EDGE ARTICLE

Check for updates

Characterization of the T4 gp32–ssDNA complex by native, cross-linking, and ultraviolet photodissociation mass spectrometry†

Molly S. Blevins, [a] Jada N. Walker, [a] Jeffrey M. Schaub, [b] Ilya J. Finkelstein [b]
and Jennifer S. Brodbelt *[a]

Protein–DNA interactions play crucial roles in DNA replication across all living organisms. Here, we apply a suite of mass spectrometry (MS) tools to characterize a protein-ssDNA complex, T4 gp32·ssDNA, with results that both support previous studies and simultaneously uncover novel insight into this non-covalent biological complex. Native mass spectrometry of the protein reveals the co-occurrence of Zn-bound monomers and homodimers, while addition of differing lengths of ssDNA generates a variety of protein:ssDNA complex stoichiometries (1 : 1, 2 : 1, 3 : 1), indicating sequential association of gp32 monomers with ssDNA. Ultraviolet photodissociation (UVPD) mass spectrometry allows characterization of the binding site of the ssDNA within the protein monomer *via* analysis of holo ions, *i.e.* ssDNA-containing protein fragments, enabling interrogation of disordered regions of the protein which are inaccessible *via* traditional crystallographic techniques. Finally, two complementary cross-linking (XL) approaches, bottom-up analysis of the crosslinked complexes as well as MS1 analysis of the intact complexes, are used to showcase the absence of ssDNA binding with the intact cross-linked homodimer and to generate two homodimer gp32 model structures which highlight that the homodimer interface overlaps with the monomer ssDNA-binding site. These models suggest that the homodimer may function in a regulatory capacity by controlling the extent of ssDNA binding of the protein monomer. In sum, this work underscores the utility of a multi-faceted mass spectrometry approach for detailed investigation of non-covalent protein-DNA complexes.

Introduction

Protein–DNA interactions are fundamental for transcriptional regulation, DNA repair, and replication.[1,2] The genome is organized in double-stranded DNA (dsDNA); however, genetic information is accessed for transcription or replication as a single-stranded DNA (ssDNA) intermediate. ssDNA is inherently less stable and must be protected to preserve genomic integrity. All organisms, from viruses to humans, have evolved single-stranded DNA binding proteins (SSBs) to protect and stabilize ssDNA intermediates. SSBs bind ssDNA independently of sequence and with high affinity, and are critical for the sequestration and stabilization of ssDNA, which occur in preparation for DNA replication and DNA repair processes.[3–5] The bacteriophage T4 gene protein 32 (also known as T4 gp32)

is a prototypical member of this family and is required for viral replication.[3,6]

Gp32 is divided into two subdomains, subdomain I and subdomain II, which are linked through a connecting region.[7] An oligonucleotide/oligosaccharide-binding (OB) fold forms the structural core.[7] An electropositive cleft consisting of aromatic and basic residues within the connecting region facilitates non-specific ssDNA binding.[7–9] Gp32 discriminates between ssDNA and dsDNA based on the hydrophobic interactions that are created between a pocket of aromatic side chains of gp32 within the electropositive cleft and the bases of a ssDNA substrate.[9] An X-ray crystal structure of gp32 revealed a zinc finger consisting of His64, Cys77, Cys87 and Cys90 within the C-terminal tail.[7] This motif maintains the subdomain I structure.[7] To bind to ssDNA, the acidic C-terminal domain of the gp32 protein must undergo a conformational change that exposes the positively-charged region of its core domain, which in turn, interacts with the negatively charged ssDNA backbone.[10,11]

Gp32, and all other SSBs, bind ssDNA transiently.[9] In the apo (unbound) state, gp32 is a mixture of monomers and dimers, owing largely to the inhibitory effect of the C-terminal domain on the binding cleft.[9] Gp32 binds ssDNA in a cooperative manner, in which the binding of one gp32 molecule increases

[a]*Department of Chemistry, University of Texas at Austin, Austin, TX 78712, USA. E-mail: jbrodbelt@cm.utexas.edu*

[b]*Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA*

the binding affinity such that additional gp32 monomers bind to the ssDNA substrate in a contiguous process.[9,12] These findings indicate that gp32 binding to ssDNA likely occurs *via* self-dissociation of gp32 dimers and subsequent monomer-by-monomer binding.[12,13] Subdomain II (N-terminal region) facilitates cooperative binding on ssDNA *via* electrostatic interactions between the core domains of adjacent gp32 molecules.[9,11]

Here, we re-examine gp32 self-assembly and ssDNA interactions *via* a suite of advanced mass spectrometry (MS) methods.[14–17] Electrospray ionization (ESI) facilitates the transfer of intact proteins and protein–DNA complexes from the solution into the gas phase for subsequent mass spectrometric interrogation, in many cases preserving both the structure and stoichiometry of the non-covalently bound complexes.[18,19] Many studies have exploited ESI-MS as a means to preserve and facilitate analysis of native interactions, including DNA–small molecule ligand complexes,[20–23] DNA-templated silver clusters,[24–27] DNA duplex and quadruplex complexes,[28–30] and protein–DNA complexes.[31–39] Native ESI-MS studies of DNA complexes have enabled the determination of DNA–ligand stoichiometry,[32,38] ligand selectivity,[21] and DNA-binding pathways.[35] Analyses of native complexes such as these are contingent upon the exceptional performance of high-resolution and high-mass accuracy mass spectrometers.[40] UV-induced cross-linking MS has been developed for interrogation of protein–DNA complexes *via* bottom-up LC-MS/MS analysis of the DNA–peptide crosslinks after proteolytic digestion.[41] In general, compared to ESI-MS of multiprotein complexes, fewer studies have reported the analysis of protein–DNA complexes.[31–39] Analysis of protein–nucleic acid complexes containing large DNA or RNA strands (>20 nucleotides) in positive-ion mode typically generates spectra that are complicated by the presence of highly heterogeneous ion populations arising from cation adduction.[42] However, the addition of volatile salts (*e.g.*, ammonium acetate) minimizes the prevalence of cation adduction for protein–DNA complexes.[43,44]

While single-stage mass spectrometry (MS1) experiments of native DNA–protein complexes in volatile salt solutions provide basic mass information and binding stoichiometry, MS/MS, often *via* collisionally activated dissociation (CAD), is typically used to glean more detailed structural information. The use of MS/MS for structural interrogation of oligonucleotide-containing complexes has emerged as useful for a few studies focused on ligand localization and conformational changes.[22,26,28,31,34,45] However, in the case of non-covalent complexes, collisional activation methods, although providing some sequence information, do not afford extensive information related to ligand localization and overall complex structure. To combat these deficits, alternative activation methods that are more suitable for the analysis of non-covalent macromolecular complexes have been developed, including surface-induced dissociation (SID),[46] electron activation methods,[47] and ultraviolet photodissociation (UVPD).[48] The latter uses high-energy UV photons to promote cleavages along multiple positions of the protein backbone, generating a diverse series of fragment ions including a, $a + 1$, b, c, x, $x + 1$, y, $y − 1$, and z ions.[49,50] Through access to these higher-energy fragmentation pathways, UVPD yields both ligand-containing (holo) ions and ligand-free (apo) product ions, thus enabling the determination of ligand-binding sites and revealing conformational reorganization.[51,52]

Complementary to the information generated *via* UVPD of native protein-ligand complexes is the insight that can be obtained from cross-linking mass spectrometry (XL-MS). XL-MS has been previously utilized to unveil protein–protein interaction networks and to monitor conformational changes of multimeric complexes.[53,54] The covalent linkages created by crosslinking allows mapping of protein–protein binding interfaces,[55] especially when used in conjunction with molecular docking software.[56] Crosslinking of non-covalent protein complexes can be analyzed using traditional bottom-up approaches to gain distance information between neighboring residues, or *via* analysis of the intact crosslinked complexes to decipher the overall complex stoichiometry of crosslinked subunits.

Herein, we showcase the use of native UVPD-MS data to characterize viral protein–ssDNA complexes comprised of bacteriophage T4 gene product 32 (gp32) and ssDNA substrates (dT12, dT20). Complementary information gathered from native UVPD-MS and XL-MS unravel the unique characteristics of gp32's ssDNA binding mechanism. Using these approaches, we establish with individual amino acid precision how gp32 monomers bind a ssDNA substrate. In addition, we show the non-sequence specific gp32 binding to ssDNA. Integrating insight obtained from bottom-up methods and MS1 analysis of intact crosslinked complexes demonstrates that gp32 does not bind ssDNA as a multimer (*e.g.* dimer or trimer), and rather dissociates into monomers prior to ssDNA binding. Bottom-up XL-MS enables generation of two model dimer structures in which the ssDNA-binding region of each monomer overlaps with the homodimer interface. This work substantiates native MS, UVPD-MS and XL-MS as complementary techniques for probing protein–nucleic acid interactions with unprecedented structural resolution.

