Novel molecular aspects of the CRISPR backbone protein ‘Cas7’ from cyanobacteria

Prakash Kalwani, Devashish Rath* and Anand Ballal*

1Molecular Biology Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400085. 2Homi Bhabha National Institute, Anushaktinagar, Mumbai 400094

*Address correspondence to:

Dr. Devashish Rath, Ph: +91-22-25590796; email: devrath@barc.gov.in

Dr. Anand Ballal, Ph: +91-22-25593290; e-mail: aballal@barc.gov.in; ORCID: 0000-0002-8776-3021

Keywords

Cas7, Cyanobacteria, Anabaena, CRISPR type-ID, RNA-binding protein, oligomerization
Abstract

The cyanobacterium *Anabaena* PCC 7120 shows the presence of Type I-D CRISPR system that can potentially confer adaptive immunity. The Cas7 protein (Alr1562), which forms the backbone of the Type I-D surveillance complex, was characterized from *Anabaena*. Alr1562, showed the presence of the non-canonical RNA recognition motif and two intrinsically disordered regions (IDRs). When overexpressed in *E. coli*, the Alr1562 protein was soluble and could be purified by affinity chromatography, however, deletion of IDRs rendered Alr1562 completely insoluble. The purified Alr1562 was present in the dimeric or a RNA-associated higher oligomeric form, which appeared as spiral structures under electron microscope. With RNaseA and NaCl treatment, the higher oligomeric form converted to the lower oligomeric form, indicating that oligomerization occurred due to the association of Alr1562 with RNA. The secondary structure of both these forms was largely similar, resembling that of a partially folded protein. The dimeric Alr1562 was more prone to temperature-dependent aggregation than the higher oligomeric form. *In vitro*, the Alr1562 bound more specifically to a minimal CRISPR unit than to the non-specific RNA. Residues required for binding of Alr1562 to RNA, identified by protein modeling-based approaches, were mutated for functional validation. Interestingly, these mutant proteins, showing reduced ability to bind RNA were predominantly present in dimeric form. Alr1562 was detected with specific antiserum in *Anabaena*, suggesting that the Type I-D system is expressed and may be functional *in vivo*. This is the first report that describes the characterization of a Cas protein from any photosynthetic organism.
**Introduction**

Clustered regularly interspaced short palindromic repeats (CRISPR) along with CRISPR-associated proteins (Cas) are responsible for adaptive immunity against invading nucleic acids and are found in several bacterial and archaeabacterial species [1, 2]. The adaptive feature comes from CRISPR system’s ability to record memory of previous infection by integrating small pieces of invading DNA into the genome in the form of spacers between the repeats. This genetic memory is utilized to recognize and neutralize subsequent infections involving the same invading species. This immunity is executed in three steps: adaptation or spacer acquisition, expression of Cas proteins and CRISPR RNAs (crRNAs) and finally, interference with the invading DNA. The CRISPR locus, including the repeats and the spacers, is transcribed as pre-crRNA which is then processed into mature crRNA by Cas proteins [1]. Mature crRNAs combine with Cas proteins to form active ribonucleoprotein complexes that eliminate foreign nucleic acid [3].

CRISPR systems are classified in two major classes i.e. 1 and 2. Class 1 CRISPR systems are usually multi-subunit protein complexes consisting several Cas proteins (e.g. *E. coli* cascade), while class 2 CRISPR systems are less complex, where only one protein (such as Cas9 and Cpf1) along with crRNA forms the interference module [4]. Class 1 system is further subdivided into type I, type III and type IV systems while Class 2 includes type II and type V systems. Very recently, the two major classes of CRISPR systems have been classified into 6 types (type I–VI) and 33 subtypes [5].

Type I and type III interference complexes share several structural features and are characterized by the presence of some common (Cas6, Cas5, Cas7 and Cas8) proteins and some distinct proteins (Cas3 in type I and Cas10 in type III). The Cas7 protein, which form the backbone of the large interference complexes is central to the functioning of type I and type III CRISPR interference machineries [6]. The Cas7 protein contains non-canonical RNA recognition motifs (RRM)/ ferredoxin like domain that shows very less identity to other RNA binding domains. Cas7 multimerically attaches to the mature crRNA and nucleates formation of the surveillance complex [7]. Cas7 homologue of *E. coli* forms spiral structure with crRNA, acting as a porting sites for other Cas proteins which, ultimately define complex length, activity and target recognition [7, 8, 9]. Cas7 along with Cas5 is also responsible for stability and maintenance of the crRNA [10].
Cyanobacteria are morphologically and metabolically diverse group of bacteria that are present ubiquitously across the globe [11]. Cyanobacteria have a unique evolutionary status as they are the progenitors of chloroplasts, and for this reason they are believed to be a good model system to understand plant responses to abiotic stresses [12, 13, 14]. Some strains of cyanobacteria are diazotrophic, i.e. they can convert the atmospheric nitrogen directly into ammonia. In fact, strains of filamentous, heterocystous, nitrogen-fixing cyanobacteria such as Anabaena are economically important as they are widely used as natural eco-friendly biofertilizers in paddy fields of Southeast Asia [15].

Many cyanobacteria show the presence of CRISPR repeats and Cas genes, which largely belong to type I and type III CRISPR system [16, 17]. Among these too, the type I-D system is particularly abundant in cyanobacteria and some thermophiles [4, 16]. Notably, although several cyanobacteria have been genetically modified by employing exogenous CRISPR systems (e.g. Cas9 from S. pyogenes), the endogenous Cas proteins from these organisms have not been characterized at all [17]. The genome of filamentous, heterocystous, N₂-fixing cyanobacterium Anabaena PCC 7120 showed the presence of all proteins related to the type I-D CRISPR system. Moreover, the crRNAs associated with the type I-D system were expressed in vivo and processed into mature crRNAs in Anabaena [17]. Also, as described later in the text, using the specific antiserum, the Cas7 protein of this type I-D system, Alr1562 (encoded by alr1562), was detected in Anabaena extracts, indicating that this CRISPR system may indeed be functional in vivo. To gain insights into the structure-function of the hitherto unexplored cyanobacterial CRISPR proteins, Alr1562 the Cas7 homolog from Anabaena PCC 7120 was characterized in this study.

Materials and methods

Bioinformatic analysis

Anabaena genome was examined for CRISPR repeats and Cas proteins employing CRISPi database (https://crispi.genouest.org/). All the ORFs in vicinity of CRISPR locus 3 were subjected to CD search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and were given a putative annotation (Figure 1A). Disorder prediction for Alr1562 protein was performed using d2p2 server [18] (http://d2p2.pro/search), which provided disorder consensus for Arl1562 using different disorder prediction servers. Model of Alr1562 was generated through I-TASSER and
the best model with a low C-score was selected for future analysis [19]. Identification of residues involved in binding to RNA was based on conservation with other homologues, I-TASSER predictions and structural overlapping of the Alr1562 model (generated through I-TASSER) to the known *E. coli* cascade complex. Multiple sequence alignment to assess conservation was accomplished using Clustal omega and MUSCLE [20] whereas the structural overlap of the Alr1562 model and the *E. coli* cascade was performed using Chimera (1.13.1) [21].

