Saturation mutagenesis charts the functional landscape of *Salmonella* ProQ and reveals a gene regulatory function of its C-terminal domain

Alisa Rizvanovic, Jonas Kjellin, Fredrik Söderbom and Erik Holmqvist

Department of Cell and Molecular Biology, Biomedical Centre, Uppsala University, Uppsala, S-75124, Sweden

Received March 16, 2021; Revised August 03, 2021; Editorial Decision August 04, 2021; Accepted August 10, 2021

**ABSTRACT**

The global RNA-binding protein ProQ has emerged as a central player in post-transcriptional regulatory networks in bacteria. While the N-terminal domain (NTD) of ProQ harbors the major RNA-binding activity, the role of the ProQ C-terminal domain (CTD) has remained unclear. Here, we have applied saturation mutagenesis coupled to phenotypic sorting and long-read sequencing to chart the regulatory capacity of *Salmonella* ProQ. Parallel monitoring of thousands of ProQ mutants allowed mapping of critical residues in both the NTD and the CTD, while the linker separating these domains was tolerant to mutations. Single amino acid substitutions in the NTD associated with abolished regulatory capacity strongly align with RNA-binding deficiency. An observed cellular instability of ProQ associated with mutations in the NTD suggests that interaction with RNA protects ProQ from degradation. Mutation of conserved CTD residues led to overstabilization of RNA targets and rendered ProQ inert in regulation, without affecting protein stability *in vivo*. Furthermore, ProQ lacking the CTD, although binding competent, failed to protect an mRNA target from degradation. Together, our data provide a comprehensive overview of residues important for ProQ-dependent regulation and reveal an essential role for the enigmatic ProQ CTD in gene regulation.

**INTRODUCTION**

The ability of bacteria to survive and thrive in dynamic environments relies on precise and rapid regulation of gene expression. RNA-binding proteins (RBPs) are key factors in post-transcriptional regulatory networks and often work in conjunction with small regulatory RNAs (sRNAs) (1,2). At the molecular level, specific sequence and/or structural RNA motifs are recognized by RBPs through RNA-binding domains, among which the S1 domain, the cold-shock domain, the Sm domain and the K homology domain are well-studied examples (3). The arrangement and combination of RNA-binding domains define the regulatory activity of an RBP (4). To understand how RBPs regulate their RNA targets, it is therefore important to know how these domains function.

Recent advances in global RBP identification have assigned RNA-binding activity to many proteins and domains not previously implicated in RNA–protein interactions (5–16). Among bacterial RBPs, ProQ has recently emerged as a global RNA binder in *Salmonella enterica* (S. enterica) and *Escherichia coli* (E. coli) (13,14,16–18) and is one of the founding members of the ProQ/FimO protein family (PF04352 Pfam) (14). The members of this family are ubiquitously found in proteobacterial species (19–22), and, in addition to ProQ, include plasmid-encoded FimO proteins (23–25), RocC (19,26) and Lpp1663 (27,28) in *Legionella pneumophila*, NMB1681 in *Neisseria meningitidis* (21,29) and *S. enterica* FopA (22). All of these proteins possess the well-conserved ProQ/FimO domain, which has been suggested to harbor the major RNA-binding activity in each respective protein (19,22,28,30–35). Many ProQ/FimO family members have additional N- or C-terminal domains or extensions (19,22,28–30,36), but their contribution to RNA-binding and/or gene regulatory activity largely remains unknown (19,25,30,32,34).