Experimental

Materials and reagents

Gp32 (pIF89) and gp32-ΔCTD (amino acids 1–254, pIF898) were cloned with a C-terminal intein–chitin binding domain. pIF89 was transformed into BL21 ArcticExpress *E. coli* cells (Agilent) and supplemented with 50 μg mL^{-1} carbenicillin and 20 μg mL^{-1} gentamycin and grown at 30 °C. Cells were induced with 0.5 mM IPTG and cultured overnight at 12 °C. Resulting pellets were resuspended in resuspension buffer (50 mM Tris–HCl pH 7.5, 500 mM NaCl, 1 mM EDTA, 10% sucrose (w/v), 1 mM PMSF) and sonicated to lyse. Lysate was loaded on 5 mL chitin resin (NEB) equilibrated with buffer A (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA). Resin was extensively washed with buffer B (50 mM Tris pH 7.5, 1 M NaCl, 1 mM EDTA). Intein cleavage was performed on resin by incubating with buffer B supplemented with 50 mM DTT overnight at 4 °C. Resulting elution was concentrated with a 10 kDa MWCO concentrator (Amicon) and dialyzed in storage buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol (v/v)). Protein was flash frozen in liquid nitrogen

and stored at -80 °C. Final protein molecular weight is approximately 33.5 or 28.5 kDa based on SDS-PAGE results (Fig. S1†). The molecular weight results obtained from SDS-PAGE depend not only on the protein theoretical molecular weight but also on amino acid composition,[57] and are in agreement with gp32 SDS-PAGE results provided by various T4 gp32 protein vendors.

Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA). LC solvents including LC-MS-grade water, formic acid, and acetonitrile were acquired from Sigma-Aldrich (St. Louis, MO). Bis(sulfosuccinimidyl)suberate (BS3), DMTMM, MS-grade trypsin, formic acid (99.5+%) and Pierce™ C18 Spin Columns utilized for bottom-up XL-MS sample clean-up were purchased from Thermo-Fisher Scientific (Waltham, MA). LC analytical columns (15 cm, 75 μm inner diameter) containing C18 stationary phase (3 μm diameter) were packed in-house. Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad Laboratories Inc., Hercules, CA) were used for desalting, buffer exchange, and size exclusion chromatography (SEC).

Sample preparation

For native MS, gp32 and gp32-ΔCTD solutions were diluted to 10 μM in 50 mM ammonium acetate (AmAc), whereas denatured samples were diluted to 10 μM in 50/50 acetonitrile/water with 0.1% formic acid. Similarly, dT12 and dT20 strands were diluted to 10 μM in water and added to the gp32- and gp32-ΔCTD-AmAc solutions for an incubation period of approximately 10 minutes at 25 °C. The resulting gp32 and gp32 + dT solutions were desalted, buffer exchanged, and subjected to SEC using 6 kDa SEC filters. For XL-MS experiments, gp32 was reconstituted to 0.1 mM into a $1\times$ phosphate buffered saline solution at pH 7.2. 10 mM BS3 or 10 mM DMTMM stock was diluted to 5 mM in water and then allowed to react with gp32 at protein/crosslinker molar ratio of 1 : 10 (1 hour incubation at 25 °C) for the samples subjected to proteolytic digestion and bottom-up LCMS/MS analysis, or 1 : 100 (10 minutes incubation at 25 °C) for the MS1 analysis of the intact complexes.[58,59] For the MS1 analysis of the intact complexes, the crosslinking reactions were quenched with 0.5% formic acid after the 10 minute cross-linking reactions to avoid excessive crosslinking. Excess (unreacted) crosslinker and monomer were removed by passing the reaction solution through 30 kDa SEC filters. The cross-linked samples prepared for bottom-up analysis were quenched with AmAc in $50\times$ excess of the cross-linker. In preparation for trypsin digestion, these samples were further diluted with 150 mM ammonium bicarbonate. Trypsin was added at a protein/protease molar ratio of 1 : 40 and incubated for 16 h at 37 °C.[58] Prior to LC separation, samples were cleaned up with C18 spin columns, dried with a SpeedVac, and reconstituted in 2% acetonitrile.

Direct infusion experiments

Equimolar (~10 μM each) gp32 or gp32-ΔCTD + ssDNA solutions were loaded into Au/Pd-coated nanospray borosilicate static tips (prepared in-house) for nESI. A heated capillary set to 200 °C was used to desolvate the protein–DNA complexes,

aiding their transmission into the gas phase. All direct infusion experiments were conducted on a Thermo Scientific Q Exactive UHMR mass spectrometer customized for the implementation of UVPD as described earlier,[60] except for the MS/MS experiments undertaken on denatured proteins which were performed on a Thermo Orbitrap Elite mass spectrometer, also equipped with a 193 nm excimer laser for UVPD as previously described.[61] The robust sensitivity of the UHMR mass spectrometer to ions of high m/z and its optimized optics for the retention of electrostatic complexes were ideal for analysis of intact complexes and afforded information about the heterogeneity of protein complexes with high molecular weights.[62–64] Modifications made to the higher-energy collisional dissociation (HCD) cell allowed the implementation of UVPD using a 193 nm ArF excimer laser (Excistar, Coherent, Santa Cruz, CA). Modulation of the injection flatapole and interflatapole voltages allowed in-source trapping (IST), enabling front-end collisional activation that improved analysis of native proteins and protein–DNA complexes.[60,65–68] MS1 analysis of intact cross-linked proteins were analyzed on the UHMR instrument. MS1 spectra were collected at $R = 1563$ and averaged over 25 scans. All HCD and UVPD mass spectra were collected using an isolation width of 5 m/z, a resolution of 200 000 (@m/z 200), a trapping gas value of 1 and IST of -50 V. UVPD was performed with a single pulse at 1–3 mJ per pulse. UVPD conditions were optimized to maximize coverage, and 1 pulse at 3 mJ per pulse was found to provide the highest number of fragment ions (Fig. S2 and S3†). All UVPD mass spectra were collected in triplicate for each laser energy condition.

Bottom-up LC-MS/MS of crosslinked samples

Tryptic digests of cross-linked proteins were separated on a Thermo Scientific Dionex UltiMate 3000 nano-LC system equipped with a house-packed C18 trap (3 cm × 100 μm i. d.) and analytical columns (15 cm, 75 μm i. d.) and analyzed using a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer. The resolution was set to 60 000 and 30 000 (@m/z 200) for MS1 and MS2 spectra, respectively. MS/MS data collection was performed at top-speed mode with a 3 s cycle time, 1×10^5 intensity threshold, 50 ms maximum injection time, fixed mode HCD, 25% collision energy, 5×10^4 AGC target, and 2 μscans per scan. Dynamic exclusion was utilized for ions of the same m/z observed 3 times in the MS1 spectra within a rolling 20 s elution window, with an exclusion duration of 20 s and a mass tolerance of 25 ppm. Cross-linked peptides were separated using a gradient of 2% B to 35% B to 90% B over the course of 60 minutes. Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. A flow rate of 300 nL min^{-1} was used throughout the 60 minute separation.

Data analysis

Deconvolution of all high-resolution MS/MS spectra was performed using the Xtract algorithm (Thermo Fisher) at a S/N of 3, while deconvolution of low-resolution MS1 spectra was performed using UniDec.[69] ProSight Lite was used for all sequence

coverage analysis using an error tolerance of 10 ppm. Mass shifts of +79.94, +61.91 and +3589.62 Da, +6024.00 Da (with 10 ppm tolerance) and combinations of these mass shifts were used for the C-terminal covalent S_2O modification, Zn^{2+} cofactor (e.g., addition of one Zn atom and loss of two hydrogen atoms), dT12 (Fig. S4a†), and dT20 (Fig. S4b†), respectively. To determine the backbone cleavage yields generated by UVPD, the abundances of the holo ions and their corresponding apo ion series were collectively summed from triplicate runs acquired using three different UVPD conditions (one laser pulse applied at 1 mJ, 2 mJ or 3 mJ). To allow direct comparison across all spectra, the identified holo/apo ions were normalized to the total ion current of the spectrum, as previously described.[60,70,71] All structural representations of gp32 are based on the X-ray crystal structure produced by Shamoo et al., which contains residues 22–239 of the entire 301 amino acid sequence (PDB 1GPC).[7] Backbone cleavages derived from the identified holo fragment ions were plotted onto the gp32 crystal structure (PDB 1GPC) using UV-POSIT[72] and a series of Python and MATLAB (MathWorks) scripts. For this analysis, backbone cleavages derived from holo ions (containing both Zn^{2+} and DNA) were included only if seen at least twice within the three sets of UVPD data (obtained from triplicate runs using 1 pulse, 3 mJ per pulse).

Bottom-up XL-MS data were searched against appropriate protein databases using Byonic software (Protein Metrics, San Carlos, CA) with a tolerance of 10 ppm for precursor and fragment ions, a maximum of two missed cleavages, and a 1% FDR cutoff. Strict qualifications were utilized during the K–K and D/E–K crosslink filtering process. In brief, peptides were grouped by unique peptides, filtered with a score cutoff of 300 or higher, and were further substantiated with a Pep2D of 9.9×10^{-5} or lower. The cross-linked peptide hits of the gp32 homodimer were verified using ClusPro 2.0 protein–protein docking software (Boston, MA) in conjunction with PyMOL Molecular Graphics System, Version 2.0 (Schrodinger, LLC). ClusPro-generated gp32 homodimer structures that displayed inter-protein K–K crosslinks of 30 Å or lower and D/E–K crosslinks of 15 Å or lower were selected.

