

SCIENTIFIC REPORTS



OPEN

Structural characterization of a novel KH-domain containing plant chloroplast endonuclease

Ashok K. Rout¹, Himanshu Singh^{1,4,6}, Sunita Patel^{4,5}, Vandana Raghvan², Saurabh Gautam⁷, R. Minda², Basuthkar J. Rao^{2,8} & Kandala V. R. Chary^{1,3,4}

Chlamydomonas reinhardtii is a single celled alga that undergoes apoptosis in response to UV-C irradiation. UVI31+, a novel UV-inducible DNA endonuclease in *C. reinhardtii*, which normally localizes near cell wall and pyrenoid regions, gets redistributed into punctate foci within the whole chloroplast, away from the pyrenoid, upon UV-stress. Solution NMR structure of the first putative UV inducible endonuclease UVI31+ revealed an α_1 - β_1 - β_2 - α_2 - α_3 - β_3 fold similar to BoIa and type II KH-domain ubiquitous protein families. Three α -helices of UVI31+ constitute one side of the protein surface, which are packed to the other side, made of three-stranded β -sheet, with intervening hydrophobic residues. A twenty-three residues long polypeptide stretch (D54-H76) connecting β_1 and β_2 strands is found to be highly flexible. Interestingly, UVI31+ recognizes the DNA primarily through its β -sheet. We propose that the catalytic triad residues involving Ser114, His95 and Thr116 facilitate DNA endonuclease activity of UVI31+. Further, decreased endonuclease activity of the S114A mutant is consistent with the direct participation of Ser114 in the catalysis. This study provides the first structural description of a plant chloroplast endonuclease that is regulated by UV-stress response.

Chlamydomonas reinhardtii is a single celled alga that swims with its two flagella and undergoes apoptosis in response to UV-C irradiation¹. The alga shows classical hallmarks of animal cell apoptosis and hence can be used as a model system for studying its molecular mechanism in a plant-like environment. Several candidate molecules such as apoptosis protease activating factor (APAF), a caspase-3 like protein, and a defender against apoptotic death (*DADI*) were studied in *C. reinhardtii*, and found to exhibit a distinct activation pattern correlating with the onset of death upon UV irradiation^{1,2}. One of these candidate molecules, which is yet to be characterized, is referred to as UVI31+ (encoded by *uvi31+* gene, which was originally identified in *Schizosaccharomyces pombe* as an UV-inducible gene)^{3,4}. The *uvi31+* gene is induced by UV light and its expression remains unaltered by other DNA damaging or cytotoxic agents³. However, *uvi31+* does not show any significant homology with any known DNA repair genes⁴. Its expression has been found to be cell-cycle and growth-phase dependent³⁻⁵. The cellular and biochemical characterization of UVI31+ protein revealed interesting biological facets: UVI31+ in *C. reinhardtii* exhibits DNA endonuclease activity and it is induced upon UV stress^{3,6}. We have previously observed that UVI31+ is induced in *C. reinhardtii* when grown in the dark, a common stressor, whereby the protein localization is enhanced near the pyrenoid⁶. UVI31+ localized near the cell wall and pyrenoid regions gets redistributed into punctate foci within the whole chloroplast, away from the pyrenoid, upon UV stress⁶. The observed induction upon UV-stress and the endonuclease activity suggested a plausible role of UVI31+ in DNA repair^{3,6} or some aspects of cellular adaptations.

The study of DNA repair in algae and in higher plants has been slow in comparison with other eukaryotic organisms including yeast. Interestingly, *C. reinhardtii* is routinely used as a model system to study the DNA repair

¹Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai, 400005, India. ²Department of Chemical and Biological Sciences, Tata Institute of Fundamental Research, Mumbai, 400005, India. ³Indian Institutes of Science Education and Research, Berhampur, 760010, Odisha, India. ⁴Tata Institute of Fundamental Research, Center for Interdisciplinary Sciences, Hyderabad, 500075, India. ⁵UM-DAE Centre for Excellence in Basic Sciences, Mumbai University Campus, Mumbai, India. ⁶Department Chemistry and Pharmacy, Ludwig-Maximilians-University, Butenandtstr. 5-13, 81377, Munich, Germany. ⁷Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110093, India. ⁸Indian Institutes of Science Education and Research, Tirupati, 517501, Tirupati, India. Ashok K. Rout and Himanshu Singh contributed equally. Correspondence and requests for materials should be addressed to K.V.R.C. (email: chary@tifr.res.in)

processes, primarily due to the availability of several mutants deficient in various repair processes and their ease of induction and isolation⁷. The most characterized DNA repair process in *C. reinhardtii* is the photo-reactivation repair that repairs thymine dimers in DNA⁸. UV irradiation causes the formation of thymine dimers in DNA, which are repaired in plants through two major mechanisms: photo-reactivation repair and excision repair. The participation of DNA photolyase in repair process is reported in *C. reinhardtii*^{7–9}. DNA endonuclease activity was detected in UVI31+ from *C. reinhardtii*, both endogenously and in its purified form⁶. The UVI31+ from *C. reinhardtii*, when heterologously expressed in *E. coli*, confers to the bacteria around 1000 folds higher resistance to UV⁶. The observed UVI31+ induction upon UV-stress, increased UV resistance upon overexpression in *E. coli*, and the localization near pyrenoid and chloroplast regions suggest a plausible role of UVI31+ in DNA repair^{6,9}.

Given its dynamic cellular localization changes as a function of various stress treatments (UV and dark incubation) and associated endonuclease activity, seeking the structural biology of UVI31+ as an interesting plant protein became highly imminent. Therefore, in the current study, we sought the basic structural biology of UVI31+ from *C. reinhardtii* as a target in order to glean the structural and functional insights of this protein. Moreover, we surmised that the study would enhance our understanding of DNA repair system in plants. Here, we set out to determine the 3D structure of UVI31+ using multi-dimensional solution NMR spectroscopy and study its functional aspects through various biophysical methods and biochemical assays. The study thus provides the structural basis for DNA recognition and endonuclease catalysis by a plant protein UVI31+.

Results and Discussion

Primary structure of UVI31+ reveals short stretches of intrinsically disordered regions. The UVI31+ primary structure revealed three short stretches of intrinsically disordered regions (M24–P31, H59–A73 and T120–Q123) as predicted by *CASP9* and *Metadisorder* web servers (Fig. S1)¹⁰. Rest of the primary structure displayed higher propensity for ordered structure. The primary structure further shows 24% sequence identity and 40.8% sequence similarity with a protein of *BolA* (*pdb id: 2DHM*) family (Scheme S1 and S2)¹¹.

UVI31+ exists as a stable monomer. The UVI31+ was purified using affinity chromatography followed by size-exclusion chromatography and analyzed by SDS-PAGE, mass spectroscopy (MALDI-TOF) and dynamic light scattering (DLS) experiments. During the size-exclusion chromatography, UVI31+ was eluted at a molecular weight corresponding to its monomeric state (Fig. S2A). UVI31+ showed up as a single-band in SDS-PAGE (Fig. S2B), an indication of its purity and exhibited robust DNA endonuclease activity as assessed through “in-gel” assays⁶. Further, the MALDI data showed monomeric state of UVI31+, with a molecular weight (M^r) of 13429.85 Da w.r.t. the expected value of 13300.93 Da (Fig. S2C). The DLS measurements confirmed the monomeric state of the protein with 2.31 nm hydrodynamic radius (R_h) (Fig. S2D). The CD spectrum of UVI31+ suggested that the protein is well-folded (Fig. S2E), while the temperature dependent far-UV CD spectra of the protein revealed its stability with a T_m value of ~54 °C (Fig. S2F).

3D structure of UVI31+. The 3D structure of UVI31+ was determined by simulated annealing procedure with the torsion angle dynamics protocol using the NMR constraints mentioned in Methods section and the program CYANA 3.0^{12,13}. A total of 100 conformers were calculated, from which ten conformers with lowest target function and no distance or angle violations were selected. These ten conformers with lowest target function were further subjected to molecular dynamics simulation in explicit water with NMR-derived distance restraints, angle restraints using CNS 1.21 program¹⁴ and the standard water shell refinement protocol was used^{15,16}. This step improved the Ramachandran plot statistics and also the Z-score for the ordered residues. The Ramachandran plot of the superimposed ensemble showed that 95.2% of residues form part of the most favorable regions of torsion angle space, while 4.7% of the residues were found to be in allowed regions and 0.1% were found in additionally allowed regions. The program PROCHECK-NMR¹⁷ and PSVS-1.4 (http://www.psvs-1_4.nesg.org) were used to validate the quality of the selected ensemble of lowest energy structures of UVI31+. The 3D coordinates of individual atoms of UVI31+ thus obtained were deposited in the PDB (*pdb id: 5ZB6*).