ProQ itself was initially identified as a positive factor for proline uptake and osmoprotection (37–39). It is now well established that ProQ interacts with hundreds of different RNA targets, primarily sRNAs and 3′ untranslated regions (UTRs) of mRNAs (13,14,16,18,32). ProQ recognizes RNA stem-loop structures, flanked by 5′-located A-rich and 3′ located U-rich sequences, which often overlap with intrinsic terminators (13,32). Despite the detailed mapping of ProQ–RNA interactions, the functional consequences following binding are in most cases still unclear. Transcriptomic studies have shown that deletion or overexpression of *proQ* leads to global changes in gene expression (18,40,41), but a direct link between RNA binding and gene regulation has only been described for a few RNA targets. For instance, ProQ
protects the *cspE* mRNA from 3'-dependent RNA degradation by RNase II (13), blocks RNase III-mediated cleavage of the duplex formed by sRNAs RybB and RbsZ (18) and promotes translation inhibition exerted by the sRNA RaiZ on *hupA* mRNA (40). Congruent with ProQ’s role as a global regulator, this protein has been linked to diverse cellular responses, including adaptation to osmotic stress (37–39), biofilm formation (42), adaptation to resource limitation (43), bacterial virulence (44) and motility (40,41).

ProQ folds into two globular N- and C-terminal domains (NTD and CTD) bridged by an extended, unstructured linker (30). The NTD, spanning residues 1–119, constitutes the ProQ/FinO domain (20,30,44,45) and appears to harbor the core RNA-binding activity; the NTD alone is sufficient for binding to the majority of tested RNA ligands in *vitro* (32). Similarly, screening for RNA binding-defective ProQ mutants using a bacterial three-hybrid strategy identified many critical residues within the NTD but none in the CTD (31). In contrast to the ProQ NTD, the ProQ CTD is exclusively found in ProQ but absent from other proteins in the ProQ/FinO family. The CTD, spanning residues 180–228, is mostly composed of β-sheets that form a barrel-like structure, which partially resembles the Tudor domain (30), known to mediate protein–protein interactions via methylated amino acid residues (46). However, Tudor domains are usually found in eukaryotic proteins (30), and no evolutionary relationship between the ProQ CTD and Tudor domains has been convincingly established. Compared to the NTD, the function of the CTD has been convincingly established. The ProQ CTD and Tudor domains has been convincingly established. Compared to the NTD, the function of the CTD (30) is understudied. Even though it may participate in RNA interactions (31), it is considered an important determinant of ProQ stability, since the protein is stabilized through interaction with RNA. In particular, in vivo in the NTD rendered ProQ unstable in *vivo*. These residues largely overlap with those affecting ProQ–RNA interactions (31), indicating an intimate relationship between RNA-binding and RNA strand exchange (30,32,39), the contribution of such activities for ProQ’s overall RNA-binding activity and/or regulatory function is so far unclear.

To better understand the relationship between the RNA-binding activity and regulatory function of ProQ, we developed a screen that allowed for systematic identification of ProQ mutant proteins with impaired regulatory activity. We identified >20 residues in the NTD that are critical for ProQ-dependent gene regulation *in vivo*. These residues largely overlap with those affecting ProQ–RNA interactions (31), indicating an intimate relationship between RNA-binding and RNA gene regulation. Interestingly, mutations in the NTD rendered ProQ unstable in *vivo*, suggesting that the protein is stabilized through interaction with RNA. In addition to the NTD, our screen identified residues in the CTD which, when mutated, lead to overstabilization of several RNA targets and rendered ProQ inactive in activating a reporter gene, without affecting protein stability in *vivo*. In line with this, we show that ProQ lacking the CTD, although binding competent, critically impacts the regulatory activity of ProQ.

**MATERIALS AND METHODS**

**Bacterial strains and growth conditions**

All bacterial strains used in this study are listed in Supplementary Table S1. Bacteria were routinely grown aerobically at 37°C in LB medium with shaking at 200 rpm. Where indicated, the growth medium was supplemented with kanamycin (50 μg/ml), tetracycline (15 μg/ml), chlortetracycline (30 μg/ml), ampicillin (100 μg/ml) or isopropanyl β-D-1-thiogalactopyranoside (IPTG) at 5, 50, 100, 250, 500, 1000 or 2000 μM.