**Organisms and growth conditions**

*E. coli* cells were cultured in Luria-Bertani (LB) medium containing appropriate antibiotics at 37°C with shaking (150 rpm). The antibiotics used were neomycin 50 µg ml⁻¹ and chloramphenicol 34 µg ml⁻¹. *Anabaena* PCC 7120 was cultured in BG-11 liquid medium (pH 7.0) with or without combined nitrogen (17 mM NaNO₃) under continuous illumination (30 µEm⁻²s⁻¹) as a still culture at 27±2°C.

**Overproduction, purification of Alr1562 proteins and Western blotting**

The *alr1562* ORF was PCR-amplified from the genomic DNA of *Anabaena* PCC 7120 using Alr1562_Fwd and Alr1562_Rev DNA primers (Supplementary Table 1). The amplified PCR product was digested with the restriction enzymes BamHI and Sall and subsequently ligated to a similarly digested pRSF-1b vector. The pRSF*alr1562* was transformed into *E. coli* NovaBlue cells. PCR-based amplification using suitable primers and *Anabaena* DNA was used to generate DNA fragments encoding the truncated Alr1562 proteins. These DNA products were also ligated to pRSF-1b as mentioned above. Site directed mutagenesis to obtain point mutants of Alr1562 was performed using KOD plus mutagenesis kit (from TOYOBO Co. Ltd.), employing the wildtype pRSF-*alr1562* as template. Six site directed mutants (R29A, F37A, R38A, R72A, K73A and Y119A) were generated by converting all the respective residues to alanine. All the above-mentioned plasmids were sequenced to confirm their nucleotide sequence and subsequently transformed into BL-21 DE3 (pLysS) cells for overexpression of the desired proteins.

All respective plasmids (Supplementary Table 2) were transformed into *E. coli* BL21 DE3 pLysS for overproduction of the his-tagged Alr1562. *E. coli* cells were initially grown in LB medium at 37°C (with shaking, 150 rpm) till the culture density reached OD₆₀₀ ~1.0. Subsequently, the culture was shifted to 20°C and after 30 min, IPTG (0.5 mM) was added.
Induction was carried out for 16 h at 20°C. Cells were pelleted by centrifugation (4000xg, 10 min). Cells were suspended in buffer A (50 mM Tris-Cl, 200 mM NaCl, 5 mM imidazole; pH 8.0), lysed by sonication (2s burst at 200 to 300 W with a 2s cooling period between each burst) and centrifuged at 15000xg to separate the inclusion bodies from the soluble fraction. Purification of Alr1562 from the soluble fraction was performed with Ni-NTA agarose (Sigma) in a batch process. After thoroughly washing the resin with buffer A, the desired protein was eluted by increasing the concentration of imidazole in buffer A to 50-250 mM. Alr1562 protein eluted between 50 mM to 250 mM imidazole, with maximal elution at 100 mM imidazole. RNaseA (2 units/10 ml homogenate) and NaCl (1.5 M) treatment for removal of RNA was performed during homogenization of E. coli cells prior to incubation with the Ni-NTA beads.

The purified Alr1562 or its variants were resolved on sodium dodecyl sulphate–polyacrylamide gels (12% or 15%) or on native polyacrylamide gels (8% acrylamide) and stained with Coomassie Brilliant Blue G-250 or ethidium bromide. The Alr1562 protein was also resolved on native polyacrylamide gels using standard 0.5X TBE buffer at 4°C. Alr1562 protein was also employed to raise specific antiserum in rabbits at a commercial facility (Genei, India)

The total cellular proteins from Anabaena cultures were extracted with Laemmli’s buffer [22], electrophoretically separated on denaturing polyacrylamide gel (15%), blotted on to a nitrocellulose membrane (Sigma), and probed with the Alr1562 antiserum (1:5000 dilution). ALP conjugated anti-rabbit IgG was used as the secondary antibody and blots were developed with the substrate NBT/BCIP (Roche). All Western blots were repeated at least thrice with consistent results.

RNA isolation

RNA was isolated using TRIZol™ LS reagent (Invitrogen) by following the manufacturer’s protocol. RNA associated with Alr1562 obtained from different imidazole elution fractions (i.e. 50 mM, 100 mM or 250 mM imidazole) was resolved on 1.4% denaturing formaldehyde gel. RNA was subjected to RNAseq analysis at a commercial facility (Agrigenome, India)

Electrophoretic mobility shift assay (EMSA)

Binding of Alr1562 to different RNA substrates was evaluated using γ-32P-labelled ssRNA oligos. Three different custom-made RNA substrates (Sigma) i.e. CRISPR repeat-spacer-repeat
(RSR) motif, only CRISPR repeat, and one non-specific 20 mer were employed. These substrates were end labelled with $\gamma$-32P-ATP employing polynucleotide kinase. Labelling was done using 10 pmol RNA/DNA oligo with 6.6 pmol $\gamma$P32 ATP and 40 units of polynucleotide kinase (NEB) in the buffer provided with the enzyme. This 10 µl reaction was incubated at 37°C for 45 minutes, 40 µl of RNase free water was added and the reaction contents were denatured at 65°C for 20 min. After heat denaturation, the unbound $\gamma$-P32 ATP was removed from the labelled oligos using Microspin™ G-50 Columns (Merck). Binding was carried out in reaction buffer (10 mM HEPES, 15 mM MgCl$_2$, 200 mM KCl, 10 mM DTT, 25% Glycerol, pH 7.0) and 200 nmol of substrate was used per reaction. Protein and RNA substrates were incubated for 30 min at room temperature, mixed with a tracking dye and separated on 8% TBE gel (80V for 1h). Gel was then vacuum dried and exposed to X-ray film for 2h to overnight. X-ray films were developed using developer and fixer solutions in a dark room. For competition experiments, 25X, 75X or 200X of different RNA oligos were added.

**Circular Dichroism (CD) spectroscopy and Dynamic Light Scattering**

For CD spectra analysis, nearly, 0.1 mg/ml concentration of protein was used as described earlier [23]. The purified protein was diluted in buffer containing 20 mM Tris (pH 7.6) and spectra were recorded using a CD spectrophotometer (Biologic spectrometer MOS500). CD spectra of different forms of protein was taken from 30°C to 80°C to analyze melting temperature (T$_m$). Hydrodynamic size of Alr1562 was determined by dynamic light scattering (DLS) using the Malvern Zetasizer nanoseries instrument. Individual protein fractions (in 20 mM Tris, pH 8), were analyzed in 0.2 ml or 1 ml cuvettes supplied with the instrument.