Results and discussion

To investigate the stoichiometry, structure and mechanism of this viral protein–ssDNA system, native MS was used to analyze solutions containing gp32 or gp32·ssDNA complexes. These species were characterized using MS/MS, including both HCD and UVPD. Crosslinking experiments were undertaken to explore the gp32 homodimer interface, including: (1) direct infusion of the intact complexes formed via crosslinking of the gp32 homodimer and subsequent incubation with ssDNA, and (2) crosslinking of gp32 followed by a traditional bottom-up workflow with tryptic digestion and LC-MS/MS analysis for identification of cross-linked peptides.

We observed gp32 monomers and homodimer ions via native ESI-MS (Fig. 1a), in agreement with a prior FRET study.[10] Each gp32 monomer retained one zinc atom, resulting 1 : 1 or 2 : 2 gp32·Zn complexes (see Fig. S5† for deconvoluted

spectrum). Zn(II) chelation is directly correlated to gp32·ssDNA binding, indicating that gp32 remains folded during MS analysis.[73] MS/MS characterization of the gp32·Zn complex (11+) resulted in 55% sequence coverage by UVPD (1 pulse, 3 mJ) (Fig. 1b and S6†) and 5% coverage by HCD (Fig. S7†) based on production and consideration of both apo sequence ions (no Zn) and holo sequence ions (Zn retained). In agreement with previous studies,[74–76] we observed little change in UVPD sequence coverage based on precursor charge state (Fig. S8a and b†). C-terminal modification of gp32 was also identified based on the UVPD mass spectrum and verified via subsequent bottom-up LC-MS/MS analysis (Fig. S9†) and additional top-down MS/MS analysis of denatured gp32 (Fig. S10†). This C-terminal modification is consistent with incorporation of a disulfur monoxide moiety at Cys302 and is attributed to the intein reaction which occurs during protein purification.

We detected a disulfide bond between Cys87 and Cys90 based on the sequence maps derived from the UVPD and HCD mass spectra of denatured gp32 (Fig. S10a–c†). Upon addition of the reducing agent TCEP (Fig. S10d†), this disulfide bond is disrupted, resulting in the observation of two new backbone cleavages occurring between the two previously linked cysteines (Fig. S10e and f†). Additionally, markedly increased sequence coverage was obtained for the disulfide-reduced proteins upon HCD and UVPD (32% and 41% sequence coverage, respectively) (Fig. S10e and f†) compared to the non-reduced proteins (HCD 21% sequence coverage, UVPD 33% sequence coverage) (Fig. S10b and c†), consistent with previous reports indicating that the presence of disulfide bonds impedes fragmentation of intact proteins.[77] The ion signal originally dispersed among oxidized and reduced proteoforms may be concentrated into the reduced form after addition of the reducing agent, thus increasing the abundance of the precursor ion available for MS/MS analysis. This "concentration" of the precursor into a more homogeneous form likely contributes to the increased sequence coverage observed for the reduced protein compared to the non-reduced protein. These results, along with the notable absence of a disulfide bond in the native protein (Fig. S6b†), suggest that the Cys87–90 disulfide bond forms upon denaturation of gp32 and loss of the Zn^{2+} atom, which is consistent with previous studies which highlight the role of Zn in the coordination of residues Cys87 and Cys90 in the native gp32 protein and stabilization of subdomain I.[73]

We next turned to examination of gp32 in complex with ssDNA oligonucleotides. Addition of ssDNA oligonucleotide, dT12, to the gp32 solution resulted in production of 1 : 1 gp32·dT12 complexes, as observed in several charge states (all containing one Zn as seen in Fig. 1c – see Fig. S11a† for deconvoluted mass spectrum). Nearly all of the 1 : 1 and 2 : 2 gp32·Zn complexes originally seen in Fig. 1a shifted to the ssDNA-bound species. Increasing the oligonucleotide length to dT20 resulted in detection of 2 : 1 and 3 : 1 gp32·dT20 complexes (with two and three Zn, respectively) (Fig. 1d and deconvoluted spectrum in Fig. S11b†). UVPD mass spectra of the 2 : 1 and 3 : 1 gp32·dT20 complexes are displayed in Fig. S12,† both of which show the production of numerous large-size fragment ions, some of which may correspond to
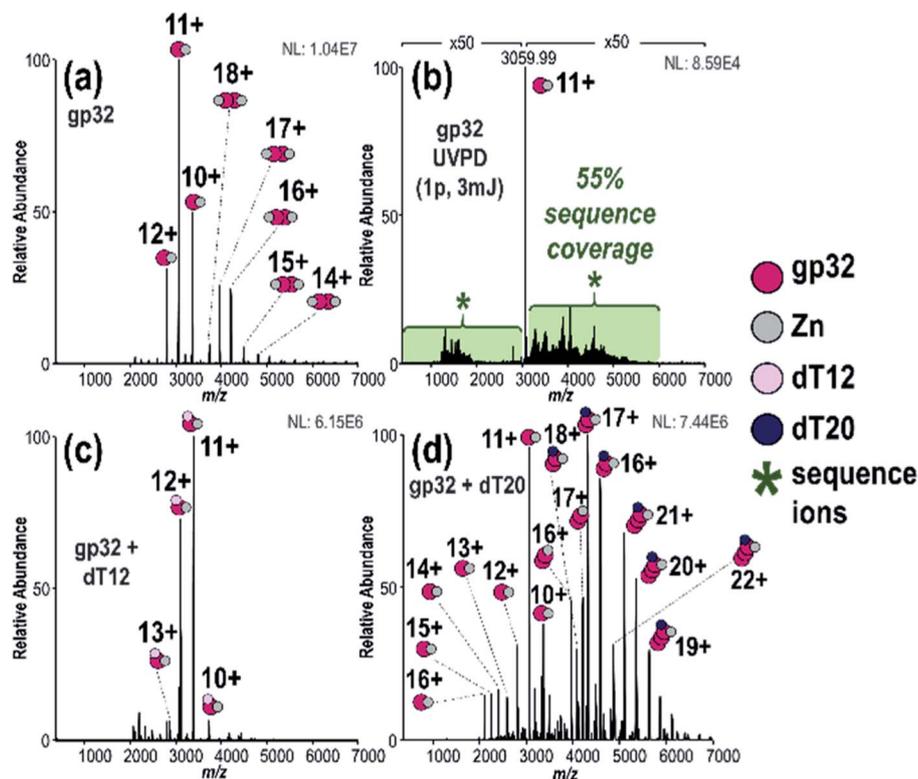
Fig. 1 (a) MS1 spectrum of gp32 showing both monomer and dimers, (b) 193 nm UVPD spectrum of gp32 monomer (11+) (expanded UVPD spectrum is provided in Fig. S6a†), (c) MS1 spectrum of solution containing gp32 + dT12 showing 1 : 1 gp32·dT12 complexes, (d) MS1 spectrum of solution containing gp32 + dT20 showing 2 : 1 and 3 : 1 gp32·dT20 complexes.

portions of two gp32 proteins held together by non-covalent interactions in addition to ones with and without dT20 and Zn. Given the high probability of false positives when searching for ions that include sub-portions of two or more molecules of the same protein, these types of ions were excluded from the searches. Only fragment ions corresponding to sequence ions from apo gp32, gp32 + Zn, gp32 + dT20, or gp32 + Zn + dT20 were considered, as shown in Fig. S12c and f.† These results indicate the ability of native MS to monitor the oligomerization of gp32 via binding to ssDNA as the length of the oligonucleotide increases, and are consistent with an approximately 6,7-nucleotide footprint, as reported in previous gp32 studies.[10,78,79] These results likewise establish that native MS can distinguish the stoichiometries of native protein–DNA complexes and reveal the oligomerization of gp32 and its binding to ssDNA as a function of the DNA strand length.

Next, we used UVPD to interrogate the structures of the gp32·ssDNA complexes. UVPD generates both apo (without DNA) and holo (with DNA) fragment ions that enable localization of the gp32·dT12 interactions with individual amino acid precision.[51,52] To minimize mis-assignment of ions, we focused on the holo fragment ions that retained the entire mass of the DNA sequence. Restricting the holo fragment ion searches to only those that retain the entire ssDNA streamlines the searches and fixes the mass shift of the holo fragment ions to a defined value (in this case +3589.62 Da for dT12), much in the same way that searches and localization of specific post-translational

modifications are successfully executed in other MS/MS analyses of proteins.[80,81] Prior UVPD studies of other protein–ligand complexes have either not observed fragmentation of the ligand, as the amide backbone of the protein is a significant UV chromophore, or have not undertaken searches for fragment ions containing sub-portions of the ligands owing to the enormous search space and potential for false positives.[52,60,70,71,82,83] In essence, considering only holo fragment ions which contain the entire mass of the ssDNA ligand may exclude the identification of some meaningful ions but importantly avoids increasing the false positive rate for identification of mis-assigned spurious holo fragment ions.