**Cloning**

Plasmids and oligonucleotides used in this study are listed in Supplementary Tables S2 and S3, respectively. To construct transcriptional reporter plasmid pPBlE-gfp (pAR022), plasmid pUA66 was cut with Xhol and BamHI, and ligated to a PCR product amplified from SL1344 genomic DNA (primers EHO-1144/1283) and cut with Xhol and BamHI. The ProQ-plasmid (pAR009) was constructed in three steps. First, promoter PBad was replaced with promoter PlacO on plasmid pH450, yielding plasmid pAR001. Briefly, the PlacO sequence was amplified (primers EHO-1006/1007) from plasmid pZE12-luc and ligated to the PCR product (primers, EHO-1004/1005) amplified from pH450. The PCR products were cut with NsiI and BglII prior to ligation. Second, the lacI repressor gene was added to pAR001 to construct plasmid pAR007. The lacI sequence was amplified (primers EHO-1047/1048) from plasmid pACYCDest-luc, cut with NsiI and ligated to NsiI-digested pAR001. Finally, the proQ ORF, amplified (EHO-1000/1001) from plasmid pZE12-proQ, was inserted into plasmid pAR007. Both the insert and the vector were cut with Ndel and BglII prior to ligation. The pdTomato-ProQ plasmid (pAR018) was constructed in two steps. First, the dTomato gene was amplified (primers EHO-992/993) from strain MH235 and inserted into pAR007 to construct plasmid pAR013. Both insert and vector were cut with Ndel and EcoRI prior to ligation. Second, the proQ sequence was amplified (EHO-1000/1001) from pZE12-proQ and inserted into pAR013. Both the insert and the vector were cut with Ndel and NotI prior to ligation. pAR052 was constructed by adding a 3xFLAG sequence to the 3′-end of the proQ ORF on plasmid pAR009. The pAR009 plasmid was amplified using primers EHO-1614/1615 and ligated. The parental plasmid was digested with DpnI prior ligation. To construct plasmids expressing ProQ truncations, the sequences representing residues 1–129, 1–179 and 130–228 of proQ were amplified (primers EHO-1625/1565, EHO-1625/1626, EHO-1627/1628, respectively) from pAR052, followed by DpnI digestion to remove the parental plasmid, and ligation to construct plasmids pAR053, pAR054 and pAR055, respectively. To construct plasmids expressing 3xFLAG-tagged mutant ProQ proteins, the pAR052 plasmid was amplified using primers EHO-1551/1552, EHO-1560/1561, EHO-1558/1559 and EHO-1562/1563, and ligated, generating plasmid pAR058, pAR059, pAR060 and pAR061, respectively.

**Error-prone PCR**

Error-prone PCR of *proQ* on plasmid pdTomato-ProQ (pAR018) was carried out in a two-step PCR reaction using the GeneMorph EZClone Domain Mutagenesis Kit (Stratagene). The first PCR reaction (Mutant Megaprimer Synthesis) contained 500 ng of plasmid pAR018, dNTPs (200 μM), primers EHS-1123 and EHS-1124, Mutazyme II DNA Polymerase and Mutazyme II Reaction Buffer in a
total volume of 50 μl. The cycling conditions were: 120 s at 95°C, 30 cycles of 30 s at 95°C, 30 s at 56°C, 60 s at 72°C and 10 min at 72°C. The PCR reaction was analyzed on a 1% agarose gel and purified with the GeneJET PCR Purification Kit (ThermoScientific). The second PCR reaction (EZClone Reaction) contained 50 ng of plasmid pAR018, 250 ng of the PCR product from the first PCR reaction (used as primers here), EZClone Enzyme Mix and EZClone Solution in a total volume of 50 μl. The cycling conditions were: 60 s at 95°C, and 25 cycles of 50 s at 95°C, 50 s at 60°C, 840 s at 68°C. After 2 min on ice, 1.5 μl of DpnI was added for 2 h at 37°C to remove template plasmid.

**ProQ mutant library preparation**

*Escherichia coli* TOP10 cells (Invitrogen) were made competent by washing exponentially growing cells three times in cold sterile H2O, after which 50 μl of competent by washing exponentially growing cellsthreetimesin *Escherichia coli* ProQ mutant library preparation.