**Negative stain transmission electron microscopy (TEM)**

The protein (in 20 mM Tris, pH 7.6) was diluted to 1:10 ratio with 20 mM Tris, and 5 µl of sample was spotted on UV-activated carbon coated copper grids (200 mesh). After 5 min, grids were stained with 2.0 % aqueous uranyl acetate solution for 2 min. The excess stain was removed using a filter paper wick and the grid was allowed to dry. Samples were viewed on a Carl Zeiss Libra 120 TEM operating at an accelerating voltage of 120 kV.

**Gel filtration chromatography**
Gel filtration chromatography experiments were performed on Superdex 200 10/300 mm GL column (GE Healthcare, US) equilibrated with 20 mM Tris, pH 7.2 buffer containing 50 mM NaCl, at a flow rate of 0.5 ml/min at 25°C. The column was connected to a ÄKTA pure protein purification system - GE Healthcare Ltd. The column was calibrated with Sigma gel filtration standards viz. Cytochrome C (12.4 kDa), Carbonic anhydrase (29 kDa), Bovine serum albumin (66 kDa), Alcohol dehydrogenase (150 kDa), α-amylase (200 kDa) and Blue dextran (2000 kDa).

Results

Bioinformatic analysis of alr1562/Alr1562 from Anabaena PCC 7120

A comprehensive in silico search for CRISPR loci in Anabaena PCC 7120 (NC_003272) genome with CRISPRdisco (http://github.com/crisprlab/CRISPRdisco) showed the presence of 8 CRISPR loci. Of these, a complete putative Type-ID system was observed in CRISPR locus 3 (Figure 1A). The CRISPR RNA from this locus (CrRNA-3) formed the typical stem-loop structure, which is a characteristic of functional CrRNA (Supplementary Figure 1A). The CRISPR type I-D system is mostly associated with the cyano sub-type of CRISPR/Cas locus, usually found in cyanobacteria and other archaeal species [4]. The alr1562 ORF (1035-bp, encoding a protein of 344 amino acids), which was the second ORF of CRISPR type-ID system, was flanked by alr1561 (2694-bp, encoding a Cas10d protein of 897 amino acids) and alr1563 (759-bp, encoding a Cas5 protein of 252 amino acids) (Figure 1A). In the sequence database (http://genome.microbedb.jp/cyanobase/GCA_0009705.1genes/alr1562), Alr1562 is annotated as a protein with unknown function. BLAST search showed the Alr1562 protein to be virtually identical (93-94 % identity) to the putative Csc2/Cas7 protein present in filamentous cyanobacterial species such as Nostoc punctiforme and Nostoc sp. PCC 7107. Alr1562 was designated as Csc2 i.e. CRISPR/Cas Subtype Cyano protein 2, a repeat associated mysterious protein (RAMP), in the conserved domain database (CDD). Unlike its cyanobacterial counterparts, Alr1562 showed very poor sequence identity to the other Cas7 proteins that have been functionally characterized (Supplementary Figure 1B). Bioinformatic analysis for prediction of disordered regions for Alr1562 (http://d2p2.pro/search), which predicts a combined consensus of disorder in proteins from several online servers, showed Alr1562 to contain higher
degree of disorder at N-terminal (1-100 amino acids) and C-terminal ends (300-344 amino acids) (Figure 1B).

Heterologous over-expression, purification and in vivo detection of Alr1562 in Anabaena

In order to characterize Alr1562, the full-length protein or its variants were overexpressed and purified from E. coli. To express proteins with a N-terminal hexa-histidine tag, the complete alr1562 ORF or the three truncated versions of the alr1562 ORF, encoding the Alr1562 proteins that lacked the disordered regions (Figure 1B), were ligated into the E. coli expression vector pRSF-1b. The overproduction of all the respective proteins, in line with their expected mol. wt., was observed in E. coli (Supplementary Figure 2A, 2B, 2C, 2D). Interestingly, only the full-length Alr1562 was found in the soluble fraction (even at 37°C) (Figure 1C), whereas all the three truncated variants were exclusively present in the inclusion body fraction of E. coli. In fact, lowering the temperature of incubation (post-induction) to 20°C did not help, and these overproduced proteins continued to remain insoluble. The purified Alr1562 protein was employed to raise specific antiserum in rabbits which, was used to detect the production of the Alr1562 protein in Anabaena (Figure 1D). With Anabaena cell-free protein extracts, the expected-size signal corresponding to Alr1562 could be visualized on Western blots. The slightly lower band seen in the nitrogen-supplemented lane could be due to the processing of Alr1562 under these conditions in Anabaena. These results indicate that the Alr1562 protein is indeed expressed in Anabaena.

The purified Alr1562 protein exists in dimeric as well as higher oligomeric forms

The full-length Alr1562 protein could be purified to near homogeneity by affinity chromatography employing the Ni-NTA matrix (Figure 1C). When resolved on native PAGE, the 100/250 mM imidazole fractions showed the presence of higher oligomeric forms (seen close to the well) as well as a lower oligomeric form as shown in the (Figure 2A). Incidentally, when the 50 mM imidazole fraction was resolved on native PAGE, the lower oligomeric form of Alr1562 was mostly observed (Figure 2A). Gel filtration of Alr1562 (250 mM imidazole fraction) showed two distinct peaks (Figure 2B); peak I (PI) at around 8 ml retention volume and peak II (PII) at around 15 ml retention volume. The molecular weight of lower oligomeric form of Alr1562 present in the PII peak was found to be around ~90 kDa, whereas the PI peak was
>600 kDa (Supplementary Figure 3A). Thus, the lower oligomeric form is very likely to be the dimeric form of Alr1562.

Both PI and PII fractions showed the same-sized band (~40 kDa) in when resolved on SDS PAGE. However, on native PAGE, a distinct difference in their profile was observed (Figure 2C). The PI fraction showed the presence of the higher oligomeric form of Alr1562 whereas only the lower oligomeric form of the protein was present in the PII fraction. When the 50 mM elution fraction was subjected to gel filtration, the PII peak was dominant and only a small peak corresponding to PI was observed (Supplementary Figure 3B), indicating the prevalence of lower oligomeric form in this fraction. DLS profile of Alr1562 corresponding to peaks PI or PII showed hydrodynamic radii of 30 nm and 6 nm, respectively (Figure 2D). PI showed a broader peak, indicating the heterogenous nature of the higher oligomeric form. Circular dichroism profile of either the PI or the PII form was found to be similar (Figure 2E, 2F), showing the presence of 43% distorted alpha helix, 1.36% beta sheet and 46% random coils.