Each gp32·ssDNA complex can generate three types of holo fragment ions: ones containing only Zn (e.g., loss of DNA and retention of Zn), ones containing only DNA (e.g., loss of Zn and retention of DNA), and ones containing both Zn and DNA. Zn-free product ions may arise from backbone cleavages in stretches of the protein remote from the Zn binding site, or may arise from ejection of Zn from Zn-coordinated fragment ions during UVPD. The summed distributions of apo ions and these three types of holo ions were categorized based on whether they contain the N-terminus (all $a/a + 1$, b, c ions) or C-terminus (all $x/x + 1$, $y/y − 1$, z ions) and directly compared to the distribution of apo and holo (Zn) ions produced from the DNA-free gp32 precursor (Fig. 2a). While the fragment ion distributions from gp32 (11+) display a nearly equal distribution of apo N- and C-terminal fragment ions, gp32·ssDNA (11+) shows a shift

towards preferential production of apo N-terminal ions. With respect to the production of holo fragment ions (*i.e.*, Zn-containing product ions from the gp32 precursor or fragment ions containing Zn, DNA, or DNA + Zn from the gp32·ssDNA complex), the number of holo ions containing the C-terminus is significantly greater than the number of holo ions containing the N-terminus upon UVPD of the gp32·ssDNA complex, whereas gp32 produces more holo fragment ions containing the N-terminus compared to those containing the C-terminus. These results suggest that Zn in gp32 is coordinated closer to the N-terminus, whereas ssDNA is coordinated closer to the C-terminus in the gp32·dT12 complex.

All of the holo ions produced from the 1 : 1 gp32·dT12 complexes were further grouped by holo ion type (*i.e.*, containing DNA, DNA + Zn, or Zn) (Fig. 2b) as well as by specific ion type ($a/a + 1$, b, c, $x/x + 1$, $y/y − 1$, z) based on the UVPD spectra shown in Fig. S13.† In general, the various ions types ($a/a + 1$, b, c, $x/x + 1$, $y/y − 1$, z) are observed both with and without retention of Zn, as well as with and without retention of dT12. It is this large array of assignable fragment ions that facilitates localization of dT12 as the gp32·dT12 complexes disassemble

and release fragments that retain dT12. These patterns of different holo ions again highlight that dT12 is coordinated closer to the C-terminus, as there are a significantly greater number of C-terminal holo ions than N-terminal holo ions. This contrasts the apo ion data shown in Fig. 2b for UVPD of the 1 : 1 gp32·dT12 complexes in which N-terminal apo ions are nearly twice as abundant as C-terminal apo ions. Fig. 2c shows an even more detailed breakdown of the various types of fragment ions produced from the gp32·ssDNA complexes – *a*-type ions are most abundant for the apo fragment ions, whereas *x*- and *y*-type ions are most abundant for the various holo ion types. The observed shift towards a higher number of C-terminal ions for the holo fragment ions *vs.* apo fragment ions produced from gp32·ssDNA (Fig. 2b) as well as the shifts in distribution of holo and apo fragment ions upon UVPD of gp32 *versus* gp32·ssDNA (Fig. 2a) underscores how the presence of dT12 changes the fragmentation pattern of this protein.

In order to localize the ssDNA ligand binding site, the DNA-containing holo ions (ones containing a portion of the gp32 sequence plus the entire dT12 sequence, with or without retention of Zn) were collectively summed based on the UVPD
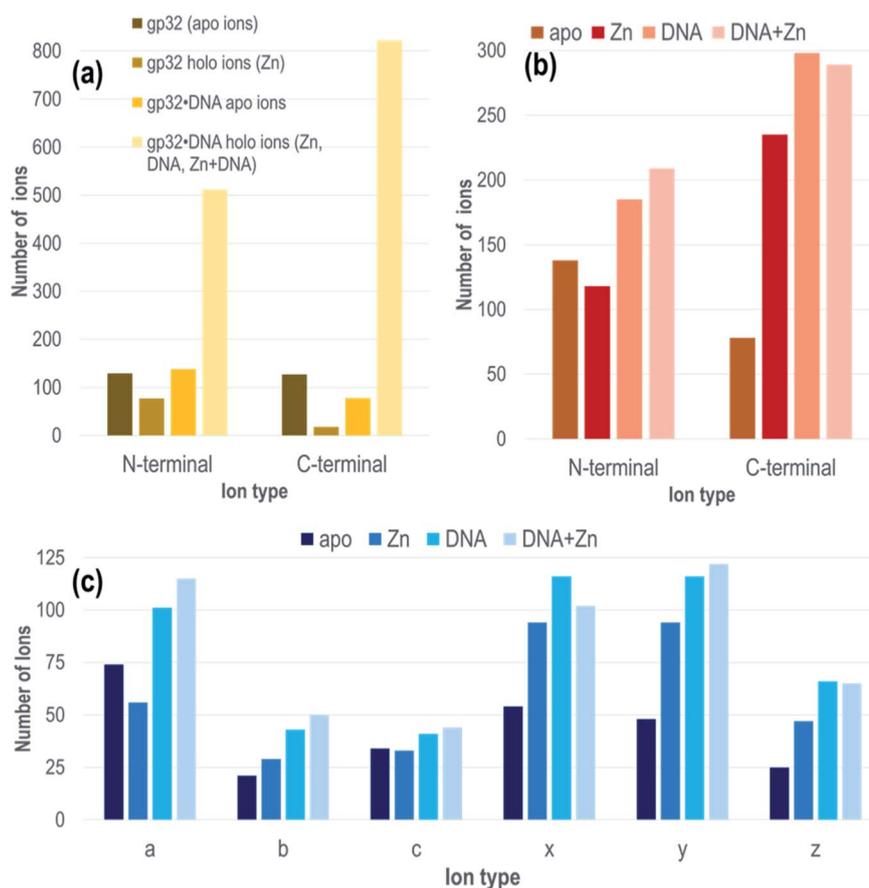

Fig. 2 Distribution of N–terminal (*a,b,c*) and C–terminal (*x,y,z*) fragment ions (a) fragment ions without Zn or with Zn from gp32 (*m/z* 3061, 11+ charge state, containing 1 Zn) or fragment ions containing neither Zn nor DNA or containing Zn, DNA, or Zn + DNA from 1 : 1 gp32·dT12 complex (*m/z* 3386, 11+ charge state, containing 1 Zn), (b) based on ion type containing neither Zn nor DNA, or containing Zn or DNA or both DNA + Zn from 1 : 1 gp32·dT12 complex (*m/z* 3386, 11+ charge state, containing 1 Zn), (c) based on specific ion type (*a/a + 1, b, c x/x + 1, y/y − 1, z*) for 1 : 1 gp32·dT12 complex (*m/z* 3386, 11+ charge state, containing 1 Zn) for fragment ions containing neither Zn nor DNA (apo) or those containing Zn, DNA, or DNA + Zn.

mass spectra acquired for the 1 : 1 gp32·dT12 complex (11+). This holo ion mapping method with UVPD has been previously used to determine binding sites for protein–small molecule ligand complexes but is here used for the first time to interrogate noncovalent protein–DNA interactions.[52,84] The sequence map that illustrates the backbone cleavage sites that lead to the resulting dT12-containing fragment ions is shown in Fig. 3a, demarcated based on whether the fragment ions contain the C-terminus (red) or N-terminus (blue), or in some cases complementary N-terminus and C-terminus ions (green). The corresponding sequence coverage map showing all the backbone cleavage sites that generate all the identified DNA-containing holo ions is displayed in Fig. S14.† As observed from the map, most of the C-terminal holo ions contain over 200 residues. Most N-terminal holo ions, although few in total, are equally large, containing over 200 amino acids. It is reasonable that these long stretches of the protein would retain the DNA. However, there are also a number of fragment ions whose compositions are consistent with just a few residues of gp32 bound to the entire dT12, such as $(y_6 \cdot dT12)^{2+}$ (a short C-terminal holo ion) and $(a_5 \cdot dT12)^{3+}$ (a short N-terminal holo ion) (see confirmations of these ions *via* isotopic fits in Fig. S15.†).

The backbone cleavage sites from which the C-terminal and N-terminal holo ions originate were plotted onto the gp32 crystal structure (PDB ID: 1GPC)[7] (Fig. 3b, and sequence motif map in Fig. S16†) and color-coded to match the cleavage sites in the companion sequence map. Red, blue and green correspond to the backbone positions marking the C-terminal, N-terminal, and bidirectional DNA-containing holo ions, respectively. An almost continuous series of C-terminal holo ions is observed near the protein N-terminus in addition to a number of N-terminal holo ions at the protein C-terminus, indicating a large breadth of ssDNA contact throughout the entire central structural region of the protein (Fig. 3). These results indicate that gp32 binds ssDNA with the majority of residues of sub-domain I, the connecting region, and subdomain II.[7] Furthermore, the array of small bidirectional holo ions observed at both the N- and C-terminus of the protein suggests that both the N-terminus and C-terminus of gp32 interact with dT12. Overall, this data sheds new light on the extent and reach of ssDNA binding with gp32, expanding the binding site from the electropositive cleft, and implicates interactions of both the N- and C-terminal tails with dT12, adding new insight to the partially-solved X-ray structure of gp32 (1GPC).[7] These results highlight the ability of MS-based structural methods to interrogate disordered regions of proteins which may be inaccessible *via* traditional techniques.