A superposition of ten lowest-energy conformers and a representative lowest energy structure of UVI31+ with an all backbone RMSD of 0.4 Å (ordered region) are shown in Fig. 1A,B, respectively. The 3D structural statistics are provided in Table 1. The overall structural organization observed in the 3D structure of UVI31+ consisted of three α -helices (α_1 : G34–A43 (with an RMSD = 0.12 Å), α_2 : L91–I98 (0.08 Å), α_3 : E105–A108 (0.16 Å)), three β -strands (β_1 : H48–N53 (with an RMSD = 0.14 Å), β_2 : F77–S83 (0.13 Å), β_3 : A112–K118 (0.11 Å)) and a twenty-three residues long flexible loop (D54–H76). Thus, UVI31+ adopts an α_1 - β_1 - β_2 - α_2 - α_3 - β_4 fold, with a 23 amino acid residues long highly flexible loop (D54–H76) between the β -strands β_1 and β_2 . Figure 1C shows residue-wise average backbone RMSD (in Å) derived from the mean structure of the ten energy-minimized conformations of UVI31+, which signify greater flexibility for the polypeptide stretch connecting β_1 and β_2 strands.

UVI31+ is similar to the members of BolA and K Homology domain protein families. *C. reinhardtii* UVI31+ shows substantial structural homology with *BolA* protein from *E. coli* with both of them adopting α_1 - β_1 - β_2 - α_2 - α_3 - β_3 structural fold (Fig. S3A)¹¹. A structure comparison using DALI webserver (http://ekhidna.biocenter.helsinki.fi/dali_lite/start) results a Z-score of 3.3. The structured parts of the two proteins are superimposed with an RMSD of 3.8 Å. *E. coli BolA* and its homologues constitute a widely-conserved *BolA*-like protein family and they have divergent functions^{18–21}. These families of proteins have the ability to impart round cell morphology when over-expressed in bacterial cells following a stress response²². In addition, *BolA* family proteins exert cell division septation control and also show DNA binding ability²³. *BolA* is found in prokaryotes and eukaryotes including *Homo sapiens*. On the basis of functional genomics and 3D structural data, *BolA* was predicted to be a reductase that interacts with a mono-thiol glutaredoxin²⁴. Recently, researchers have found that

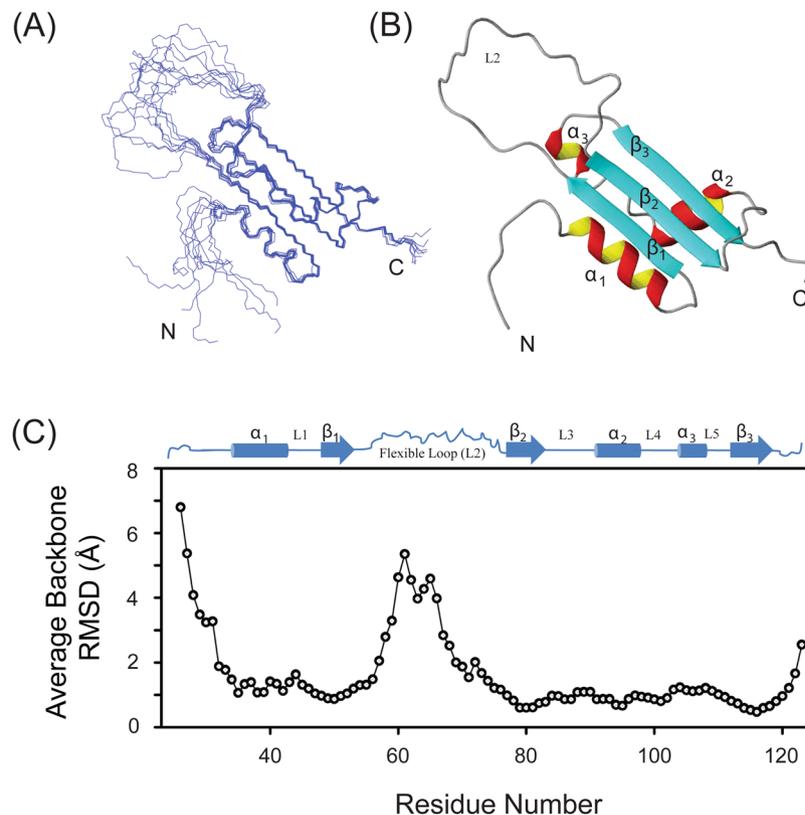


Figure 1. Solution NMR structure of UVI31+. (A) An ensemble of 10 superimposed minimum energy NMR-derived structures of UVI31+. (B) Representative lowest energy structure of UVI31+. (C) Residue-wise plot of RMSD for UVI31+ with NMR derived secondary structural elements shown at the top of this panel.

deletion of *BolA* in *Plasmodium falciparum* resulted in its slow growth, morphological defects and accumulation of high levels of reactive oxygen species²⁵. Thus, BolA-like proteins are considered as potential precursors to new anti-malarial drugs in therapeutics. This highlights the importance of this family of proteins and their fold. Further, BolA proteins have a helix-turn-helix motif, which is a major structural motif with an ability to bind DNA, supporting the postulated function of DNA repair^{11,26}. However, the molecular function of BolA remains unknown till date¹⁹. It is worth to mention here that while the BolA protein has 16 residues long-loop between β_1 and β_2 strands, UVI31+ has a 23 residues long-loop (D54-H76) between β_1 and β_2 strands (Fig. S3A). This loop is poorly defined, primarily due to insufficient nOes seen within this polypeptide stretch, a consequence of greater flexibility. The flexibility of this loop is further confirmed from the ¹⁵N-relaxation data and [¹⁵N, ¹H]-nOe data, as discussed below (Fig. 2). These flexible long-loops of UVI31+ and BolA (*pdb id: 2DHM*) could be superimposed poorly with an RMSD of 2.3 Å and the loop in UVI31+ was found to adopt a more open conformation as compared to that in BolA¹¹.

Another structural homologue of UVI31+ is found to be the K Homology (KH) domain, which also adopts α_1 - β_1 - β_2 - α_2 - α_3 - β_3 structural fold (Fig. S3B)²⁷⁻²⁹. A structure comparison using DALI webserver results a Z-score of 2.3 and have an RMSD of 4.5 Å for the structured region. This KH domain was first identified in the human heterogeneous nuclear ribonucleoprotein (hnRNP) K³⁰. An evolutionarily conserved sequence of around 70 amino acid residues, the KH domain is present in a wide variety of nucleic acid-binding proteins associated with transcriptional and translational regulation, along with other cellular processes^{27-29,31}. KH domains bind either RNA or single stranded DNA^{11,32-34}. The nucleic acid is bound in an extended conformation across one-side of the domain with nonspecific contacts, contributing to its binding specificity. There are two types of KH domains. The type I KH domain has a three-stranded β -sheet with its β -strands in anti-parallel orientation with respect to one another. On the other hand, in the type II KH domains, two of the β -strands belonging to the three-stranded β -sheet are in a parallel orientation^{11,33,35}. Besides, it has been highlighted that β - β - α metal topology seen in KH domains acts as prevalent nuclease domain in many nucleases³⁶⁻⁴⁰.

The orientation of all α -helices and β -strands present in UVI31+ is almost similar to the arrangement in BolA and K-homology type II domain proteins (Fig. S3A,B). While β_1 and β_2 strands are in an anti-parallel orientation, β_2 and β_3 strands are found to be in parallel orientation with respect to each other (Fig. 1B). The long loop connecting β_1 and β_2 is missing in both BolA and KH domain. The orientation of the α -helices (α_1 , α_2 and α_3) in the UVI31+ is similar to that of BolA and K-homology domain protein family^{11,28}. In UVI31+, the three α -helices form one side of the protein surface, while the other side is formed by three β -strands and a strong hydrophobic core packed between them, providing a compact 3D structure (Fig. 1A,B). Most of the hydrophobic

Conformationally restricting restraints	
Distance Restraint List	
Total	1254
Intraresidue (i = j)	298
Sequential (i - j = 1)	646
medium-range	204
long-range	106
Hydrogen bonds	42
Dihedral angle restrains (ϕ and ψ)	102
Disulfide restraints	0
No. of restraints per residue	12.66
Residual restraint violations	
Average no. of distance violations per structure > 0.5 Å	0
Average no. of dihedral angle violations per structure > 5°	0
Average no. of Van der Waals violations per structure > 0.5 Å	0
Model quality	
Rmsd backbone atoms (Å)	0.4
Rmsd heavy atoms (Å)	0.9
Rmsd bond lengths (Å)	0.004
Rmsd bond angles (°)	0.7
MolProbity Ramachandran statistics^a	
Most favored region (%)	95.2
Allowed region (%)	4.7
Additionally allowed region (%)	0.1
Disallowed region (%)	0.0
Global quality scores (raw/Z score)	
Verify3D	0.15/−4.98
PROCHECK (ϕ - ψ)	−0.11/−0.12
PROCHECK (all)	−0.66/−3.90
MolProbity clash score	27.60/−3.21
Model contents	
Total no. of residues	100
BMRB accession number	16864
PDB ID code	5ZB6

Table 1. NMR structural statistics for the ensemble of 10 refined conformers of UVI31+ (*pdb id:5ZB6*).

^aOrdered residues: 33–53, 76–122.

residues (Ala, Leu, Ile, Val, Pro, Phe, Trp and Met) form the central core of the protein and are buried at the interface of α -helices and β -sheet. The highly flexible loop (D54-H76) between β_1 and β_2 strands is devoid of branched hydrophobic residues and it has only four Ala residues spaced widely apart suggesting its intrinsically disordered nature.