The higher oligomeric form of Alr1562 is associated with RNA

The 250 mM imidazole elution fraction of Alr1562 when resolved on native acrylamide gels showed a distinct ethidium bromide fluorescence at the top of the well, indicating the association of nucleic acids (from *E. coli*) along with the purified protein (Figure 3A). When treated with DNase I, this fluorescence did not disappear, but when treated with RNaseA, a distinct reduction in this fluorescence was observed, indicating that it was RNA that was associated with the Alr1562 protein (Figure 3B). Moreover, the $A_{260}/A_{280}$ ratio of PI and PII (Figure 2B) was ~1.3 and ~0.6 respectively, further corroborating that the association of RNA with the PI fraction.

RNA was isolated from 50 mM/100 mM/250 mM imidazole elution fractions that were obtained from the Ni-NTA matrix during purification of Alr1562 from *E. coli*. As shown in Figure 3C, RNA could be obtained only from 100 mM or 250 mM fraction that contained both forms i.e. the higher oligomeric as well as the lower oligomeric form of Alr1562. But, the 50 mM imidazole elution fraction, which mostly contained the lower oligomeric form, did not yield any significant amount RNA. These results indicate that the higher oligomeric form of Alr1562 is generally associated with RNA. RNAseq analysis found the RNA associated with the Alr1562 protein (250 mM imidazole elution) to be the abundantly expressed RNAs from *E. coli* and no sequence
specificity was observed. Most of reads were from Alr1562 itself, ribosomal RNA and other transcripts from other housekeeping genes of E. coli (Supplementary Table. 3).

**Higher oligomeric form of Alr1562 converts to lower oligomeric form on treatment with NaCl & RNaseA**

The 250 mM imidazole elution fraction showed the preponderance of higher oligomeric forms as compared to the lower oligomeric i.e. dimeric form (Figure 3D). In fact, the higher oligomeric forms of Alr1562 appeared as smear, which started right from the well of the stacking gel (Figure 3D). On treatment with 1.5 M NaCl and RNaseA, the overall content of the higher oligomeric form decreased, as evident from the reduction of smearing in the stacking gel and a distinct increase in the dimeric form was observed in the lower portion of the gel. In gel filtration analysis, compared to the untreated Alr1562 preparation, the content of the higher oligomeric form (i.e. PI) decreased and a concomitant increase in the lower monomeric form (PI) was observed in the NaCl & RNaseA-treated preparation. Also, with the above-mentioned treatment, a broadening of the PI peak was observed (Figure 3E). These results clearly indicate the conversion of higher oligomeric form to lower oligomeric forms on treatment with NaCl-RNaseA. Clearly, Alr1562 associated with the endogenous RNA of E. coli to form the higher oligomers. Hence, conditions that disrupt RNA-protein interaction (high salt/RNaseA), release the Alr1562 protein from RNA present in the higher oligomeric form, consequently leading to increase in the content of the dimeric form.

**Alr1562 aggregates on heat treatment, releasing the bound RNA**

The 50 mM imidazole elution fraction that mostly contains the dimeric form of Alr1562 was heated (40°-80°C) and subsequently resolved on native gels. Compared to the protein incubated at 40°C, the content of the dimeric Alr1562 band reduced considerably at 50°C and the protein was observed at the top of the lane, indicating aggregation. The size of the aggregates increased on further increasing the temperature, and the Alr1562 protein was observed only in the stacking gel at temperature above 70°C (Figure 4A). The CD profile of the Alr1562 protein present in this fraction showed temperature-dependent loss of secondary structure, and temperature over 70°C, it was completely denatured (Figure 4B).
The higher oligomeric form (PI peak, Figure 2B) did not show aggregation up to 60°C (Figure 4C). In the corresponding CD profile too, there was no apparent loss in the secondary structure till 60°C (Figure 4D). In 40°-60°C-treated samples, RNA was observed in the upper half of the gel, associated with the Alr1562 protein. However, at higher temperatures (i.e. 70°C and above), most of the protein was observed in the stacking gel, demonstrating aggregation. Likewise, CD analysis showed the protein to have completely lost its secondary structure at higher temperatures. Interestingly, in contrast to the 40°-60°C-heated samples, on incubation at 70°C and 80°C, the majority of RNA was observed in the lower half of the gel, indicating the inability of the aggregated Alr1562 protein to bind to RNA (Figure 4E).

**The Alr1562 protein binds to ssRNA in vitro**

The CRISPR RNA (crRNA) consists of two repeats separated by a spacer. These repeats in crRNA form specific secondary structures (Supplementary Figure 1A), which facilitates binding of the Cas7 protein. Ability of Alr1562 to bind to CRISPR repeat-spacer-repeat (RSR) RNA motif from *Anabaena* was assessed by electrophoretic mobility shift assay (EMSA). For competition, along with the unlabeled RSR, two different RNA substrates viz. *Anabaena* CRISPR repeat (R2) and a 20 nucleotide RNA unrelated to *Anabaena* CRISPR (20-mer) were employed (Figure 5A). The dimeric form i.e. PII peak obtained during gel filtration, which is not associated with RNA, was employed for the assays.

The Alr1562 bound to the $^{32}$P-RSR RNA in a protein concentration-dependent manner (Figure 5B). Competition assays showed the cold RSR RNA, inhibit binding of Alr1562 to the $^{32}$P-RSR RNA very efficiently. Reduction in binding was also observed in the presence of cold R2 RNA, but it was not as efficient as that observed with the RSR RNA. Interestingly, the cold 20-mer RNA could not effectively compete with the $^{32}$P-RSR and a considerable amount of $^{32}$P-RSR-Alr1562 complexes could be observed even with 75X molar excess of cold 20-mer RNA. EMSA assays indicate that Alr1562 interacts most efficiently with RSR RNA, followed by R2 RNA and 20-mer RNA in that order (Figure 5C).

Electron microscopy of Alr1562 (250 mM imidazole fraction) showed several spiral structures, ranging from 30-120 nm (Figure 5D). Along with these, many indistinct structures were also viewed in the background. However, only the indistinct structures were seen when the dimeric (i.e. RNA-free) form of Alr1562 was observed under the electron microscope (Figure 5E). As
the higher oligomeric forms of Alr1562 are associated with RNA, the spiral structures are likely to be the RNA-Alr1562 complexes.

**Structural modelling of Alr1562 to identify residues involved in binding to RNA**

The closest homolog of Alr1562 with a known structure was Cas7 from *Thermofilum pendens* (PDB ID: 4TXD). This homolog showed 30% similarity with a query coverage of 0.71 with Alr1562. However, the structure obtained with Swiss-Model (https://swissmodel.expasy.org/) employing 4TXD was of low quality, and hence a fold-recognition approach for modeling of Alr1562 was employed. I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER) provided a qualitatively better model of Alr1562 (Figure 6A) with detailed secondary structure formations, especially in the C-terminal region, which was lacking in the structure generated with Swiss-Model. I-TASSER predicted several ligand binding sites which showed the residues F37, R38, T39, T70, R72, K73, V75, A76, R79, and Y119 to be involved in binding to nucleic acids.