Next, we investigated the hypothesis that gp32 homodimers dissociate into monomers prior to binding to ssDNA.[9,12] First gp32 was incubated with Lys–Lys crosslinker BS3, and the resulting MS1 spectrum shown in Fig. 4a substantiates that the intact crosslinked products are predominantly dimers with an estimated average of 8 cross-linked and/or dead-end modified lysines per molecule of gp32 (MW_{exp} of intact cross-linked gp32 dimer = 69 755 Da, see deconvoluted spectrum in Fig. S17†). The similar mass shifts of each dead-end modification, which add a mass shift of +158 Da, and each crosslink (+138 Da), prevents specific differentiation of the two modifications given the large mass of the dimer and the resolution of the mass
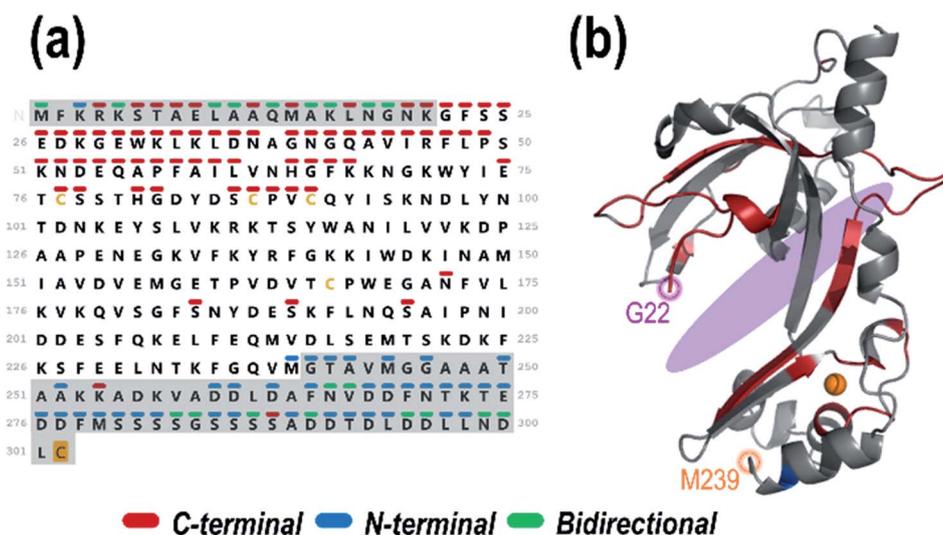

Fig. 3 (a) Sequence of gp32 with the backbone cleavage sites leading to N-terminal ((*a*, *a* + 1, *b*, *c*) blue), C-terminal ((*x*, *x* + 1, *y*, *y* − 1, *z*) red) or bidirectional (green) dT12-containing holo fragment ions generated upon UVPD (combined data from 1 pulse at 3 mJ) of the 1 : 1 gp32·dT12 complex (*m/z* 3386, 11+ charge state, containing 1 Zn). Greyed out sequence areas correspond to regions of the sequence that are unresolved in the crystal structure but contain many confirmatory backbone cleavages that lead to assignable sequence ions. The C-terminal cysteine is shaded in gold to denote the disulfur monoxide modification. (b) Schematic model of the gp32 crystal structure (PDB ID: 1GPC, residues 22−239) with the residues corresponding to backbone cleavage sites that lead to holo fragment ions shaded in the respective colors (red for backbone cleavages that lead to C-terminal sequence ions, blue for backbone cleavages that lead to N-terminal sequence ions, and green for backbone cleavages that result in both C-terminal and N-terminal ions). The approximate ssDNA binding cleft is shaded in purple.
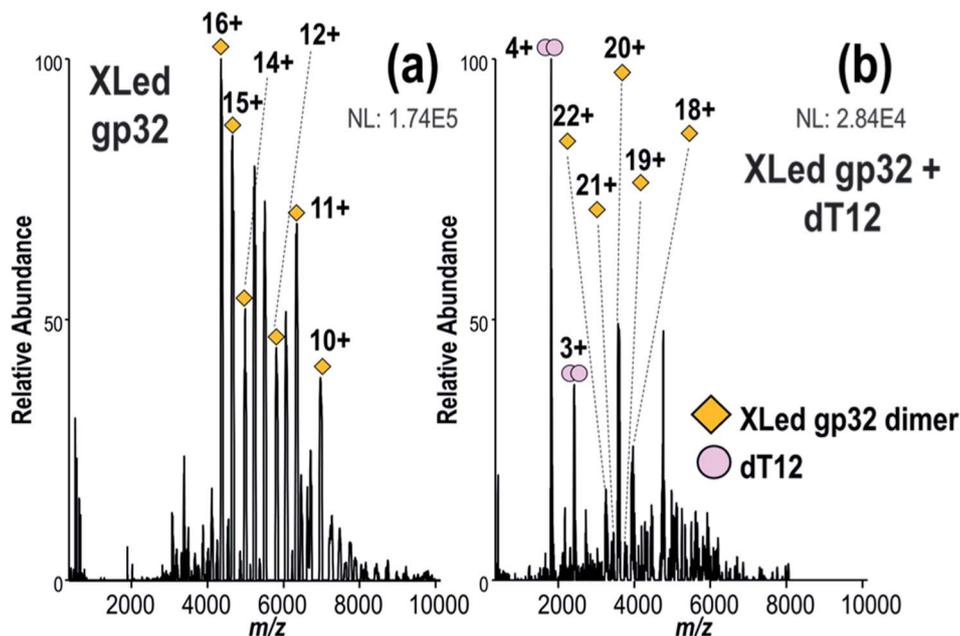
Fig. 4 Native MS1 spectra of (a) the solution containing gp32 after BS3 cross-linking and (b) the solution containing gp32 after BS3 cross-linking followed by addition of dT12, showing the absence of gp32·dT12 complexes. Clean-up of the solutions using 30 kDa SEC filters after crosslinking removed most of the monomeric gp32 and unreacted crosslinker.

spectrometer. The deconvoluted MS1 spectrum (Fig. S17†) indicates that the crosslinked dimer contains a maximum of 17 crosslinks (with no dead-ends) or a maximum of 15 dead-ends (no crosslinks), or various combinations of crosslinks and dead-ends between these two extremes. Because of the different possible mass additions related to the crosslinking reactions (*i.e.*, dead-end modification, protein cross-link), it is not possible to determine whether the cross-linked gp32 dimer retains Zn. Given the four crosslinks identified later in Fig. 5, the cross-linked sample likely contains on average 4 crosslinks and 12 dead-end modifications per dimer. The MS1 spectrum obtained after addition of ssDNA (dT12) to the BS3-crosslinked gp32 solution is displayed in Fig. 4b. No gp32·dT12 complexes are observed, instead several dT12 dimers and crosslinked gp32 dimers are evident (Fig. 4b, S18a and b†). The inability of the cross-linked gp32 homodimer to form complexes with dT12 suggests that the binding cleft of gp32 is blocked or disrupted upon crosslinking, thus disabling interactions with dT12 in a way that prevents the formation of the gp32·dT12 complexes. Additionally, the charge state distribution of the cross-linked gp32 dimer shifts towards the lower *m/z* range (higher charge states) upon the addition of dT12 to the solution (Fig. 4b). We hypothesize that this shift in the charge state distribution may be due to the relatively high concentration of free dT12 in solution compared to the heterogeneous population of cross-linked gp23 dimers. Cross-linked gp32 dimer was also analyzed in the presence of dT20 (Fig. S19†) with similar results, in that no binding was observed between dT20 and the cross-linked gp32 homodimer. In this case, the addition of dT20 did not result in a notable shift in the charge state distribution of the cross-linked gp32 dimer, and free dT20 is not observable

in the mass spectrum. The corresponding MS1 spectrum acquired for a denaturing solution containing crosslinked gp32 and dT12 is shown in Fig. S20,† showing both denatured gp32 monomer and dimer. The charge state distributions of denatured, cross-linked gp32 dimers in Fig. S20a† (∼14+ to 20+) is somewhat broadened relative to the charge state distribution of the native-like non-cross-linked gp32 dimers (Fig. 1a, ∼14+ to 18+), and the charge state distribution of denatured cross-linked gp32 monomers (∼11+ to 14+ in Fig. S20a†) are extremely shifted relative to the charge states of denatured gp32 monomers (Fig. S21,† ∼18+ to 31+). The intramolecularly crosslinked gp32 monomers may better maintain a more
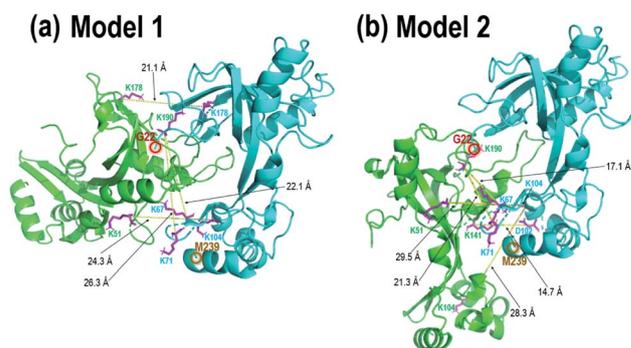

Fig. 5 gp32 homodimer models from bottom-up crosslinking data and protein–protein docking results with BS3 (K–K) and DMTMM (D/E–K) crosslinks delineated as magenta residues with yellow lines, with (a) four BS3 crosslinks and (b) four BS3 crosslinks and one DMTMM crosslink. Comparison of these two homodimer models with the gp32-ssDNA structure shown in Fig. 3b shows clear overlap of the two homodimer interfaces with the monomeric gp32 ssDNA binding cleft.

compact structure, preventing unfolding and charging, in addition to the conversion of the basic lysine side-chains to various less basic hydrolyzed deadend groups.