Modulation of conformational dynamics of UVI31+. The ¹⁵N relaxation data aid in probing milliseconds to picoseconds motions and provide information about the overall and internal motions in a given protein and thus are very crucial for understanding the protein dynamics⁴¹. The ¹⁵N-longitudinal relaxation rates (R_1) and [¹⁵N, ¹H]-nOe values provide information about fast time-scale dynamics in the range of nanoseconds to picoseconds time-scale. On the other hand, the ¹⁵N-transverse relaxation rates (R_2) are largely sensitive to slow motions and reflect slower exchange processes that occur in the milliseconds to microseconds time scale⁴². The R_1 and R_2 values determined for individual residues present in the UVI31+ show large variations all along the sequence (Fig. 2A,B). The average R_1 and R_2 values for the residues involved in highly structured regions were found to be $1.14 \pm 0.17 \text{ s}^{-1}$ and $13.77 \pm 1.20 \text{ s}^{-1}$, respectively, while for the loop connecting β_1 and β_2 strands these values were $1.69 \pm 0.33 \text{ s}^{-1}$ and $10.69 \pm 2.39 \text{ s}^{-1}$, respectively. The relative higher R_1 values and lower R_2 values for the loop connecting β_1 and β_2 strands indicate that the loop is highly flexible, as discussed. This observation is further supported by a low average R_2/R_1 value of 6.85 ± 2.46 for the loop as compared to that of the structured regions, which was found to be 12.35 ± 2.04 (Fig. 2C). The flexibility of the polypeptide stretch D54-H76 was also evident from the observation of a low average value for the [¹⁵N, ¹H]-nOe, which was found to be 0.64 ± 0.14 as compared to that of the structured regions, which was found to be 0.87 ± 0.05 (Fig. 2D). In addition, the negative and low nOe values seen near the N- and C-terminals ends of UVI31+ indicate their flexibility too. Taken

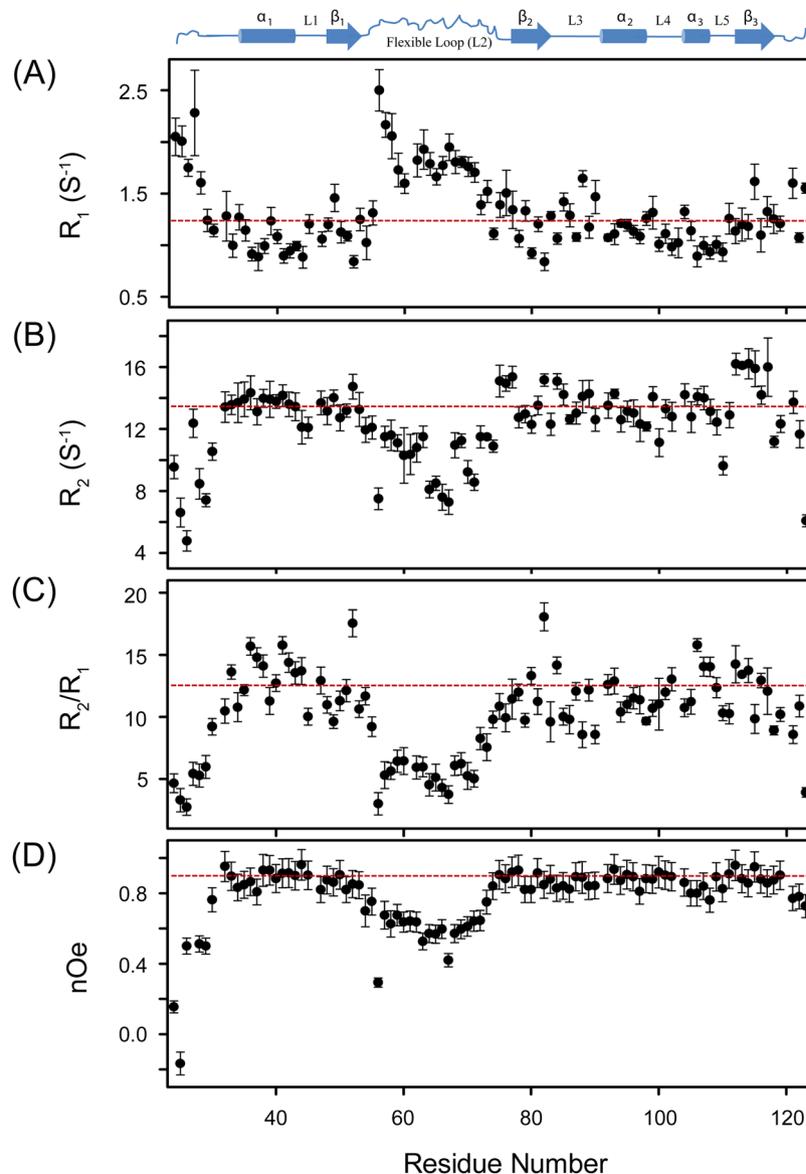


Figure 2. The ^{15}N relaxation data of UVI31+ recorded on a Bruker Avance 800 MHz NMR spectrometer at 25°C. (A) Longitudinal relaxation rates, R_1 ($1/T_1$), (B) Transverse relaxation rates, R_2 ($1/T_2$), (C) Residue-wise plot of R_2/R_1 , (D) ^{15}N , ^1H -nOe defined as $I_{\text{sat}}/I_{\text{eq}}$, where I_{sat} and I_{eq} are the measured intensities of peaks in the 2D experiments with and without proton saturation, respectively. The relaxation data could not be estimated for Pro residues (P31, P46, and P120) and for residues H61, L91, and S103 due to spectral overlap or weak intensities resulting improper fit. NMR derived secondary structural elements of UVI31+ are shown at the top of the panel A.

together, both the NMR derived 3D structure and ^{15}N -relaxation data of UVI31+ clearly suggest that the protein is quite ordered in entirety except for the terminal ends and the polypeptide stretch (D54-H76) connecting β_1 and β_2 strands present in the protein.

Electrostatic surface charge potential distribution in UVI31+. In order to map potential DNA binding surface of the protein, we calculated the electrostatic surface charge potential for UVI31+. As shown in the Fig. 3, the highly flexible disordered loop (D54-H76) showed significant amount of negatively charged surface potential (shown in red). On the other hand, we observed pronounced positively charged surface potential (in blue) right at the center of the protein, forming a distinct cleft (Fig. 3). The residues in and around this cleft belong to the N-terminal segment (M24-A35), proximal α_1 helix and β_1 strand (H48-N53). The residues that significantly contribute to the positively charged cleft are K37, K39, K50, K57, R78 and K107. This cleft is the most plausible binding site for the negatively charged DNA/RNA, facilitating the endonuclease activity of the protein, which we probed further.

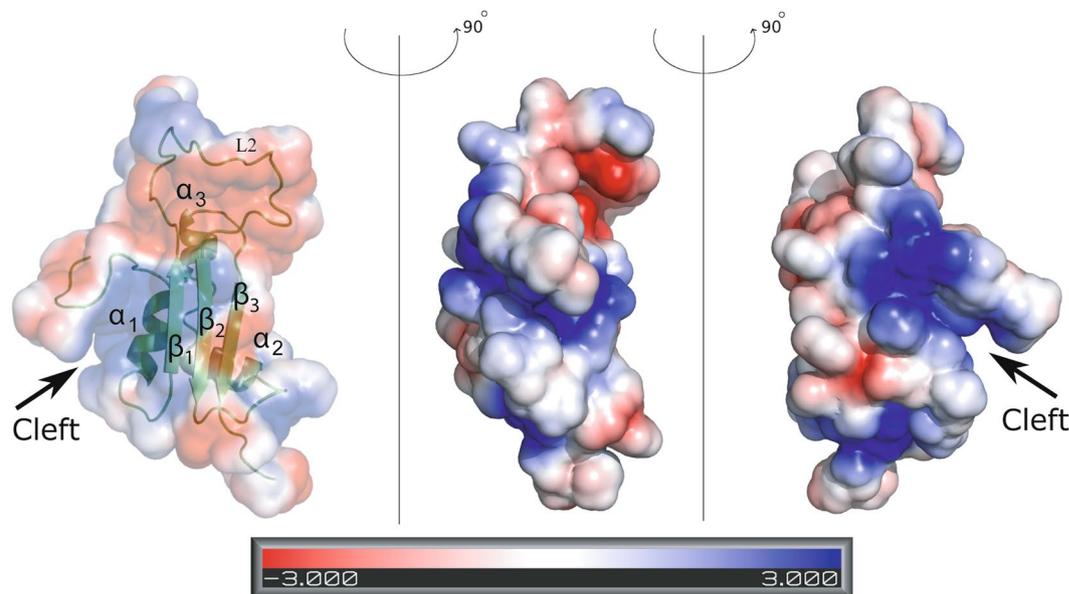


Figure 3. Surface electrostatic charge potential of UVI31+. The spectral bar at the bottom indicates the electrostatic potential in electron volts (eV). Red color represents negative charge potential (-3 eV) while blue color represents positive charge potential (3 eV).

Residues belonging to the β -sheet and the long loop regions undergo perturbation upon DNA binding.

To test the DNA binding affinity of UVI31+, ITC experiments were performed with a self-complementary ds-(CGCGAATTCGCG) as a template to measure apparent dissociation constant (K_d) (Fig. 4). The ITC isotherms showed that UVI31+ binds to ds-DNA exothermically (Fig. 4 and Table S1), with effective K_d value of $5.36 \mu\text{M}$. The binding occurs at two sets of sites (Table S1).