Multiple sequence alignment of Cas7 homologs carried out with two different programs (MUSCLE and Clustal omega) showed a majority (6 out of 10) of these residues to be conserved. As type I surveillance complexes are known to have conserved structures [16, Supplementary Figure 5] the predicted structure of Alr1562 was superimposed on the Cas7 of the RNA bound Cascade complex structure from *E. coli* (PDB ID: 4u7u). F37 and R38 residues were not only position-wise conserved in *E. coli* Cas7, but they also interacted with the crRNA. Most of the other residues (T39, S42, R72, K73, V75 and A76) were spatially very close to the crRNA suggesting their involvement in binding to the RNA (Figure 6B).

**Alr1562 point mutants are prevalently present in lower oligomeric forms and display reduced capacity to bind RNA**

To experimentally verify the importance of the above-mentioned amino acids in the actual binding to RNA, 6 of these (R29, F37, R38, R72, K73 and Y119) were converted to alanine by site directed mutagenesis and the corresponding proteins purified after over-expression in *E. coli*. Interestingly, unlike the wild-type Alr1562 that co-purified with lot of *E. coli*-RNA, all these mutant proteins showed considerably reduced amount of associated RNA. OD\textsubscript{260/280} ratio of these...
ranged from 0.62 to 0.69, whereas that of the wild-type Alr1562 was 1.36 (Supplementary Table 4).

Gel filtration analysis showed all the mutant proteins to have a substantially reduced PI peak but the PII peak was considerably elevated, which was the reverse of that observed with the wild-type Alr1562 (Figure 6C). On native acrylamide gels too, the lower forms of the mutant Alr1562 proteins predominated. The CD spectrum of the Alr1562 point mutants was mostly similar to that of the wild-type protein (Supplementary Figure 4), indicating that the mutant proteins were indeed properly folded. EMSA assays showed the mutant Alr1562 proteins to have a substantially decreased ability to bind to the RSR substrate. Only 26-52 µM of the wild-type Alr1562 was enough to saturate the RSR, whereas in contrast, even 176 µM of any of the mutant proteins was unable to completely shift RNA employed in the reaction (Figure 6D).

Discussion:
The CRISPR systems are adaptive immune systems of prokaryotes that memorize earlier assaults of foreign genetic material and use this stored information to fend off attacks from similar elements in the future [24]. This process is brought about by the coordinated action of several CRISPR-associated i.e. ‘Cas’ proteins. In the CRISPR-mediated interference process, the leader sequence activates transcription of several repeat-spacer-repeat (RSR) units i.e. the pre-crRNA, which is subsequently processed by the Cas6 endonuclease to yield the mature crRNA [25]. Different Cas proteins recognize this mature crRNA and form the interference complex around it. The main backbone protein of this complex is Cas7, which forms a helical assembly around the crRNA. Cas7 associates with Cas8, and Cas5 to cover the 5′-terminal of the crRNA. Small subunits coat this helical backbone to form the CASCADE complex, which then enables binding of crRNA to the invading DNA [26]. Ultimately, the target destruction is brought about by Cas3 (nuclease/helicase) protein [2]. Thus, the Cas7 protein plays a key role in this process and is therefore vital for interference.

The Cas7 proteins from different bacterial species show very poor sequence homology to each other. Due to this immense diversity, it is likely that these proteins have evolved different properties and hence, biochemical/biophysical analysis of these is indispensable to understand their function or mechanism of action. The genome of filamentous, heterocystous, N₂-fixing cyanobacterium Anabaena PCC 7120 showed the presence of all proteins related to the type ID
CRISPR system [16]. Moreover, the crRNAs associated with the type ID system were expressed \textit{in vivo} and processed into mature crRNAs in \textit{Anabaena} [17]. Also, using the specific antiserum, the Cas7 protein of this type I-D system, Alr1562, was detected in \textit{Anabaena} extracts (Figure 1D), indicating that this CRISPR system may indeed be functional \textit{in vivo}.

Unlike completely structured proteins, many proteins have regions of intrinsic disorder, which are important for their function [27]. The presence of intrinsically disordered regions (IDRs) may affect the stability, solubility or its ability to crystallize [28]. Experiments with \textit{E. coli} or insect cell-free extracts have shown that higher content of intrinsic disorder in proteins is associated with the increased solubility of the expressed protein [29, 30]. However, in another study, the presence of low complexity IDRs was recommended to be minimised to aid solubility of proteins in \textit{E. coli} [31]. The full-length receptor protein, Cqm1, which has IDR at both its ends (akin to Alr1562), was completely insoluble when expressed in \textit{E. coli}. But deletion of these IDRs rendered the truncated Cqm1 protein soluble in this expression system [32]. In the case of Alr1562, deletion of either of the N-terminal or the C-terminal IDR caused this protein to be insoluble (Figure 1A), indicating the requirement of these regions for expression of the soluble Alr1562 in \textit{E. coli}.

Many CRISPR-associated RNA binding proteins, including Cas7, belong to the newly identified class of repeat-associated mysterious proteins (RAMPs). The RAMPs are characterized by the presence of two RRMs (RNA recognition motifs) of $\beta$-$\alpha$-$\beta$ type. Like the other RAMP proteins, Alr1562 also showed the presence of $\beta$-$\alpha$-$\beta$-$\beta$-$\alpha$-$\beta$ topology, which are usually associated with binding to nucleic acids. The typical $\beta_1$-$\alpha_1$-$\beta_2$-$\beta_3$-$\alpha_2$-$\beta_4$ topology of Alr1562 RRM was interrupted by hairpins and loops. Notably, the central domain that contains $\beta$-sheets were similar in all known homologues of Cas7 proteins (Supplementary Figure 5) even when there was very little sequence similarity between them. Although known to bind RNA using conventional modes of protein-RNA interaction, RAMPs are capable of binding to a wide variety of RNA motifs [33].

Alr1562 bound to a nonspacer-repeat (i.e. nonspecific, NS) RNA (Figure 5C), indicating that it had an inherent ability to bind to any RNA molecule. Similarly, the Cas7 proteins from \textit{Solfolobus solfataricus} and \textit{Methanopyrus kandleri}, have also been shown to bind to RNA nonspecifically [6, 7]. However, among the different RNA tested, Alr1562 bound the best to
RSR (repeat-spacer-repeat), revealing an increased specificity towards this substrate (Figure 5C). When purified from *E. coli*, Alr1562 was found to be associated with the endogenous *E. coli*-RNA. Incidentally, when identified, these RNAs molecules turned out to be the most abundantly expressed RNAs in *E. coli*, again denoting ability of the RAMPs such as Alr1562 to bind to non-specific RNAs.