In order to localize the crosslinks in the gp32 dimer, a bottom-up LC-MS/MS strategy was used. Gp32 was incubated with either Lys–Lys crosslinker BS3 or Asp/Glu–Lys crosslinker DMTMM. The resulting solutions containing crosslinked gp32 were then subjected to tryptic digestion and MS/MS analysis to map the locations of the crosslinks. The base peak LC-MS chromatograms and the MS/MS spectra of the crosslinked peptides are shown in Fig. S22 and S23.† Four BS3-crosslinked peptides were identified, with crosslinks occurring between Lys residues 51 and 104, 104 and 104, 178 and 178, 190 and 67, and 190 and 71, while one DMTMM-crosslinked peptide was identified with the linkage between D102 and K141 (Fig. S23a†). The crosslinked peptides were fit to potential gp32 dimer structures generated by protein–protein docking. Fig. 5 shows the two optimized gp32 homodimer structures based on validation of the five cross-linked peptide assignments. Model 1 incorporates 4 out of the 5 identified BS3 crosslinks, whereas Model 2 incorporates 4 out of the 5 identified BS3 crosslinks plus the DMTMM crosslink (D102-K141). All potential intramolecular crosslinked peptide hits were discarded based on comparing Lys–Lys and Asp/Glu–Lys distances in the model monomer *vs.* dimer structures with BS3 or DMTMM crosslinking distance constraints. Comparison of the two homodimer models in Fig. 5 to the gp32 crystal structure in Fig. 3b reveals overlap of the homodimer interfaces in each of the two models with the gp32 binding cleft, suggesting that a large portion of the gp32 ssDNA-binding cleft is involved in the gp32 homodimer interface. This result implies that the ssDNA binding cleft is not accessible for ssDNA binding when gp32 is homodimeric. Our MS data supports a dimerization model that gp32 dimers are unable to bind ssDNA and that dimer dissociation is required prior to binding. We additionally purified a gp32 construct with a C-terminal tail deletion (gp32-ΔCTD, amino acids 1–254) to test whether the C-terminal acidic tail impacts dimerization, which is unresolved in the gp32 crystal structure. We still observed dimerization in the gp32-ΔCTD constructs *via* native MS1 (Fig. S24a†), but did not observe evidence of ssDNA binding (Fig. S24b†). Our data suggests that the acidic tail is dispensable for dimer formation, similar to the T7 phage gp2.5,[85] but is essential for ssDNA binding. We favor a model where the gp32 acidic tail mediates multiple protein–protein interactions, including with T4 gp59.[86]

Conclusions

In summary, combining native MS, UVPD-MS and XL-MS expanded our molecular understanding of how T4 gp32 binds ssDNA and how this binding is inhibited for the crosslinked homodimeric gp32 complex. Native UVPD-MS combined with XL-MS provides direct evidence that the gp32 dimer dissociates into monomers prior to binding to ssDNA. The presence of 3 : 1 gp32·dT20 complexes observed in Fig. 1d also indicates a sequential association of gp32 monomers with ssDNA owing to the odd number of gp32 molecules that are bound to the

ssDNA. Additionally, we characterized the ssDNA binding cleft size and developed a model structure for the ssDNA-inactive gp32 dimer. We conjecture that gp32 dimers act as an inactive storage compartment and may also regulate the extent of ssDNA binding. Cellular processes that sequester gp32 into dimers also prevent ssDNA binding by this protein. Based on the successful characterization of the well-studied T4 gp32–ssDNA complexes, we anticipate that this multi-pronged MS approach will provide further in-depth characterization of other novel protein–DNA systems which may have unsolved structures and unknown binding modes.

Data availability

Extensive data is provided as ESI.†

Author contributions

MSB, IJF, and JSB contributed to conceptualization of this project. JMS and IJF provided T4 gp32 protein constructs (resources), while JSB provided all other resources. JMS performed SDS-PAGE experiments (data curation), while MSB and JNW performed all mass spectrometry experiments (data curation), formal analysis, investigation and methodology. JSB coordinated project administration. JSB and IJF are responsible for funding acquisition for this project. MSB and JSB prepared figures (visualization) and the manuscript (writing – original draft). All authors contributed to the final manuscript (writing – review & editing).

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

References

1 B. Dey, S. Thukral, S. Krishnan, M. Chakrobarty, S. Gupta, C. Manghani and V. Rani, DNA–Protein Interactions: Methods for Detection and Analysis, *Mol. Cell. Biochem.*, 2012, **365**, 279–299.

2 S. M. Gustafsdottir, J. Schlingemann, A. Rada-Iglesias, E. Schallmeiner, M. Kamali-Moghaddam, C. Wadelius and U. Landegren, *In Vitro* Analysis of DNA–Protein Interactions by Proximity Ligation, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 3067–3072.

3 D. P. Giedroc, K. M. Keating, K. R. Williams, W. H. Konigsberg and J. E. Coleman, Gene 32 Protein, the Single-Stranded DNA Binding Protein from Bacteriophage

T4, Is a Zinc Metalloprotein, *Proc. Natl. Acad. Sci. U.S.A.*, 1986, **83**, 8452–8456.

4 R. D. Shereda, A. G. Kozlov, T. M. Lohman, M. M. Cox and J. L. Keck, SSB as an Organizer/Mobilizer of Genome Maintenance Complexes, *Crit. Rev. Biochem. Mol. Biol.*, 2008, **43**, 289–318.

5 A. H. Marceau, Functions of Single-Strand DNA-Binding Proteins in DNA Replication, Recombination, and Repair, *Methods Mol. Biol.*, 2012, **922**, 1–21.

6 M. Z. Islam, A. Fokine, M. Mahalingam, Z. Zhang, C. Garcia-Doval, M. J. van Raaij, M. G. Rossmann and V. B. Rao, Molecular Anatomy of the Receptor Binding Module of a Bacteriophage Long Tail Fiber, *PLoS Pathog.*, 2019, **15**(12), e1008193.

7 Y. Shamoo, A. M. Friedman, M. R. Parsons, W. H. Konigsberg and T. A. Steitz, Crystal Structure of a Replication Fork Single-Stranded DNA Binding Protein (T4 Gp32) Complexed to DNA, *Nature*, 1995, **376**, 362–366.

8 A. K. Szczepaska, E. Bidnenko, D. Płochocka, S. McGovern, S. D. Ehrlich, J. Bardowski, P. Polard and M.-C. Chopin, A Distinct Single-Stranded DNA-Binding Protein Encoded by the Lactococcus Lactis Bacteriophage BIL67, *Virology*, 2007, **363**, 104–112.

9 D. Jose, S. E. Weitzel, W. A. Baase and P. H. von Hippel, Mapping the Interactions of the Single-Stranded DNA Binding Protein of Bacteriophage T4 (Gp32) with DNA Lattices at Single Nucleotide Resolution: Gp32 Monomer Binding, *Nucleic Acids Res.*, 2015, **43**, 9276–9290.

10 C. Phelps, B. Israels, D. Jose, M. C. Marsh, P. H. von Hippel and A. H. Marcus, Using Microsecond Single-Molecule FRET to Determine the Assembly Pathways of T4 SsDNA Binding Protein onto Model DNA Replication Forks, *Proc. Natl. Acad. Sci. U.S.A.*, 2017, **114**, E3612–E3621.

11 K. Pant, B. Anderson, H. Perdana, M. A. Malinowski, A. T. Win, C. Pabst, M. C. Williams and R. L. Karpel, The Role of the C-Domain of Bacteriophage T4 Gene 32 Protein in SsDNA Binding and DsDNA Helix-Destabilization: Kinetic, Single-Molecule, and Cross-Linking Studies, *PloS One*, 2018, **13**, e0194357.

12 P. H. von Hippel and A. H. Marcus, The Many Roles of Binding Cooperativity in the Control of DNA Replication, *Biophys. J.*, 2019, **117**, 2043–2046.

13 J. R. Casas-Finet, K. R. Fischer and R. L. Karpel, Structural Basis for the Nucleic Acid Binding Cooperativity of Bacteriophage T4 Gene 32 Protein: The (Lys/Arg)3(Ser/Thr)2 (LAST) Motif, *Proc. Natl. Acad. Sci. U.S.A.*, 1992, **89**, 1050–1054.

14 T. D. Veenstra, Electrospray Ionization Mass Spectrometry: A Promising New Technique in the Study of Protein/DNA Noncovalent Complexes, *Biochem. Biophys. Res. Commun.*, 1999, **257**, 1–5.

15 E. Duijn, Current Limitations in Native Mass Spectrometry Based Structural Biology, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 971–978.

16 A. C. Leney and A. J. R. Heck, Native Mass Spectrometry: What Is in the Name?, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 5–13.

17 V. Gabelica, C. Vreuls, P. Filée, V. Duval, B. Joris and E. D. Pauw, Advantages and Drawbacks of Nanospray for Studying Noncovalent Protein-DNA Complexes by Mass Spectrometry, *Rapid Commun. Mass Spectrom.*, 2002, **16**, 1723–1728.

18 A. J. R. Heck and R. H. H. van den Heuvel, Investigation of Intact Protein Complexes by Mass Spectrometry, *Mass Spectrom. Rev.*, 2004, **23**, 368–389.

19 H. Hernández and C. V. Robinson, Determining the Stoichiometry and Interactions of Macromolecular Assemblies from Mass Spectrometry, *Nat. Protoc.*, 2007, **2**, 715–726.