The gel-filtration chromatography of the purified UVI31+ protein showed that its biochemical endonuclease activity was associated with monomeric form of the protein, as was reported earlier⁶. However, the identity of the residues participating in the DNA recognition was not known. Nuclease activity of UVI31+ with DNA fragments of varying lengths showed that the protein binds strongly to 12-bp DNA and the binding was not detectable when DNA size was less than 10-bp and as discussed below, UVI31+ has no strong target preference for either ssDNA or dsDNA (Fig. S4). With this in the backdrop, the 12 mer ds-DNA mentioned above was chosen for interaction studies with UVI31+. In the present study, we could identify these residues by recording a 2D [¹⁵N, ¹H]-so-fast-HMQC, with an acquisition time of 2 s. An overlay of [¹⁵N, ¹H]-so-fast-HMQC spectra of UVI31+ and UVI31+:DNA complex (Fig. 5A) showed up significant CSPs indicating the interaction of UVI31+ with the ds-DNA. Residue-wise CSPs calculated as described in the Methods section enabled us to identify the interacting residues of UVI31+ with ds-DNA. As is evident from the CSP plot (Fig. 5B), residues belonging to β_1 (F49 and K50), β_2 (L79), β_3 (K118 and T119) strands, four residues (K57, H58, A59 and H76) of the flexible long-loop (D54-H76) between β_1 and β_2 strands and two residues (A35 and K37) belonging to the α_1 helix undergo significant CSPs upon DNA binding (Fig. 5C). We highlight here that out of these eleven residues, six are positively charged. Three of these residues K37, K50 and K57 contribute significantly to the total positive surface-charge potential and form part of the cleft mentioned earlier (Fig. 3). These residues interact with the negatively charged DNA. As discussed earlier, though the protein possesses three α -helices, only two residues belonging to the α_1 -helix were found interacting with DNA, suggesting no involvement of α_2 - and α_3 -helices in DNA recognition. Other residues, which showed CSPs, are those belonging to the flexible loop (K57, H58, A59 and H76). Topologically, these residues are situated in close proximity to the β -sheet (Fig. 5C). The K35 belonging to the α_1 -helix is in close proximity to K50 of β_1 strand and L79 of β_2 -strand with the corresponding C $^\alpha$ -C $^\alpha$ distances 11.0 and 7.6 Å, respectively. The residues belonging to α_1 -helix, which show chemical shift perturbations, are in close proximity to the flexible loop. All these observations taken together suggest that the DNA recognition is with the loop and β -sheet domain of UVI31+.

Taking into account the fact that few residues belonging to the flexible long-loop (D54-H76) undergo CSPs upon interaction with the DNA, we attempted to study the influence of the flexible long-loop on the endonuclease activity. In this endeavor, shortening of the 23 residues long-loop connecting β_1 and β_2 strands to a 4 residues loop (⁵⁴DSGG⁵⁷) (Fig. S5A), as described in the Methods section, did not affect the endonuclease activity. The zymogram convincingly showed that the DNA endonuclease activity of “loop-null” mutant is comparable to that of the wild-type UVI31+ (Fig. S5B).

S114A mutation drastically reduced endonuclease activity of UVI31+. With the 3D structure in hand and using the computational method called CLASP (CataLytic Active Site Prediction), we attempted to uncover putative residues, namely Q96, Y99, E105 and S114 that showed good congruent matches for nuclease

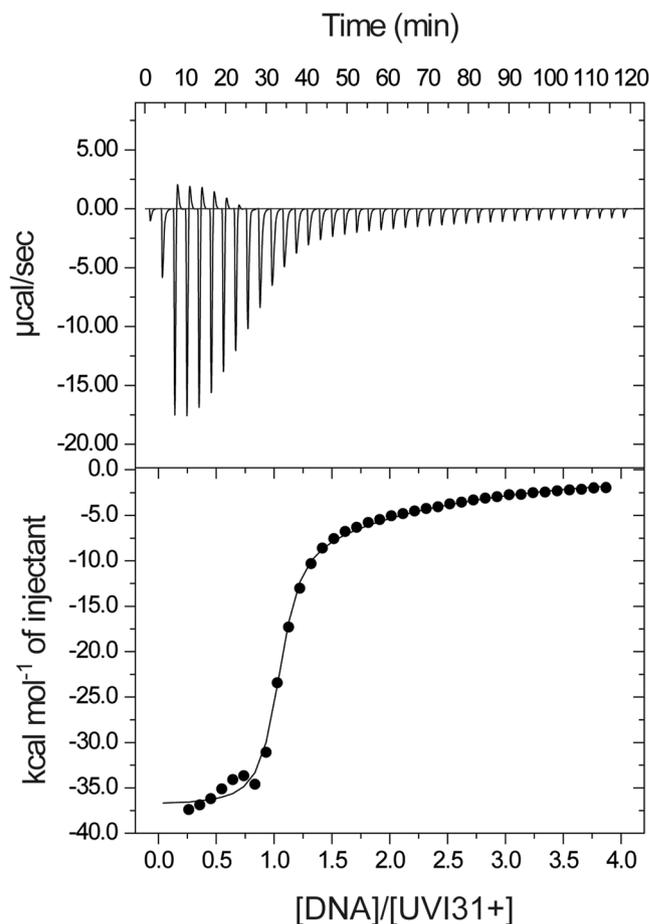


Figure 4. ITC data of the binding of UVI31+ to ds-DNA. The thermodynamic parameters describing the fit are presented in Table S1.

activity⁴³. This is based on the spatial and electrostatic properties of the probable catalytic residues⁴³. Out of all these, we speculated that the S114 could be responsible for the endonuclease activity as the Ser hydroxyl group is implicated as a nucleophile in the catalysis of nucleases action³⁶.

On the other hand, it is well known fact that an acid-base-nucleophile catalytic triad is a group of three amino-acid residues that are found in and around the active site of several nucleases and some proteases⁴⁴. Such a triad is a common motif involved in generating a nucleophilic residue that is needed for covalent catalysis⁴⁴. The side-chain of the nucleophilic residue performs covalent catalysis of the substrate. The lone-pair of electrons present on the oxygen or sulphur attacks the electropositive carbonyl/phosphoryl group. The “Ser-His-Glu/Asp/Thr” motif is one of the most thoroughly characterized catalytic triad^{45–49}. As mentioned earlier since the side-chain hydroxyl group of a Ser is implicated as a nucleophile in the catalysis of nucleases action, one could in principle identify a possible catalytic triad by searching around for the partner residues (His and Glu/Asp/Thr) in the vicinity of any Ser present in a given protein. For example, in UV-damage endonuclease (UVDE, *pdb id*: 2J6V) protein, the S234, which was earlier implicated in the endonuclease activity, was considered to be part of a catalytic triad (Ser-His-Glu/Asp/Thr)⁵⁰. This residue was used as a reference in identifying other two partner residues (His-Glu/Asp/Thr) in its vicinity. Such a search revealed H244 and E269 as partner residues. Likewise, in the present study, the putative endonuclease active site in UVI31+ was predicted by identifying a potential catalytic triad (Ser-His-Glu/Asp/Thr) likely to be present in its 3D structure. In UVI31+, there are eight Ser residues. By taking each of these Ser residues as a reference, search was made to identify the putative catalytic triad (Ser-His-Glu/Asp/Thr). This resulted in the prediction of the residues S114, H95 and E80/T116 as the only probable constituents of the catalytic triad (Ser-His-Glu/Thr). A closer examination of the 3D structure revealed that S114, H95 and T116 are in the cleft of UVI31+ as discussed above while E80 is protruding out of the β_3 -strand of the three-stranded β -sheet and exposed to the solvent. Thus, we speculated S114-H95-T116 as the catalytic triad and set out to carry out functional assessment of the promiscuous S114 belonging to the S114-H95-T116 triad by mutating Ser114 to Ala. It is worth to mention here that ‘His’ side-chains play an important role in most of the enzymatic catalytic processes. The pKa of side-chain imidazole proton of free ‘His’ is around 6.2. However, during catalysis it has been observed to cover a pKa range from 6.9 to 7.9. In the case of the catalytic triad involved in DNA binding, ‘His’ pKa value was reported to be 7.3^{51–54}. In our proposed catalytic triad, we are expecting ‘His’ pKa value to be in the same range.

