial support by the DFG (RTG 2467, project number 391498659 “Intrinsically Disordered Proteins-Molecular Principles, Cellular Functions, and Diseases”, INST 271/404-1 FUGG, INST 271/405-1 FUGG, and CRC 1423, project number 421152132), the Federal Ministry for Economic Affairs and Energy (BMW, ZIM project KK5096401SK0), the region of Saxony-Anhalt, and the Martin Luther University Halle-Wittenberg (Center for Structural Mass Spectrometry). CI acknowledges financial support by European Union—NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041—VITALITY—CUP E13C22001060006.

Author contributions

A.D.I. and M.K. performed experiments, C.T. performed computational modeling, M.K., C.H.I., C.I., and J.K. analyzed data, C.A., P.L.K., and A.S. supervised the study, C.T., C.A., and A.S. wrote the manuscript with input from all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35437-5>.

Correspondence and requests for materials should be addressed to C.A. or A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023