95/H9262 total volume of 50 μl. The cycling conditions were: 60 s at 95°C, and 25 cycles of 50 s at 95°C, 50 s at 60°C, 840 s at 68°C. After 2 min on ice, 1.5 μl of DpnI was added for 2 h at 37°C to remove template plasmid.

**Fluorescence-activated cell sorting (FACS)**

Sample aliquots of the mutant library and SL1344 control cells deleted for proQ, harboring the PφfIE-gfp reporter (pAR022) and either the non-mutated pdTomato-ProQ plasmid (pAR018) or an empty control (pAR007), were diluted 100-fold in fresh LB-medium and grown overnight at 37°C. Cells were pelleted by centrifugation for 5 min at 3700 × g, resuspended in RNAlater (ThermoScientific) and diluted 100-fold in sterile PBS. FACS was carried out using a MoFlo Astrios EQ (Beckman Coulter, USA) cell sorter with 488 nm and 532 nm lasers for excitation of GFP and dTomato, respectively. The trigger was set on forward scatter at a threshold of 0.05%. Sorted cells (2 × 106 cells/pool) were re-analyzed by the MoFlo Astrios EQ cell sorter and the MACSQuant VYB flowcytometer. Data analysis of approximately 20 000 cells/pool was done using the FlowJo™ Software. The remaining sorted cells were pelleted by centrifugation and resuspended in 100 μl of H2O.

**High-throughput sequencing**

Each pool of sorted cells was used as template in a PCR reaction to generate sequences for high-throughput sequencing analysis. Before PCR, sorted cells were denatured at 95°C for 5 min. Each PCR reaction contained 2 μl of sample template, dTNPs, 10 μM barcode-specific primers (forward primer: EHO-1436–1447, reverse primer: EHO-1448), Phusion High-Fidelity (HF) DNA Polymerase (ThermoScientific) and Phusion HF buffer (ThermoScientific) in a total volume of 50 μl. PCR program: 30 s at 98°C, 30 cycles of 10 s at 98°C, 30 s at 70°C, 15 s at 72°C and 5 min at 72°C. PCR products were analyzed on 1% agarose gels and purified using the GeneJET PCR Purification Kit, followed by DNA concentration measurements using Qubit Fluorometric Quantitation (ThermoScientific). All samples were pooled, of which 500 ng was used for sequencing-library preparation. Library preparation and sequencing of the pooled DNA sample (barcoded amplicons) were performed at the Uppsala Genome Centre (UGC), a national facility within the National Genomic Infrastructure (NGI) hosted by Science for Life Laboratory (SciLifeLab) in Uppsala, Sweden. Sequencing libraries were constructed using the SMRTbell™ Template Prep Kit 1.0 (500 bp to 20 kb) and purified with the AMPure Clean-Up kit (PacBio). The library quality was evaluated using a Bioanalyzer Chip (Agilent) before sequencing. Sequencing was performed on a Sequel III SMART® Cell using the Sequel II system (PacBio).

**Bioinformatic analysis**

* Primer sequences were trimmed from the circular consensus reads (CCS) obtained from the PacBio sequencing with cutadapt v. 2.26.0 (47). To assure that all the reads were output in the same orientation, cutadapt was run with the --revcomp flag. Further sequencing read filtering and subsequent quantification of mutations were performed with python scripts...
Site-directed mutagenesis

Identified mutations were re-constructed by site-directed mutagenesis. The PCR reactions contained 10 ng of template plasmid pAR018 or plasmid pAR009, dNTPs (200 μM), mutagenic primers listed in Supplementary Table S3, Phusion High-Fidelity DNA Polymerase and Phusion HF buffer in a total volume of 50 μL. The PCR products were analyzed on 1% agarose gels and purified with the GeneJet PCR Purification Kit. After removal of template plasmid by inserting the ORF encoding the red fluorescent protein levelssimilartoendogenousexpressionfromthenativeRNA (tmRNA, EHO-867) detection are shown in Supplementary Table S3.