The Alr1562 protein oligomerized in a helical manner along RNA (Figure 5D). Similarly, Cas7 proteins from *E. coli* or *S. solfataricus* [6] also formed a helical arrangement on RNA. This kind of organization appears to help the other proteins to associate with Cas7 in building up the Cascade complex. The *S. solfataricus* or *M. kandleri* Cas7 proteins were shown to be monomeric in solution and required other Cas proteins and crRNA for helical assembly [7]. The Alr1562 protein was observed to be dimeric and formed higher oligomers/helical assembly only in the presence of RNA. This was best exemplified by the Alr1562 mutants which were largely present in their dimeric form as they showed substantially reduced binding to RNA. This also suggests that ability of Alr1562 to dimerize is independent of its capability to bind RNA.

2D topology of I-TASSER model of Alr1562 provided further structural insights for comparisons with the other Cas7 homologs. Mutation of the R-21 residue of the *M. kandleri* Cas7 led to complete loss of binding activity [7]. Mutation of the corresponding residue of Alr1562, R-29, also led to reduced binding to RNA, indicating the importance of this residue in Cas7 function. Residues F37, R38, R72 and K73 were identified to be important for binding of Alr1562 to RNA in this study. As these four residues were not reported in previous studies, it would be interesting to see if the corresponding residues in the other Cas7 homologs are similarly involved in binding to RNA.

In conclusion, the Alr1562 protein showed the ability to exist in dimeric or higher oligomeric form when purified from *E. coli*. Detailed analysis showed that oligomerization occurred due to the association of Alr1562 dimers with endogenous RNA of *E. coli*. Interestingly, Alr1562 bound more specifically to repeat-spacer-repeat (RSR) RNA than to the non-specific RNA. Using a combination of bioinformatic approaches, residues required for binding of Alr1562 to RNA were identified and subsequently mutated by site directed mutagenesis. These mutated proteins not only showed the prevalence of the dimeric form, but also displayed considerably reduced binding to RNA, demonstrating that dimer formation was independent of binding to
RNA. Moreover, Alr1562 was expressed in *Anabaena*, suggesting that the CRISPR Type-ID system may be indeed functional *in vivo*. To our knowledge, this is the first report that describes the characterization of a Cas protein from any photosynthetic organism.

**Acknowledgements**

We would like to thank Dr. H.S. Misra, Head Molecular Biology Division for his constant support and encouragement. Help rendered by Dr. Alka Gupta, Dr. Manisha Banerjee, Mr. Dhiman Chakravarty and Ms. Namrata Waghamare during the course of this work is gratefully acknowledged.

**Author contribution**

AB and DR designed the overall study. AB, PK & DR planned the research; PK performed the experimental work; AB, PK & DR, interpreted the results. AB, PK & DR wrote the paper.

**Financial sources**

Work was funded by Department of Atomic Energy, India.

**References**

1. Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochemie 117, 119-128.

2. Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S., van der Oost, J., … Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature, 477, 486–489. doi:10.1038/nature10402

3. Jackson, R. N., Golden, S. M., van Erp, P. B., Carter, J., Westra, E. R., Brouns, S. J., van der Oost, J., Terwilliger, T. C., Read, R. J., … Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. Science (New York, N.Y.), 345(6203), 1473-9.

4. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., &Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews. Microbiology, 13, 722–736.
5. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S., … Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nature reviews. Microbiology, 18, 67–83. doi:10.1038/s41579-019-0299-x
6. Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., … Lawrence, C. M. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). Journal of Biological Chemistry, 286, 21643-56.
7. Hrle, A., Su, A. A., Ebert, J., Benda, C., Rendau, L., & Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. RNA Biology, 10, 1670-8.
8. Manica, A., & Schleper, C. (2013). CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus Sulfolobus. RNA Biology, 10, 671–678. doi:10.4161/rna.24154
9. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H., & Fernández, I. S. (2020). Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. Nature, 577, 271–274. doi:10.1038/s41586-019-1849-0
10. Lisa-Katharina Maier, Aris-Edda Stachler, Jutta Brendel, Britta Stoll, Susan Fischer, Karina A. Haas, Thandi S. Schwarz, Omer S. Alkhnbashi, Kundan Sharma, Henning Urlaub, Rolf Backofen, Uri Gophna & Anita Marchfelder (2019) The nuts and bolts of the Haloferax CRISPR-Cas system I-B, RNA Biology, 16, 469-480, doi: 10.1080/15476286.2018.1460994
11. Schirrmeister, B. E., Antonelli, A., & Bagheri, H. C. (2011). The origin of multicellularity in cyanobacteria. BMC Evolutionary Biology, 11, 45. doi:10.1186/1471-2148-11-45.
12. Raven, J. A., & Allen, J. F. (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biology, 4, 209. doi:10.1186/gb-2003-4-3-209.
13. Murata, N., & Wada, H. (1995). Acyl-lipid desaturases and their importance in the tolerance and acclimatization of cold cyanobacteria. Biochemical Journal, 308, 1–8.
14. Chakravarty, D., Banerjee, M., Bhani, S. C., & Ballal, A. (2016). A Salt-Inducible Mn-Catalase (KatB) Protects Cyanobacterium from Oxidative Stress. Plant Physiology, 170, 761–773. doi:10.1104/pp.15.01632
15. Qiu, Y. L., & Yu, J. (2003). Azolla - a model organism for plant genomic studies. Genomics, Proteomics & Bioinformatics, 1, 15–25. doi:10.1016/s1672-0229(03)01004-0
16. Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. Methods in Molecular Biology (Clifton, N.J.), 1311, 47–75. doi:10.1007/978-1-4939-2687-9_4
17. Hou, S., Brenes-Álvarez, M., Reimann, V., Alkhnbashi, O. S., Backofen, R., Muro-Pastor, A. M., & Hess, W. R. (2019). CRISPR-Cas systems in multicellular cyanobacteria. RNA Biology, 16, 518–529. doi:10.1080/15476286.2018.1493330
18. Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztányi, S., Uversky, V.N., Obradovic, Z., Kurgan, L., Dunker, A.K., Gough, J. (2013). "D2P2: Database of Disordered Protein Predictions." Nucleic Acids Research 41(D1), D508-D516. DOI:10.1093/nar/gks1226
19. Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Research, 43(W1), W174–W181. doi:10.1093/nar/gkv342
20. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research, 47(W1), W636-W641. doi:10.1093/nar/gkz268. PMID: 30976793; PMCID: PMC6602479.
21. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25, 1605-12.
22. Laemmli, U. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227, 680–685.
23. Chakravarty, D., Bihani, S.C., Banerjee, M., & Ballal, A. (2019). Novel molecular insights into the anti-oxidative stress response and structure–function of a salt-inducible cyanobacterial Mn-catalase. Plant Cell & Environment 42, 2508–2521.
24. Wiedenheft B., Sternberg S.H., and Doudna J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature, 482, 331-338.
25. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., … van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321, 960-964. doi:10.1126/science.1159689
26. Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., & Doudna, J. A. (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. Molecular Cell, 63, 840–851. doi:10.1016/j.molcel.2016.07.027