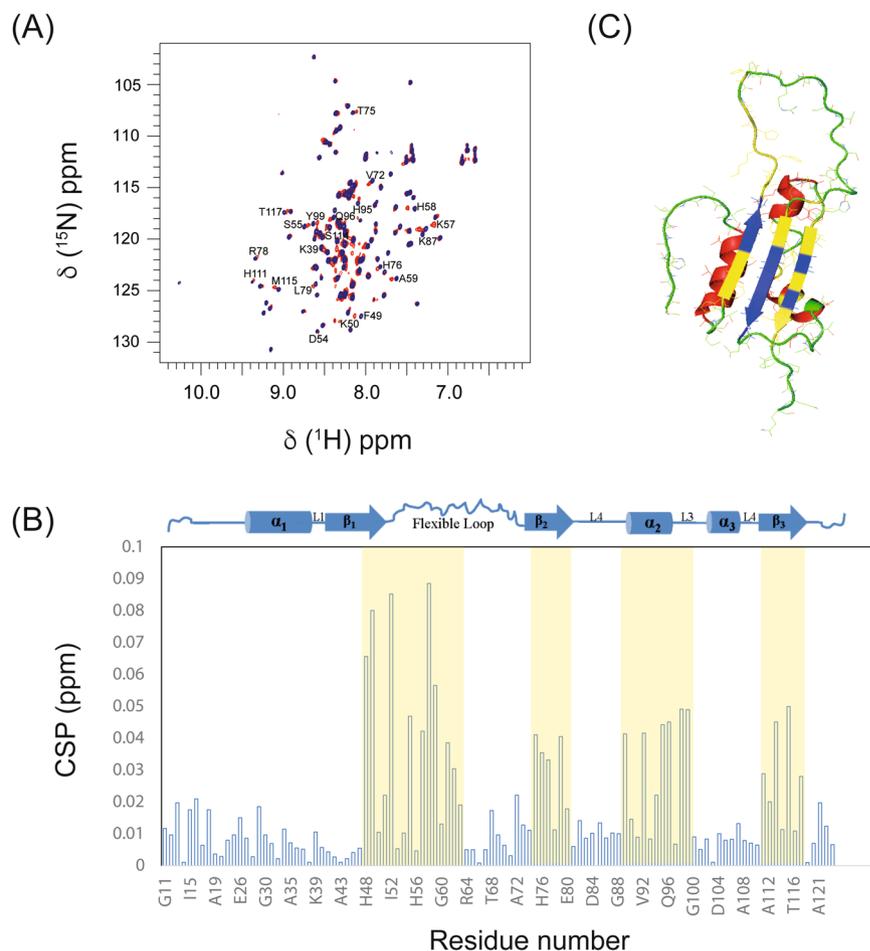


Figure 5. UVI31+ and DNA interaction by NMR. (A) Overlay of 2D [^{15}N , ^1H]-so-fast-HMQC for UVI31+ (black) and UVI31+ : DNA complex (red). (B) Residue-wise plot of CSPs upon UVI31+ : DNA complex formation. Light yellow bar represents residues involved in DNA interaction. The secondary structural elements of the NMR derived structure of UVI31+ are depicted on top of the panel B. (C) The α -helices, β -strands and loops are shown in red, blue and green, respectively. The residues undergoing perturbation upon DNA binding are shown in yellow.

CD experiments with this single mutant of UVI31+ (S114A-UVI31+) showed no change in the secondary structure as compared to the wild-type (Figs S6A and S2E). The melting temperature (T_m) of the mutant and wild-type proteins calculated from the CD data were found to be almost same ($\sim 54^\circ\text{C}$) (Figs S6B and S2F). Further, the MALDI data showed the mutant is in monomeric state with a molecular weight (M^r) of 13414.52 Da and the mass is less than 16 Da from the wild-type because of serine to alanine mutation (Figs S6C and S2C). Overlay of 2D [^{15}N - ^1H]-so-fast-HMQC of UVI31+ with that of the S114A-UVI31+ showed similar spectral signatures except for the residue mutated, indicating that there are no major structural differences between the wild-type protein and its S114A-UVI31+ (Fig. S6D). These results show that S114A mutant is structurally similar to the wild-type protein.

It is relevant to reiterate here that using a computational method called CLASP (Catalytic Active Site Prediction)⁴³, we had pinpointed S114 as a potential catalytically relevant Ser amongst eight available Ser residues in the protein. As predicted, S114A mutation abolished the endonuclease activity, thereby strengthening our prediction. With the 3D structural elucidation (current study), it became feasible to narrow down the putative catalytic triad partners of Ser114. Our structure based search uncovered that S114-H95-T116 could be the best suited triad in UVI31+ that satisfies the geometric constraints of acid-base catalysis requirement which is consistent with CLASP results as well as S114A loss of activity. Therefore, we strongly suggest that S114-H95-T116 is the putative catalytic triad involved in UVI31+ endonuclease activity.

Further, we noticed that the identified catalytic triad (S114-H95-T116) is located in the positively charged cleft, which is shown as blue surface potential in the UVI31+ 3D structure (Fig. 3), hinting its involvement in the interaction with the DNA. There are several reports in literature highlighting the involvement of Ser in hydrogen bonding interactions with the nucleotide bases^{55–58}. DNA titrations with S114A mutant showed minimal CSPs, as shown in the overlay of 2D [^{15}N , ^1H]-so-fast-HMQC spectra (Fig. S6E). Overlay of residue-wise CSPs of UVI31+ and S114A-UVI31+ with ds-DNA, showed significantly small CSPs for the residues belonging to β_1 (F49 and K50), β_2 (L79), β_3 (K118 and T119) strands, four residues (K57, H58, A59 and H76) of the flexible long-loop

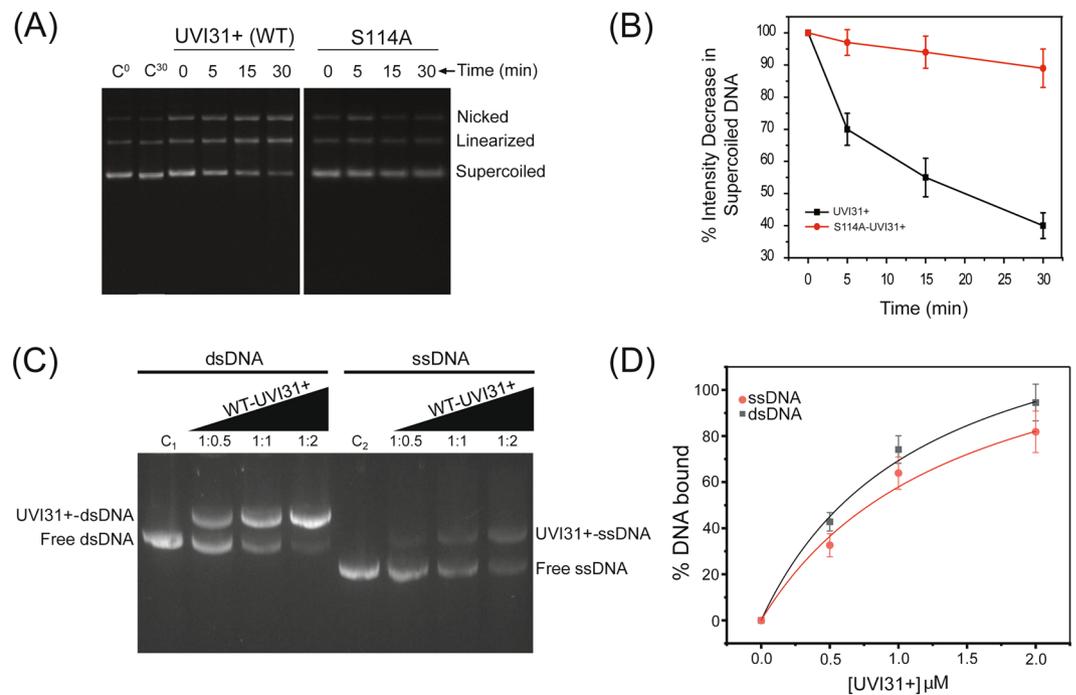


Figure 6. UVI31+ target DNA analyses: ssDNA versus dsDNA. **(A)** UVI31+ endonuclease activity as studied by agarose gel (1%) assay for wild-type UVI31+ and its S114A mutant. The image presented here is from a single agarose gel, cropped and grouped together with white space. The white space represents some of the lanes removed, which contained markers that are not needed for the results presented here. **(B)** Percentage band-intensity decrease of supercoiled DNA plotted as a function of reaction time for wild-type UVI31+ and its S114A mutant. Error bars represent the standard deviation derived from three independent experiments. C⁰ and C³⁰ shown in the gel are the plasmids at 0 and 30 min, respectively. **(C)** Electrophoresis mobility shift assay (EMSA) for DNA-UVI31+ interaction studies. Increasing concentration of UVI31+ (0.5 to 2 μM) was used for observing interaction with 50-bp dsDNA or ssDNA using 2% agarose gel electrophoresis. DNA concentration was kept fixed (1.0 μM) in both sets of experiments with increasing UVI31+ protein concentration thereby providing DNA to protein ratios of 1:0.5, 1:1, and 1:2, respectively. C₁ and C₂ lanes are the control dsDNA and ssDNA bands without protein, respectively. The image presented here is the full-length agarose gel, cropped for presentation point of view. **(D)** Quantification of the gel shift assay was performed via intensity analysis using ImageJ. Results are presented as an average of three measurements and the K_d values are presented in Table 2.

Conditions	K _d (μM)	R value
UVI31 + -ssDNA	1.4 ± 0.3	0.98
UVI31 + -dsDNA	1.1 ± 0.3	0.99

Table 2. K_d values as determined by gel shift assay (EMSA).

(D54-H76) between β₁ and β₂ strands and two residues (A35 and K37) belonging to the α₁ helix in the case of S114A-UVI31+ mutant protein with ds-DNA (Fig. S6F).