20 S. A. Hofstadler and R. H. Griffey, Analysis of Noncovalent Complexes of DNA and RNA by Mass Spectrometry, *Chem. Rev.*, 2001, **101**, 377–390.

21 C. L. Mazzitelli, J. S. Brodbelt, J. T. Kern, M. Rodriguez and S. M. Kerwin, Evaluation of Binding of Perylene Diimide and Benzannulated Perylene Diimide Ligands to Dna by Electrospray Ionization Mass Spectrometry, *J. Am. Soc. Mass Spectrom.*, 2006, **17**, 593–604.

22 C. L. Mazzitelli and J. S. Brodbelt, Probing Ligand Binding to Duplex DNA Using $KMnO_4$ Reactions and Electrospray Ionization Tandem Mass Spectrometry, *Anal. Chem.*, 2007, **79**, 4636–4647.

23 J. J. Wilson and J. S. Brodbelt, Infrared Multiphoton Dissociation of Duplex DNA/Drug Complexes in a Quadrupole Ion Trap, *Anal. Chem.*, 2007, **79**, 2067–2077.

24 S. Y. New, S. T. Lee and X. D. Su, DNA-Templated Silver Nanoclusters: Structural Correlation and Fluorescence Modulation, *Nanoscale*, 2016, **8**, 17729–17746.

25 E. Thyrhaug, S. A. Bogh, M. R. Carro-Temboury, C. S. Madsen, T. Vosch and D. Zigmantas, Ultrafast Coherence Transfer in DNA-Templated Silver Nanoclusters, *Nat. Commun.*, 2017, **8**, 15577.

26 M. S. Blevins, D. Kim, C. M. Crittenden, S. Hong, H.-C. Yeh, J. T. Petty and J. S. Brodbelt, Footprints of Nanoscale DNA–Silver Cluster Chromophores *via* Activated-Electron Photodetachment Mass Spectrometry, *ACS Nano*, 2019, **13**, 14070–14079.

27 Y. Guo, X. Pan, W. Zhang, Z. Hu, K.-W. Wong, Z. He and H.-W. Li, Label-Free Probes Using DNA-Templated Silver Nanoclusters as Versatile Reporters, *Biosens. Bioelectron.*, 2020, **150**, 111926.

28 E.-M. Schneeberger and K. Breuker, Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type Tat Peptide for Binding Site Mapping, *Angew. Chem., Int. Ed.*, 2017, **56**, 1254–1258.

29 S. M. Swasey, F. Rosu, S. M. Copp, V. Gabelica and E. G. Gwinn, Parallel Guanine Duplex and Cytosine Duplex DNA with Uninterrupted Spines of Ag^I-Mediated Base Pairs, *J. Phys. Chem. Lett.*, 2018, **9**, 6605–6610.

30 V. D'Atri and V. Gabelica, DNA and RNA Telomeric G-Quadruplexes: What Topology Features Can Be Inferred from Ion Mobility Mass Spectrometry?, *Analyst*, 2019, **144**, 6074–6088.

31 E. van Duijn, I. M. Barbu, A. Barendregt, M. M. Jore, B. Wiedenheft, M. Lundgren, E. R. Westra, S. J. J. Brouns,

J. A. Doudna, J. van der Oost and A. J. R. Heck, Native Tandem and Ion Mobility Mass Spectrometry Highlight Structural and Modular Similarities in Clustered-Regularly-Interspaced Shot-Palindromic-Repeats (CRISPR)-Associated Protein Complexes From *Escherichia Coli* and *Pseudomonas Aeruginosa*, *Mol. Cell. Proteomics*, 2012, **11**, 1430–1441.

32 D. Shiu-Hin Chan, W.-G. Seetoh, B. N. McConnell, D. Matak-Vinković, S. E. Thomas, V. Mendes, M. Blaszczyk, A. G. Coyne, T. L. Blundell and C. Abell, Structural Insights into the EthR–DNA Interaction Using Native Mass Spectrometry, *Chem. Commun.*, 2017, **53**, 3527–3530.

33 K. Saikusa, D. Kato, A. Nagadoi, H. Kurumizaka and S. Akashi, Native Mass Spectrometry of Protein and DNA Complexes Prepared in Nonvolatile Buffers, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 711–718.

34 J. Zhang, R. R. O. Loo and J. A. Loo, Structural Characterization of a Thrombin-Aptamer Complex by High Resolution Native Top-Down Mass Spectrometry, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1815–1822.

35 J. Satiaputra, L. M. Sternicki, A. J. Hayes, T. L. Pukala, G. W. Booker, K. E. Shearwin and S. W. Polyak, Native Mass Spectrometry Identifies an Alternative DNA-Binding Pathway for BirA from Staphylococcus Aureus, *Sci. Rep.*, 2019, **9**, 2767.

36 N.-T. Nguyen-Huynh, J. Osz, C. Peluso-Iltis, N. Rochel, N. Potier and E. Leize-Wagner, Monitoring of the Retinoic Acid Receptor-Retinoid X Receptor Dimerization upon DNA Binding by Native Mass Spectrometry, *Biophys. Chem.*, 2016, **210**, 2–8.

37 A. Politis, A. Y. Park, Z. Hall, B. T. Ruotolo and C. V. Robinson, Integrative Modelling Coupled with Ion Mobility Mass Spectrometry Reveals Structural Features of the Clamp Loader in Complex with Single-Stranded DNA Binding Protein, *J. Mol. Biol.*, 2013, **425**, 4790–4801.

38 A. Butterer, C. Pernstich, R. M. Smith, F. Sobott, M. D. Szczelkun and J. Tóth, Type III Restriction Endonucleases Are Heterotrimeric: Comprising One Helicase–Nuclease Subunit and a Dimeric Methyltransferase That Binds Only One Specific DNA, *Nucleic Acids Res.*, 2014, **42**, 5139–5150.

39 B. J. Caldwell, A. Norris, E. Zakharova, C. E. Smith, C. T. Wheat, D. Choudhary, M. Sotomayor, V. H. Wysocki and C. E. Bell, Oligomeric Complexes Formed by Redβ Single Strand Annealing Protein in Its Different DNA Bound States, *Nucleic Acids Res.*, 2021, **49**, 3441–3460.

40 R. J. Rose, E. Damoc, E. Denisov, A. Makarov and A. J. R. Heck, High-Sensitivity Orbitrap Mass Analysis of Intact Macromolecular Assemblies, *Nat. Methods*, 2012, **9**, 1084–1086.

41 A. Stützer, L. M. Welp, M. Raabe, T. Sachsenberg, C. Kappert, A. Wulf, A. M. Lau, S.-S. David, A. Chernev, K. Kramer, A. Politis, O. Kohlbacher, W. Fischle and H. Urlaub, Analysis of Protein-DNA Interactions in Chromatin by UV Induced Cross-Linking and Mass Spectrometry, *Nat. Commun.*, 2020, **11**, 5250.

42 P. A. Limbach, P. F. Crain and J. A. McCloskey, Molecular Mass Measurement of Intact Ribonucleic Acids *via* Electrospray Ionization Quadrupole Mass Spectrometry, *J. Am. Soc. Mass Spectrom.*, 1995, **6**, 27–39.

43 Z. Xia, J. B. DeGrandchamp and E. R. Williams, Native Mass Spectrometry beyond Ammonium Acetate: Effects of Nonvolatile Salts on Protein Stability and Structure, *Analyst*, 2019, **144**, 2565–2573.

44 S. Wittig, I. Songailiene and C. Schmidt, Formation and Stoichiometry of CRISPR-Cascade Complexes with Varying Spacer Lengths Revealed by Native Mass Spectrometry, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 538–546.

45 E.-M. Schneeberger, M. Halper, M. Palasser, S. V. Heel, J. Vušurović, R. Plangger, M. Juen, C. Kreutz and K. Breuker, Native Mass Spectrometry Reveals the Initial Binding Events of HIV-1 Rev to RRE Stem II RNA, *Nat. Commun.*, 2020, **11**, 5750.

46 S. R. Harvey, Y. Liu, W. Liu, V. H. Wysocki and A. Laganowsky, Surface Induced Dissociation as a Tool to Study Membrane Protein Complexes, *Chem. Commun.*, 2017, **53**, 3106–3109.

47 H. Zhang, W. Cui, J. Wen, R. E. Blankenship and M. L. Gross, Native Electrospray and Electron-Capture Dissociation in FTICR Mass Spectrometry Provide Top-Down Sequencing of a Protein Component in an Intact Protein Assembly, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 1966–1968.

48 J. S. Brodbelt, Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules, *Chem. Soc. Rev.*, 2014, **43**, 2757–2783.

49 J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation, *J. Am. Chem. Soc.*, 2013, **135**, 12646–12651.

50 L. A. Macias, I. C. Santos and J. S. Brodbelt, Ion Activation Methods for Peptides and Proteins, *Anal. Chem.*, 2020, **92**, 227–251.

51 M. B. Cammarata and J. S. Brodbelt, Structural Characterization of Holo- and Apo-Myoglobin in the Gas Phase by Ultraviolet Photodissociation Mass Spectrometry, *Chem. Sci.*, 2015, **6**, 1324–1333.

52 M. R. Mehaffey, M. B. Cammarata and J. S. Brodbelt, Tracking the Catalytic Cycle of Adenylate Kinase by Ultraviolet Photodissociation Mass Spectrometry, *Anal. Chem.*, 2018, **90**, 839–846.