27. Varadi, M., Zsolyomi, F., Guharoy, M., & Tompa, P. (2015). Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. PloS One, 10, e0139731. doi:10.1371/journal.pone.0139731

28. Oldfield, C. J., Xue, B., Van, Y. Y., Ulrich, E. L., Markley, J. L., Dunker, A. K., & Uversky, V. N. (2013). Utilization of protein intrinsic disorder knowledge in structural proteomics. Biochimica et Biophysica Acta, 1834, 487-498. doi:10.1016/j.bbapap.2012.12.003

29. Tokmakov, A. A., Kurotani, A., Ikeda, M., Terazawa, Y., Shirouzu, M., Stefanov, V., … Yokoyama, S. (2015). Content of intrinsic disorder influences the outcome of cell-free protein synthesis. Scientific Reports, 5, 14079. doi:10.1038/srep14079

30. Kurotani, A., Takagi, T., Toyama, M., Shirouzu, M., Yokoyama, S., Fukami, Y., & Tokmakov, A.A. (2010). Comprehensive bioinformatics analysis of cell-free protein synthesis: identification of multiple protein properties that correlate with successful expression. FASEB Journal, 24, 1095-1104. doi: 10.1096/fj.09-139527.

31. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., & McCafferty, J. (2004). Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnology, 4, 32. doi:10.1186/1472-6750-4-32.

32. Sharma, M., Lakshmi, A., Gupta, G. D., & Kumar, V., (2018). Mosquito-larvicidal binary toxin receptor protein (Cqm1): Crystallization and X-ray crystallographic analysis. Acta Crystallography. F74, 571-577v. doi:10.1107/S2053230X18010671

33. Wang, R., & Li, H. (2012). The mysterious RAMP proteins and their roles in small RNA-based immunity. Protein Science, 21, 463-470. doi:10.1002/pro.2044

**Figure Legends**

**Figure 1 Bioinformatic analysis and purification of the full-length Alr1562.** (A) Schematic representation of the cas genes present within the Type-ID CRISPR locus 3 of *Anabaena* PCC 7120. Arrowheads indicate the direction of transcription. The respective Cas-protein encoded by individual ORFs are also mentioned below each arrowhead. (B) Schematic representation of Alr1562 protein. The consensus disordered regions identified within the Alr1562 protein (using
d2p2 server) are shown in green. The proteins that were over-expressed in *E. coli* i.e. the full-length Alr1562 (Alr1562), Alr1562 lacking the N-terminal disordered region (1562_N100), Alr1562 lacking the C-terminal disordered region (1562_C44) or Alr1562 lacking both the terminally located disordered regions (1562_DT) are also schematically depicted. The deleted regions of the protein are indicated by dashed line. (C) Purification of the Alr1562 protein. The his-tagged Alr1562 protein was over-expressed in *E. coli* and purified by Ni-NTA affinity chromatography. Clarified cell lysate (Lys), flow through (FT), wash (W) and elution fractions obtained with different concentrations of imidazole (as indicated in the figure) were resolved on denaturing polyacrylamide gels and stained with Coomassie brilliant blue (CBB). M, molecular mass marker. (D) Cell-free total protein extracts obtained from *Anabaena* cultured under nitrogen supplemented (N⁺) or nitrogen deprived conditions (N⁻) were subjected to denaturing PAGE, transferred onto nitrocellulose membrane and probed with the Alr1562 polyclonal antiserum. The purified Alr1562 protein (obtained from C) is indicated by arrow.

**Figure 2. Biophysical properties of Alr1562.** (A) Native PAGE profile showing the oligomeric forms of Alr1562 protein obtained on elution with increasing concentrations of imidazole (as shown in the figure). The higher oligomeric form is indicated by a solid arrow while the lower oligomeric form is indicated by a dashed arrow. (B) Gel filtration profile of 250mM imidazole elution fraction. The two major peaks, PI and PII along with their respective OD_{260/280} ratios are indicated in the figure. (C) Native polyacrylamide gel analysis of PI or PII fraction obtained in B. The 250mM imidazole elution fraction used for gel filtration also depicted (Elution). M, native molecular mass marker. (D) DLS profile of PI or PII peak. The hydrodynamic radius corresponding to the peak position is indicated above the peak. Circular dichroism (CD) spectra of the PI peak (E) or the PII peak (F).

**Figure 3. Association of RNA with Alr1562.** (A) The purified Alr1562 protein (250mM imidazole fraction) was resolved on native polyacrylamide gel and stained with Coomassie (CBB) or ethidium bromide (Et-Br) as indicated. The higher oligomeric form is indicated by a solid arrow while the lower oligomeric form is indicated by a dashed arrow. (B) The purified Alr1562 protein (250mM imidazole fraction) was treated with DNaseI or RNaseA (as shown in the figure), resolved by non-denaturing PAGE (in 0.5 X TBE) and stained with Et-Br (upper panel) or CBB (lower panel). (C) The Alr1562 protein fractions obtained on elution with increasing concentrations of imidazole (as shown in the figure) were mixed with Trizol to isolate
the protein-associated RNA. The RNA was separated on agarose gel and stained with Et-Br. (D) Native PAGE profile of the Alr1562 protein (250 mM imidazole fraction) treated with RNase and NaCl (T) or left untreated (U) as indicated in the figure. Solid arrow shows the higher oligomeric forms while the dashed arrow indicates the lower oligomeric form. (E) Gel filtration profile of RNaseI and NaCl-treated Alr1562 protein described in D.

**Figure 4. Heat treatment of Alr1562.** (A) Native PAGE profile of the purified Alr1562 protein (50 mM imidazole elution fraction) after heating for 10 min at the indicated temperature. (B) CD spectra of the proteins described in A after incubation at the temperatures indicated. (C) Native PAGE profile of the purified Alr1562 protein (250 mM imidazole elution fraction) after heating for 10 min at the indicated temperature. (D) CD spectra of the proteins described in C after incubation at the temperatures indicated in the figure. (E) The native polyacrylamide gel described in C was also stained with ethidium bromide.

**Figure 5. Electromobility shift assays (EMSA) and electron microscopy.** (A) The three different RNA substrates used for EMSAs. Repeat-spacer-repeat (RSR), only repeat 2 (R2) and non-specific 20mer (20-mer) are schematically depicted. (B) Binding of the dimeric Alr1562 (PII fraction) to the radiolabelled RSR RNA. After electrophoresis, the gel was dried, exposed to X-ray film and subsequently developed. The free γ^{32}P-RSR is indicated by an arrow, while the protein-RNA complex is indicated by *. (C) Competitive EMSA. The dimeric Alr1562 protein (52 μM) was allowed to bind to the radiolabelled RSR RNA in the presence of cold RSR (25-fold or 75-fold molar excess) or cold R2 (25-fold or 75-fold molar excess) or cold 20-mer RNA (25-fold, 75-fold or 200-fold molar excess) as shown in the figure. Transmission electron micrograph of the purified Alr1562 protein (250 mM imidazole elution fraction) (D) or the PII fraction obtained after gel filtration of the 250 mM imidazole elution fraction (E).