Thus, the NMR data suggests that S114A mutant may cause an altered DNA-binding specificity, thereby drastically reducing the endonuclease activity, as seen in DNA endonuclease activity assay (Fig. 6A). It is evident from Fig. 6 that the super-coiled plasmid DNA pBR322 is nicked into its linear and open circular forms by wild-type protein within a time interval of 30 min. In an identical assay with S114A mutant, the nicking activity of the super-coiled plasmid DNA got reduced by ~85% as compared to that of wild-type protein (Fig. 6B). Further, we have performed gel shift assays with ssDNA and dsDNA (Fig. 6C). These gel shift assays clearly demonstrate that UVI31+ binds to both dsDNA and ssDNA. With increasing protein concentration, the intensity of the upper band (complex) got enhanced with concomitant reduction in the intensity of lower band (unbound DNA), demonstrating the formation of protein-DNA complexes with both ligands (dsDNA and ssDNA). Partial loss in the intensity of ssDNA (bound as well as free) as compared to dsDNA could be a result of higher endonuclease activity in the case of ssDNA, which is consistent with our other data (Fig. S4), as discussed in the above paragraphs. We notice that ssDNA versus dsDNA differences may not be very significant: Firstly, the native gel assay might mask internal cleavages in dsDNA and secondly, the ssDNA templates might have residual secondary structures, unless these are homo-polymeric sequences, which in turn present other challenges. Notwithstanding these caveats, direct comparison by gel-shift assay revealed only a minor binding preference of dsDNA over ssDNA. Quantification of the gel shift assay resulted in determining K_d values for ssDNA and dsDNA, which

turned out to be as $1.4 \pm 0.3 \mu\text{M}$ and $1.1 \pm 0.3 \mu\text{M}$, respectively (Fig. 6D and Table 2). All these observations put together, point towards UVI31+ protein as DNA endonuclease with no strong target preference for either ssDNA or dsDNA. It is worth mentioning here that the S114-UVI31+ equivalent is missing in BolA (*pdb id: 2DHM*) protein, rendering the protein deficient in endonuclease function although the protein is structurally akin to UVI31+. Interestingly, we did observe partial conservation of catalytic triad in BolA structure (*pdb id: 2DHM*) with an A85 in an equivalent structural position of S114 (Fig. S7).

Conclusion

We have determined NMR-derived 3D structure of the UV inducible gene product, UVI31+ protein, from *Chlamydomonas reinhardtii*, which possesses a $\alpha_1\text{-}\beta_1\text{-}\beta_2\text{-}\alpha_2\text{-}\alpha_3\text{-}\beta_3$ fold, similar to the fold found in BolA protein family and KH domain type II proteins. The structural elucidation uncovered two important features of the UVI31+ protein structure. Though the domain structure is very similar to that of BolA protein family and KH domain proteins, there were interesting differences vis-à-vis the same with respect to the flexible regions of UVI31+. Further, UVI31+ is found to recognize the DNA primarily by its β -sheet domain. The single point mutation (S114A) showed drastic loss of DNA endonuclease activity in UVI31+, whereas deletion of the long flexible loop had no effect on the endonuclease activity. Based on the 3D structure elucidated in the current study, and the geometric and distance constraints, S114-H95-T116 triad is the putative catalytic triad facilitating the endonuclease activity of UVI31+ protein. Interestingly, the catalytic triad S114-H95-T116 is strategically located in the positively charged cleft of UVI31+ 3D structure. Thus, the 3D structure of UVI31+ reveals a compact core that harbors the nuclease triad residues and a long disordered loop, which imparts the required flexibility to interact with other bio-macromolecular surface such as pyrenoids in the cell. Further, this study reveals the first structural description of a plant chloroplast endonuclease that is regulated by UV-stress response in *C. reinhardtii* algal plant. Future studies focused on details of UVI31+ endonuclease mechanism in the context of chloroplast biology will critically hinge on the current results.

Material and Methods

Cloning, overexpression, isolation and purification. The cDNA encoding *uvi31+* was cloned into a pET28a expression vector and transformed into *E. coli* strain BL21(DE3) cells for over-expression, isolation and purification of the protein as described earlier⁵⁹. Similar procedure was used to clone, overexpress, isolate and purify the S114A mutants of UVI31+⁶⁰. Besides, yet another variant of UVI31+, where in the polypeptide stretch ⁵⁴DSHKHAGHYARDGSTASDAGETH⁷⁶ was replaced with ⁵⁴DSGG⁵⁷ and overexpressed, isolated and purified following the same protocol as that of wild-type protein. All of these mutations were confirmed by DNA sequencing.

Dynamic light scattering. The hydrodynamic radii (R_h) of UVI31+ and its mutants were determined using dynamic light scattering (DLS) method on a Dynopro-LS instrument at 830 nm. The protein samples were centrifuged at 13,000 rpm for 10 min and filtered through a syringe filter of 0.45 μm (Millipore) before transferring into a quartz cuvette. The measurements showing parabolic curves with a straight baseline were considered for the estimation of the mean values of R_h and associated standard deviations. Regularization plots were used to determine the respective R_h .

Circular dichroism (CD) experiments. CD spectra were recorded on a JASCO J-810 spectropolarimeter equipped with a peltier cell temperature controller. For recording the CD spectra in the far-UV region (250–200 nm) 0.1 cm path length quartz cell cuvette was used, while in the visible region (300 to 700 nm) a 1 cm path length quartz cell cuvette was used. The CD parameters chosen were as follows: scan speed of 20 nm/min, time constant of 1.0 s, 1.0 bandwidth, and sensitivity of 100 mdeg. The CD spectra were analyzed using Yang's method to estimate the secondary structure content of UVI31+ and its mutants. All these experiments were conducted under a flow of pure nitrogen. A good signal-to-noise ratio in the CD spectra in both the spectral ranges were obtained upon data averaging over three scans. The protein concentrations used for far-UV and visible CD studies were of 20 μM .

Thermal unfolding. The temperature dependence of the visible CD was monitored to address thermal unfolding of UVI31+ and its mutants. The temperature was increased from 20 to 90 °C with a heating rate of 1 °C/min. The protein samples were equilibrated at each temperature for at least 3 min, and the reversibility of the unfolding was ensured by decreasing the temperature with a cooling rate of 1 °C/min. In thermal unfolding experiments, fractions of the unfolded protein f_u , at different temperatures (T), were calculated as the ratio of $(\Theta_t - \Theta_u)$ and $(\Theta_F - \Theta_u)$, where, Θ_t is the observed ellipticity at any temperature, Θ_F is the ellipticity of the fully folded form and Θ_u is the ellipticity of the unfolded form.

NMR spectroscopy. For NMR studies, uniformly ¹⁵N-labelled (u-¹⁵N) and ¹³C/¹⁵N-doubly-labelled (u-¹³C/¹⁵N) UVI31+ and its mutants were prepared in a mixed solvent of 90% H₂O and 10% ²H₂O (50 mM sodium phosphate, 50 mM NaCl (pH = 6.4)) as described earlier⁵⁹. All NMR experiments were carried out at 25 °C with protein concentrations between 0.5 and 0.6 mM on a Bruker Avance 800 MHz NMR spectrometer equipped with a 5 mm cryogenically cooled triple-resonance probe and a pulse-field gradient. A suite of 3D double- and triple-resonance NMR experiments were performed for sequence specific ¹H, ¹³C and ¹⁵N backbone resonance assignments as discussed earlier^{61,62}. In addition, we recorded 3D experiments such as HCCH-COSY/TOCSY, [¹⁵N, ¹H]-NOESY-HSQC ($\tau_m = 80$ ms) and [¹³C, ¹H]-NOESY-HSQC ($\tau_m = 120$ ms) for almost complete assignment of ¹H, ¹³C and ¹⁵N side-chain resonances and for the measurement of nOes used in the 3D structural calculation of the protein. The near complete ¹H, ¹³C and ¹⁵N resonance assignments of UVI31+ and its variants

were deposited earlier in BMRB (<http://www.bmrwisc.edu>; under the accession numbers 16864 (for wild-type) and 18567 (S114A mutant)^{59,60}. The backbone ^{15}N T_1 relaxation measurements at 800 MHz were acquired using 256^*1024 complex points along t_1 and t_2 dimensions, respectively, and inversion recovery delays of 50, 100, 200*, 300, 400, 500, 600*, 800 and 1020 ms. The ^{15}N T_2 measurements were carried out with the same acquisition parameters using CPMG pulse sequence with relaxation delays of 10, 30, 50*, 70, 90, 110*, 130, 150, 170 and 190 ms. In both the experiments, delays marked with an asterisk were recorded twice for error calculation. Steady-state [^{15}N , ^1H] heteronuclear nOe measurements were carried out with and without proton saturation during the relaxation delay. In these nOe experiments, a 2.5 s of proton saturation was used. The heteronuclear nOe values were determined as the ratio of the peak intensities measured from the spectra acquired with and without proton saturation. NMR spectra were processed using Felix 2002 (Accelrys Inc., San Diego) and analyzed using Topspin 2.0 and 3.1 (Bruker BioSpin: <http://www.bruker-biospin.com/>), TATAPRO⁶³ and CARA⁶⁴. The ^1H chemical shifts were referenced with respect to an external standard 2, 2-dimethyl-2-silapentene-5-sulfonates (DSS), while the ^{13}C and ^{15}N chemical shifts were referenced indirectly⁶⁵.