53 A. Sinz, Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–Protein Interactions, *Mass Spectrom. Rev.*, 2006, **25**, 663–682.

54 J. D. Chavez and J. E. Bruce, Chemical Cross-Linking with Mass Spectrometry: A Tool for Systems Structural Biology, *Curr. Opin. Chem. Biol.*, 2019, **48**, 8–18.

55 M. M. Zhang, B. R. Beno, R. Y.-C. Huang, J. Adhikari, E. G. Deyanova, J. Li, G. Chen and M. L. Gross, An Integrated Approach for Determining a Protein–Protein Binding Interface in Solution and an Evaluation of Hydrogen–Deuterium Exchange Kinetics for Adjudicating Candidate Docking Models, *Anal. Chem.*, 2019, **91**, 15709–15717.

56 J. Mintseris and S. P. Gygi, High-Density Chemical Cross-Linking for Modeling Protein Interactions, *Proc. Natl. Acad. Sci. U.S.A.*, 2020, **117**, 93–102.

57 H. Matsumoto, H. Haniu and N. Komori, Determination of Protein Molecular Weights on SDS-PAGE, *Methods Mol. Biol.*, 2019, **1855**, 101–105.

58 M. B. Cammarata, L. A. Macias, J. Rosenberg, A. Bolufer and J. S. Brodbelt, Expanding the Scope of Cross-Link Identifications by Incorporating Collisional Activated Dissociation and Ultraviolet Photodissociation Methods, *Anal. Chem.*, 2018, **90**, 6385–6389.

59 M. B. Cammarata and J. S. Brodbelt, Characterization of Intra- and Intermolecular Protein Crosslinking by Top Down Ultraviolet Photodissociation Mass Spectrometry, *ChemistrySelect*, 2016, **1**, 590–593.

60 M. R. Mehaffey, J. D. Sanders, D. D. Holden, C. L. Nilsson and J. S. Brodbelt, Multistage Ultraviolet Photodissociation Mass Spectrometry To Characterize Single Amino Acid Variants of Human Mitochondrial BCAT2, *Anal. Chem.*, 2018, **90**, 9904–9911.

61 J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation, *J. Am. Chem. Soc.*, 2013, **135**, 12646–12651.

62 W. J. H. Van Berkel, R. H. H. Van Den Heuvel, C. Versluis and A. J. R. Heck, Detection of Intact MegaDalton Protein Assemblies of Vanillyl-Alcohol Oxidase by Mass Spectrometry, *Protein Sci.*, 2008, **9**, 435–439.

63 J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson and A. J. R. Heck, Studying 18 MDa Virus Assemblies with Native Mass Spectrometry, *Angew. Chem., Int. Ed.*, 2013, **52**, 4020–4023.

64 Y. M. Abbas, D. Wu, S. A. Bueler, C. V. Robinson and J. L. Rubinstein, Structure of V-ATPase from the Mammalian Brain, *Science*, 2020, **367**, 1240–1246.

65 M. E. Belov, E. Damoc, E. Denisov, P. D. Compton, S. Horning, A. A. Makarov and N. L. Kelleher, From Protein Complexes to Subunit Backbone Fragments: A Multi-Stage Approach to Native Mass Spectrometry, *Anal. Chem.*, 2013, **85**, 11163–11173.

66 K. L. Fort, M. van de Waterbeemd, D. Boll, M. Reinhardt-Szyba, M. E. Belov, E. Sasaki, R. Zschoche, D. Hilvert, A. A. Makarov and A. J. R. Heck, Expanding the Structural Analysis Capabilities on an Orbitrap-Based Mass Spectrometer for Large Macromolecular Complexes, *Analyst*, 2018, **143**, 100–105.

67 G. Ben-Nissan, M. E. Belov, D. Morgenstern, Y. Levin, O. Dym, G. Arkind, C. Lipson, A. A. Makarov and M. Sharon, Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex, *Anal. Chem.*, 2017, **89**, 4708–4715.

68 O. S. Skinner, P. C. Havugimana, N. A. Haverland, L. Fornelli, B. P. Early, J. B. Greer, R. T. Fellers, K. R. Durbin, L. H. F. Do Vale, R. D. Melani, H. S. Seckler, M. T. Nelp, M. E. Belov, S. R. Horning, A. A. Makarov, R. D. LeDuc, V. Bandarian, P. D. Compton and N. L. Kelleher, An Informatic Framework for Decoding Protein Complexes by Top-down Mass Spectrometry, *Nat. Methods*, 2016, **13**, 237–240.

69 M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles, *Anal. Chem.*, 2015, **87**, 4370–4376.

70 M. B. Cammarata, R. Thyer, J. Rosenberg, A. Ellington and J. S. Brodbelt, Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry, *J. Am. Chem. Soc.*, 2015, **137**, 9128–9135.

71 M. B. Cammarata, C. L. Schardon, M. R. Mehaffey, J. Rosenberg, J. Singleton, W. Fast and J. S. Brodbelt, Impact of G12 Mutations on the Structure of K-Ras Probed by Ultraviolet Photodissociation Mass Spectrometry, *J. Am. Chem. Soc.*, 2016, **138**, 13187–13196.

72 J. Rosenberg, W. R. Parker, M. B. Cammarata and J. S. Brodbelt, UV-POSIT: Web-Based Tools for Rapid and Facile Structural Interpretation of Ultraviolet Photodissociation (UVPD) Mass Spectra, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 1323–1326.

73 H. Qiu, T. Kodadek and D. P. Giedroc, Zinc-Free and Reduced T4 Gene 32 Protein Binds Single-Stranded DNA Weakly and Fails to Stimulate UvsX-Catalyzed Homologous Pairing, *J. Biol. Chem.*, 1994, **269**, 2773–2781.

74 S. N. Sipe and J. S. Brodbelt, Impact of charge state on 193 nm ultraviolet photodissociation of protein complexes, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9265–9276.

75 S. A. Robotham and J. S. Brodbelt, Comparison of Ultraviolet Photodissociation and Collision Induced Dissociation of Adrenocorticotropic Hormone Peptides, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1570–1579.

76 A. Q. Stiving, S. R. Harvey, B. J. Jones, B. Bellina, J. M. Brown, P. E. Barran and V. H. Wysocki, Coupling 193 Nm Ultraviolet Photodissociation and Ion Mobility for Sequence Characterization of Conformationally-Selected Peptides, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 2313–2320.

77 M. J. P. Rush, N. M. Riley, M. S. Westphall and J. J. Coon, Top-Down Characterization of Proteins with Intact Disulfide Bonds Using Activated-Ion Electron Transfer Dissociation, *Anal. Chem.*, 2018, **90**, 8946–8953.

78 B. R. Camel, D. Jose, K. Meze, A. Dang and P. H. von Hippel, Mapping DNA Conformations and Interactions within the Binding Cleft of Bacteriophage T4 Single-Stranded DNA Binding Protein (Gp32) at Single Nucleotide Resolution, *Nucleic Acids Res.*, 2021, **49**, 916–927.

79 D. E. Jensen, R. C. Kelly and P. H. von Hippel, DNA "Melting" Proteins. II. Effects of Bacteriophage T4 Gene 32-Protein Binding on the Conformation and Stability of Nucleic Acid Structures, *J. Biol. Chem.*, 1976, **251**, 7215–7228.

80 A. Doerr, Top-down Mass Spectrometry, *Nat. Methods*, 2008, **5**, 24.

81 M. Zhou, C. Lantz, K. A. Brown, Y. Ge, L. Paša-Tolić, J. A. Loo and F. Lermyte, Higher-Order Structural Characterisation of

Native Proteins and Complexes by Top-down Mass Spectrometry, *Chem. Sci.*, 2020, **11**, 12918–12936.

82 M. R. Mehaffey, C. L. Schardon, E. T. Novelli, M. B. Cammarata, L. J. Webb, W. Fast and J. S. Brodbelt, Investigation of GTP-Dependent Dimerization of G12X K-Ras Variants Using Ultraviolet Photodissociation Mass Spectrometry, *Chem. Sci.*, 2019, **10**, 8025–8034.

83 M. R. Mehaffey, Y. C. Ahn, D. D. Rivera, P. W. Thomas, Z. Cheng, M. W. Crowder, R. F. Pratt, W. Fast and J. S. Brodbelt, Elusive structural changes of New Delhi metallo-β-lactamase revealed by ultraviolet photodissociation mass spectrometry, *Chem. Sci.*, 2020, **11**, 8999–9010.

84 J. P. O'Brien, W. Li, Y. Zhang and J. S. Brodbelt, Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry, *J. Am. Chem. Soc.*, 2014, **136**, 12920–12928.

85 T. Hollis, J. M. Stattel, D. S. Walther, C. C. Richardson and T. Ellenberger, Structure of the Gene 2.5 Protein, a Single-Stranded DNA Binding Protein Encoded by Bacteriophage T7, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 9557–9562.

86 S. W. Morrical, H. T. Beernink, A. Dash and K. Hempstead, The Gene 59 Protein of Bacteriophage T4. Characterization of Protein-Protein Interactions with Gene 32 Protein, the T4 Single-Stranded DNA Binding Protein, *J. Biol. Chem.*, 1996, **271**, 20198–20207.