**Figure 6. Structural modelling of Alr1562 and EMSAs of the Alr1562 point mutant proteins.** (A) Model of Alr1562 built through I-TASSER, showing the major domains. (B) Overlap of Alr1562 (blue) and RNA molecule (yellow). Residues of Alr1562 predicted to bind to RNA are shown in red. (C) Gel filtration profile of different purified Alr1562 point mutants (indicated in the figure). (D) Binding of the Alr1562 point mutants to the radio-labeled RSR RNA. The amount of protein employed is indicated on top of the well. After electrophoresis, the gels were dried, exposed to X-ray film and subsequently developed. The free RNA is indicated by an arrow while the position of protein-RNA complex is indicated by *.
Figure 1 Bioinformatic analysis and purification of the full-length Alr1562. (A) Schematic representation of the cas genes present within the Type-ID CRISPR locus 3 of *Anabaena* PCC 7120. Arrowheads indicate the direction of transcription. The respective Cas-protein encoded by individual ORFs are also mentioned below each arrowhead. (B) Schematic representation of the Alr1562 protein. The consensus disordered regions identified within the Alr1562 protein (using d2p2 server) are shown in green. The proteins that were over-expressed in *E. coli* i.e. the full-length Alr1562 (Alr1562), Alr1562 lacking the N-terminal disordered region (1562_N100) , Alr1562 lacking the C-terminal disordered region (1562_C44) or Alr1562 lacking both the terminally located disordered regions (1562_DT) are also schematically depicted. The deleted regions of the protein are indicated by dashed line. (C) Purification of the Alr1562 protein. The his-tagged Alr1562 protein was over-expressed in *E. coli* and purified by Ni-NTA affinity chromatography. Clarified cell lysate (Lys), flow through (FT), wash (W) and elution fractions obtained with different concentrations of imidazole (as indicated in the figure) were resolved on denaturing polyacrylamide gels and stained with Coomassie brilliant blue (CBB). M, molecular mass marker. (D) Cell-free total protein extracts obtained from *Anabaena* cultured under nitrogen supplemented (N*) or nitrogen deprived conditions (N') were subjected to denaturing PAGE, transferred onto nitrocellulose membrane and probed with the Alr1562 polyclonal antiserum. The purified Alr1562 protein (obtained from C) is indicated by arrow.
Figure 2. Biophysical properties of Alr1562. (A) Native PAGE profile showing the oligomeric forms of Alr1562 protein obtained on elution with increasing concentrations of imidazole (as shown in the figure). The higher oligomeric form is indicated by a solid arrow while the lower oligomeric form is indicated by a dashed arrow. (B) Gel filtration profile of 250mM imidazole elution fraction. The two major peaks, PI and PII along with their respective OD260/280 ratios are indicated in the figure. (C) Native polyacrylamide gel analysis of PI or PII fraction obtained in B. The 250mM imidazole elution fraction used for gel filtration also depicted (Elution). M, native molecular mass marker. (D) DLS profile of PI or PII peak. The hydrodynamic radius corresponding to the peak position is indicated above the peak. Circular dichroism (CD) spectra of the PI peak (E) or the PII peak (F).
Figure 3. Association of RNA with Alr1562. (A) The purified Alr1562 protein (250 mM imidazole fraction) was resolved on native polyacrylamide gel and stained with Coomassie (CBB) or ethidium bromide (Et-Br) as indicated. The higher oligomeric form is indicated by a solid arrow while the lower oligomeric form is indicated by a dashed arrow. (B) The purified Alr1562 protein (250 mM imidazole fraction) was treated with DNaseI or RNaseA (as shown in the figure), resolved by non-denaturing PAGE (in 0.5 X TBE) and stained with Et-Br (upper panel) or CBB (lower panel). (C) The Alr1562 protein fractions obtained on elution with increasing concentrations of imidazole (as shown in the figure) were mixed with Trizol to isolate the protein-associated RNA. The RNA was separated on agarose gel and stained with Et-Br. (D) Native PAGE profile of the Alr1562 protein (250 mM imidazole fraction) treated with RNase and NaCl (T) or left untreated (U) as indicated in the figure. Solid arrow shows the higher oligomeric forms while the dashed arrow indicates the lower oligomeric form. (E) Gel filtration profile of RNase and NaCl-treated Alr1562 protein described in D.
Figure 4. Heat treatment of Alr1562. (A) Native PAGE profile of the purified Alr1562 protein (50 mM imidazole elution fraction) after heating for 10 min at the indicated temperature. (B) CD spectra of the proteins described in A after incubation at the temperatures indicated. (C) Native PAGE profile of the purified Alr1562 protein (250 mM imidazole elution fraction) after heating for 10 min at the indicated temperature. (D) CD spectra of the proteins described in C after incubation at the temperatures indicated in the figure. (E) The native polyacrylamide gel described in C was also stained with ethidium bromide.
Figure 5. Electromobility shift assays (EMSA) and electron microscopy. (A) The three different RNA substrates used for EMSAs. Repeat-spacer-repeat (RSR), only repeat 2 (R2) and non-specific 20mer (20-mer) are schematically depicted. (B) Binding of the dimeric Alr1562 (PII fraction) to the radiolabeled RSR RNA. After electrophoresis, the gel was dried, exposed to X-ray film and subsequently developed. The free γ-32P-RSR is indicated by an arrow, while the protein-RNA complex is indicated by *. (C) Competitive EMSA. The dimeric Alr1562 protein (52 μM) was allowed to bind to the radiolabeled RSR RNA in the presence of cold RSR (25-fold or 75-fold molar excess) or cold R2 (25-fold or 75-fold molar excess) or cold 20-mer RNA (25-fold, 75-fold or 200-fold molar excess) as shown in the figure. Transmission electron micrograph of the purified Alr1562 protein (250 mM imidazole elution fraction) (D) or the PII fraction obtained after gel filtration of the 250 mM imidazole elution fraction (E).
Figure 6 Structural modelling of Alr1562 and EMSAs of the Alr1562 point mutant proteins. (A) Model of Alr1562 built through I-TASSER, showing the major domains. (B) Overlap of Alr1562 (blue) and RNA molecule (yellow). Residues of Alr1562 predicted to bind to RNA are shown in red. (C) Gel filtration profile of different purified Alr1562 point mutants (indicated in the figure). (D) Binding of the Alr1562 point mutants to the radio-labeled RSR RNA. The amount of protein employed is indicated on top of the well. After electrophoresis, the gels were dried, exposed to X-ray film and subsequently developed. The free RNA is indicated by an arrow while the position of protein-RNA complex is indicated by *.