NMR structure calculation. The 3D solution structure of UVI31+ was determined using the following NMR constraints: (i) Dihedral angle constraints derived using TALOS with the knowledge of individual $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$, ^{13}C , chemical shift values as input⁶⁶. A total of 102 ϕ and ψ dihedral angle constraints were used; (ii) Generic hydrogen bond (H-bond) constraints were imposed for residues located at well-defined α -helical and β -strand regions. An upper limit of 2.0 Å was used for H-O distance in all hydrogen bond constraints. Total number of H-bond constraints used was 42; (iii) Cross peaks in NOESY spectra were identified, assigned and the corresponding peak intensities were translated into ^1H - ^1H distances. A total of 1254 distance constraints, which included 298 intra-residue, 646 inter-residue (sequential), 204 medium-range, and 106 long-range distance constraints were used in the 3D structure calculation. With all these restraints as input, the 3D structure of UVI31+ was calculated using the program CYANA 3.0^{12,13}. The structure figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and MOLMOL⁶⁷. The electrostatic surface charge distribution plot was made from the NMR derived structure of UV31+ using APBS tool of PyMOL⁶⁸.

UVI31+ and DNA interaction by NMR. In order to study the interaction of UVI31+ and DNA, we added 2 μL of 2 mM self-complementary ds 5'-(CGCGAATTCGCG)-3' DNA (12 mer ds-DNA) with 200 μM (^{15}N)-UVI31+ and equilibrated at 25 °C and followed it by recording a series of 2D [^{15}N , ^1H]-so-fast-HMQC spectra at different concentrations of the ds-(CGCGAATTCGCG) DNA⁶⁹⁻⁷¹. The chemical shift perturbations (CSPs) for all the individual residues of UVI31+ were monitored with the knowledge of wild-type UVI31+ $^1\text{H}^{\text{N}}$ and ^{15}N resonance assignments^{59,60}. The CSPs were measured as $[(\Delta\text{H})^2 + (\Delta\text{N}/10)^2]^{1/2}$, where, ΔH and ΔN signify the changes in $^1\text{H}^{\text{N}}$ and ^{15}N chemical shifts, respectively. The factor 10 for ^{15}N chemical shift was taken as the normalization factor, since the overall range of nitrogen chemical shifts is roughly 10 times that of proton chemical shifts for the backbone amides in folded proteins.

DNA endonuclease activity assay. The DNA gel assay is a classical way of showing the endonuclease action of a protein on DNA substrate. Conversion of supercoiled DNA to nicked circular DNA followed by linear form of DNA establishes DNA nicking as a function of time⁴³. For agarose gel assay, each reaction mixture (20 μL total volume) contained 300 ng of negatively super-coiled pBR322 DNA, taken in 50 mM sodium phosphate (pH 7.6), 50 mM NaCl, 1 mM MgCl_2 and UVI31+ or its variants. It was followed by incubation of the reaction mixture at 37 °C for 30 min. Reaction was quenched by adding 5 μL stop solution (10% glycerol, 0.005% bromophenol blue, 0.1% SDS). DNA samples were analyzed by gel electrophoresis at 3 V/cm for 4 hr on a 0.8% agarose gel taken in Tris-acetate-EDTA buffer [40 mM Tris-acetate, 1 mM EDTA (pH 8.0)]. The gel was stained in ethidium bromide solution (0.5 $\mu\text{g}/\text{mL}$) for 30 min, and finally visualized on an ultra-violet trans-illuminator.

Nuclease assay on single and double stranded DNA. UVI31+ (23 μM) was incubated with 300 ng each of linear pUC19 double stranded (ds) DNA and M13mp18 single stranded (ss) DNA for various times from 0–60 minutes in a buffer containing 50 mM sodium phosphate (pH 7.6) and 50 mM NaCl, 10 mM MgCl_2 in a reaction volume of 20 μL at 37 °C. The reaction was stopped by adding 0.1% SDS and analyzed by electrophoresis in a 1% agarose gel containing 0.5 $\mu\text{g}/\text{ml}$ EtBr. The images were analyzed by FIJI software and plotted as a function of reduction of ss- or ds-DNA band intensity with time and normalized to starting amounts of 100% of each substrate.

Electrophoresis mobility shift assay (EMSA). EMSA was carried out as described previously⁷². Briefly, 1 μM DNA (single stranded or double stranded) was incubated with varying concentrations of UVI31+ (0.5, 1, and 2 μM) in 10 mM sodium-phosphate buffer (pH 7.5) containing 10 mM NaCl and 1 mM MgCl_2 in a total volume of 20 μL . The reactions were incubated for 5 min at 37 °C and were analyzed using 2% agarose gel electrophoresis by loading 10 μL from each reaction to the respective gel wells. Control reactions without any protein were also carried out for comparison. Quantification of the gel shift assay data was carried out using ImageJ⁷³. Binding data were fitted to the equation ($y = V_{\text{max}} * x / (K_d + x)$) describing a sigmoidal curve⁷² and K_d values were calculated from the fit of the curve using Origin Pro 2018 software.

Interaction of UVI31+ with DNA by Isothermal Titration Calorimetry (ITC). We used ITC to characterize quantitatively the thermodynamics of UVI31+ binding to a self-complementary 12 mer ds-DNA. ITC experiments were performed using a VP-ITC Micro-Calorimeter (MicroCal Inc., Northampton, MA, USA) with UVI31+ (taken at a concentration of 75 μM) and self-complementary ds-(CGCGAATTCGCG) DNA taken at a concentration of 2 mM. The protein and the DNA were extensively dialyzed against similar buffer, containing 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5% glycerol, 5 mM MgCl_2 , 0.1 mM EDTA, prior to the performance

of experiments, to avoid heat signals that could arise while mixing nonequivalent buffers. All solutions were carefully degassed before each titration using equipment provided with the calorimeter. Each titration consisted of 5 μ l injections of the self-complementary ds-(CGCGAATTCGCG) DNA into the 75 μ M protein-containing sample cell (~1.5 ml) at 25 °C with a mixing speed of 220 r.p.m. Heats of dilution were determined by titrating the same ds-(CGCGAATTCGCG) DNA into the dialysis buffer or into the buffer containing the protein. The data were then integrated to generate curves in which the areas under the injection peaks were plotted against the concentration ratio of DNA to protein. Analysis of the data was performed using MicroCal ITC Origin software and isotherm were fitted with the sequential binding models.

References

- Moharikar, S., D'Souza, J. S., Kulkarni, A. B. & Rao, B. J. Apoptotic-like cell death pathway is induced in unicellular chlorophyte *Chlamydomonas reinhardtii* cells following UV irradiation. *J. Phycolgy* **42**, 423–433 (2006).
- Moharikar, S., D'Souza, J. S. & Rao, B. J. A homologue of the defender against the apoptotic death gene (*dad1*) in UV-exposed *Chlamydomonas* cells is downregulated with the onset of programmed cell death. *J. Biosci.* **32**, 261–270 (2007).
- Kim, S. H. *et al.* Identification and expression of *uvi31+*, a UV-inducible gene from *Schizosaccharomyces pombe*. *Environ. Mol. Mutagen.* **30**, 72–81 (1997).
- Lee, J. K. *et al.* Isolation of UV-inducible transcripts from *Schizosaccharomyces pombe*. *Biochem Biophys Res Commun.* **202**, 1113–1119 (1994).
- Kim, M. J., Kim, H. S., Lee, J. K., Lee, C. B. & Park, S. D. Regulation of septation and cytokinesis during resumption of cell division requires *uvi31+*, a UV-inducible gene of fission yeast. *Mol. Cells* **14**, 425–430 (2002).
- Shukla, M. *et al.* UVI31+ is a DNA endonuclease that dynamically localizes to chloroplast pyrenoids in *C. reinhardtii*. *PLoS. One.* **7**, e51913 (2012).
- Vlcek, D., Sevcovicova, A., Sviezna, B., Galova, E. & Miadokova, E. *Chlamydomonas reinhardtii*: a convenient model system for the study of DNA repair in photoautotrophic eukaryotes. *Curr Genet.* **53**, 1–22 (2008).
- Setlow, R. B., Swenson, P. A. & Carrier, W. L. Thymine dimers and inhibition of DNA synthesis by ultraviolet irradiation of cells. *Science* **142**, 1464–1466 (1963).
- Chaudhari, V., Raghavan, V. & Rao, B. J. Preparation of Efficient Excision Repair Competent Cell-Free Extracts from *C. reinhardtii* Cells. *PLoS One.* **9**, e109160 (2014).
- Kozlowski, L. P. & Bujnicki, J. M. Metadisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC. Bioinformatics* **13**, 111 (2012).
- Kasai, T. *et al.* Solution structure of a BolA-like protein from *Mus musculus*. *Protein Sci.* **13**, 545–548 (2004).
- Guntert, P. *et al.* Structure determination of the Antp (C39–S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 531–540 (1991).
- Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
- Brunger, A. T. *et al.* Crystallography and NMR systems: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **1**(54), 905–921 (1988).
- Jung, J. W., Yee, A., Wu, B., Arrowsmith, C. H. & Lee, W. Solution structure of YKR049C, a putative redox protein from *Saccharomyces cerevisiae*. *J. Biochem. Mol. Biol.* **38**, 550–554 (2005).
- Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **15**(50), 496–506 (2003).
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
- Santos, J. M., Freire, P., Vicente, M. & Arraiano, C. M. The stationary-phase morphogene *bolA* from *Escherichia coli* is induced by stress during early stages of growth. *Mol. Microbiol.* **32**, 789–798 (1999).
- Zhou, Y. B. *et al.* hBolA, novel non-classical secreted proteins, belonging to different BolA family with functional divergence. *Mol. Cell Biochem.* **317**, 61–68 (2008).
- Freire, P., Moreira, R. N. & Arraiano, C. M. BolA inhibits cell elongation and regulates MreB expression levels. *J. Mol. Biol.* **385**, 1345–1351 (2009).
- Morton, G., Singh, J. & Hadi, S. Contribution of *rpoS* and *bolA* genes in biofilm formation in *Escherichia coli* K-12 MG1655. *J. Mol. Biol.* **342**, 207–213 (2010).
- Aldea, M., Hernandez-Chico, C., Campa, A. G., Kushner, S. R. & Vicente, M. Identification, cloning, and expression of *bolA*, an *ftsZ*-dependent morphogene of *Escherichia coli*. *J. Bacteriol.* **170**, 5169–5176 (1988).
- Alda, M., Garrido, T., Pla, T. & Vicente, M. Division genes in *Escherichia coli* are expressed coordinately to cell septum requirements by gearbox promoters. *J. Bacteriol. EMBO J.* **9**, 3787–3794 (1990).
- Huynen, M. A., Spronk, C. A., Gabaldon, T. & Snel, B. Combining data from genomes, Y2H and 3D structure indicates that BolA is a reductase interacting with a glutaredoxin. *FEBS Lett.* **579**, 591–596 (2005).
- Buchko, G. W. *et al.* Solution-state NMR structure of the putative morphogene protein BolA (PFE0790c) from *Plasmodium falciparum*. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 514–521 (2015).
- Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014).
- Valverde, R., Pozdnyakova, I., Kajander, T., Venkatraman, J. & Regan, L. Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein. *Structure* **15**, 1090–1098 (2007).
- Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS J.* **275**, 2712–2726 (2008).
- Grishin, N. V. KH domain: one motif, two folds. *Nucleic Acids Res.* **29**, 638–643 (2001).
- Bomsztyk, K., Denisenko, O. & Ostrowski, J. hnRNP K: one protein multiple processes. *Bioessays* **26**, 629–638 (2004).
- Brykailo, M. A., Corbett, A. H. & Fridovich-Keil, J. L. Functional overlap between conserved and diverged KH domains in *Saccharomyces cerevisiae* SCP160. *Nucleic Acids Res.* **35**, 1108–1118 (2007).
- Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21**, 1193–1198 (1993).
- Braddock, D. T., Louis, J. M., Baber, J. L., Levens, L. & Clore, G. M. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**, 1051–1056 (2002).
- Diaz-Moreno, I. *et al.* Orientation of the central domains of KSRP and its implications for the interaction with the RNA targets. *Nucleic Acids Res.* **38**(15), 5193–5205 (2010).
- Braddock, D. T., Baber, J. L., Levens, L. & Clore, G. M. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–3485 (2002).
- Pingoud, A., Wilson, G. G. & Wende, W. Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* **42**, 7489–7527 (2014).

37. Hsia, K. C., Li, C. L. & Yuan, H. S. Structural and functional insight into sugar-nonspecific nucleases in host defense. *Curr. Opin. Struct. Biol.* **15**, 126–14 (2005).
38. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Novel KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**(3), 323–332 (2000).
39. Liu, Z. *et al.* Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**(5544), 1098–10102 (2001).
40. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M. & Cusack, S. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**(7), 1055–1067 (2005).
41. Palmer, A. G., Rane, M. & Wright, P. E. Intramolecular motions of a zinc finger DNA-binding domain from xfin characterized by proton-detected natural abundance ¹³C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371–4380 (1991).
42. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy. *J. Magn. Reson.* **93**, 151–170 (1991).
43. Chakraborty, S., Minda, R., Salaye, L., Bhattacharjee, S. K. & Rao, B. J. Active site detection by spatial conformity and electrostatic analysis-unravelling a proteolytic function in shrimp alkaline phosphatase. *PLoS. One.* **6**, e28470 (2011).
44. Smith, R. M., Josephsen, J. & Szczelkun, M. D. An Mrr-family nuclease motif in the single polypeptide restriction–modification enzyme LlaGI. *Nucleic Acids Res.* **37**, 7231–7238 (2009).
45. Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. & Suzuki, M. A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* **17**, 5484–5496 (1998).
46. Pabo, C. O. & Sauer, R. T. Protein–DNA recognition. *Annu. Rev. Biochem.* **53**, 293–321 (1984).
47. Tateno, M. *et al.* DNA recognition by beta-sheets. *Biopolymers* **44**, 335–359 (1997).
48. Liao, D. I. *et al.* Structure of the IIA domain of the glucose permease of *Bacillus subtilis* at 2.2 Å resolution. *Biochemistry* **30**, 9583–9594 (1991).
49. Gutteridge, A. & Thornton, J. M. Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30**, 622–629 (2005).
50. Paspaleva, K. *et al.* Crystal structure of the DNA repair enzyme ultraviolet damage endonuclease. *Structure* **15**, 1316–1324 (2007).
51. Gallardo, I. C., Conte, R. D., Campoy, A. V., Maurino, S. M. G. & Moreno, I. D. A non-invasive NMR method based on Histidine imidazoles to analyze the pH-modulation of protein–nucleic acid interfaces. *Chemistry* **21**(20), 7588–7595 (2015).
52. Kahyaoglu, A. & Jordan, F. Direct proton magnetic resonance determination of the pKa of the active center histidine in thioSubtilisin. *Protein Sci.* **11**(4), 965–973 (2002).
53. Gallardo, I. C., Aroca, A., Persson, C., Karlsson, B. G. & Moreno, I. D. RNA binding of T-cell intracellular antigen-1 (TIA-1) C-terminal RNA recognition motif is modified by pH conditions. *Journal of Biological Chemistry* **288**, 25986–26994 (2013).
54. Giralt, E., Pons, M. & Andreu, D. Use of histidine pKa changes to study peptide DNA interactions. *Bioorganic Chemistry* **13**(3), 171–178 (1985).
55. Cheng, A. C., Chen, W. W., Fuhrmann, C. N. & Frankel, A. D. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol.* **327**, 781–796 (2003).
56. Kim, H., Jeong, E., Lee, S. W. & Han, K. Computational analysis of hydrogen bonds in protein–RNA complexes for interaction patterns. *FEBS Letters* **552**, 231–239 (2003).
57. Kondo, J. & Westhof, E. Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide–protein complexes. *Nucleic Acids Res.* **39**, 8628–8637 (2011).
58. Rangarajan, E. S. & Shankar, V. Sugar non-specific endonucleases. *FEMS Microbiology Reviews* **25**, 583–613 (2001).
59. Rout, A. K. *et al.* Sequence specific ¹H, ¹³C and ¹⁵N backbone resonance assignments of UVI31+ from *Chlamydomonas reinhardtii*. *Biomol. NMR Assign.* **4**, 171–174 (2010).
60. Singh, H., Raghavan, V., Shukla, M., Rao, B. J. & Chary, K. V. ¹H, ¹³C and ¹⁵N resonance assignments of S114A mutant of UVI31+ from *Chlamydomonas reinhardtii*. *Biomol. NMR Assign.* **8**, 71–74 (2014).
61. Bax, A., Ikura, M., Kay, L. E., Barbato, G. & Spera, S. Multidimensional triple resonance NMR spectroscopy of isotopically uniformly enriched proteins: a powerful new strategy for structure determination. *Ciba. Found. Symp.* **161**, 108–119 (1991).
62. Bax, A. & Grzesiek, S. Methodological advances in protein NMR. *Acc. Chem. Res.* **22**, 131–138 (1993).
63. Atreya, H. S., Chary, K. V. & Govil, G. Automated NMR assignments of proteins for high throughput structure determination: TATAPRO II. *Current Science* **83**(11), 1372–1376 (2000).
64. Keller, R. L. J. Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. Ph. D. thesis, ETH Zurich, Switzerland, *Thèse de doctorat, ETH Zurich Thesis No. 15947, Switzerland* 1–149 (2004).
65. Wishart, D. S. *et al.* ¹H ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).
66. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
67. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
68. Delano, W. L. The PyMOL molecular graphics system. San Carlos, CA (2002).
69. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).
70. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
71. Schanda, P. & Brutscher, B. Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR. *J. Magn. Reson.* **178**, 334–339 (2006).
72. Judith, B., Shlomo, E. & Bik-Kwoon, T. An agarose gel electrophoresis assay for the detection of DNA-binding activities in yeast cell extracts. *Methods in Enzymology* **155**, 528–537 (1987).
73. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**(7), 671–675 (2012).

Acknowledgements

The facilities provided by National Facility for High-Field NMR at Tata Institute of Fundamental Research, Mumbai, supported by Department of Science and Technology, New Delhi, Department of Biotechnology, New Delhi, Council of Scientific and Industrial Research, New Delhi and Tata Institute of Fundamental Research, Mumbai are acknowledged. KVRC and BJR thank Department of Science and Technology, Government of India, for their Sir JC Bose National Fellowships.

Author Contributions

A.K.R. and H.S. conducted experiments, analyzed the results, and wrote the paper. S.P. for computational modeling and discussions. V.R. and R.M. conducted the endonuclease activity assays and provided the plasmid. H.S. and S.G. conducted the EMSA. B.J.R. and K.V.R.C. for designing the project and critical discussions.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-31142-w>.

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018