RESULTS

Construction of a fluorescence-based reporter monitoring ProQ-dependent gene expression

ProQ is a global RNA-binding protein that affects the expression of a large number of genes (13,14,16,18,41). However, a direct connection between RNA binding and gene regulation has only been examined for a few RNA ligands (13,18,40). Regarding ProQ’s RNA-binding activity, a recent study used a three-hybrid strategy to identify ProQ residues critical for stable RNA interactions in vivo (31). As a complementary approach, we here developed a screen for high-throughput identification of ProQ mutants that abolish the protein’s gene regulatory capacity. To this end, we first aimed at constructing a fluorescence-based reporter amenable to high-throughput screening for ProQ-dependent effects on gene expression. Transcriptomic analyses have consistently associated deletion of proQ with reduced expression of motility genes (14,41); the majority of genes involved in flagellar assembly are downregulated in a proQ deletion strain (Figure 1A) (14). In Salmonella, the flagellar system comprises over 60 genes, which are organized into a transcriptional hierarchy of three promoter classes. On top of the hierarchy, class I promoters express the FlhDC master regulator that drives transcription, either directly or indirectly, of all downstream genes (Figure 1B) (51–54). Based on this well-characterized hierarchical regulation of flagellar gene expression, we hypothesized that the observed ProQ-dependent effects stem from a direct post-transcriptional event at the top of the hierarchy, which results in indirect effects on transcription downstream (Figure 1B). To test this, we constructed a transcriptional GFP reporter driven from the promoter of fliE, a class II gene that is directly activated by the FlhDC master regulator (Figure 1B, C and Supplementary Figure S1A). As expected, deletion of proQ resulted in a significant reduction in fluorescence from the PfliE-gfp construct, indicating that ProQ indirectly activates transcription of fliE. Complementing the ΔproQ strain with an IPTG-inducible ProQ overexpression plasmid resulted in a direct and positive correlation between IPTG concentration, ProQ levels and PfliE activity (Figure 1C). In the absence of IPTG, ΔproQ cells harboring the ProQ complementation plasmid had similar PfliE activity as wild-type cells with an empty plasmid. This is consistent with leaky expression from the complementation plasmid, giving ProQ protein levels similar to endogenous expression from the native proQ locus (Supplementary Figure S1B). Hence, the PfliE-gfp construct indirectly reports on ProQ-dependent regulation of gene expression and should enable screening of ProQ mutants with reduced regulatory activity.

Next, to allow monitoring of proQ expression by measuring fluorescence, we modified the ProQ expression plasmid by inserting the ORF encoding the red fluorescent dTomato protein in-frame with ProQ. While induction of dTomato alone did not affect the fliE promoter, induction of dTomato-ProQ resulted in similarly high GFP levels as did overexpression of native ProQ (Supplementary Figure...
S1C). Cells harboring both the P fizE - gfp construct and the dTomato-ProQ plasmid were subjected to single cell analysis to simultaneously monitor dTomato-ProQ expression and ProQ-dependent regulation. While the majority of cells harboring an empty plasmid gave low fluorescence levels in both the green and red channels, the majority of cells harboring the dTomato-ProQ-expressing plasmid gave high fluorescence levels for both channels (Supplementary Figure S1D). Of note, expression from the fizE promoter was heterogeneous, leading to co-existing populations of cells expressing either low or high GFP, with the distribution of cells within each state being determined by the concentration of dTomato-ProQ. This is in line with previous findings of heterogenous expression of flagellar genes, yielding co-existing on- and off-states in response to external factors (55–57). Thus, the combination of two fluorescent reporters allows for simultaneous monitoring of ProQ expression and ProQ-dependent regulation.

**Generation and sorting of a ProQ mutant library**

Having established a system in which the activity of ProQ can be easily detected using the readout of a fluorescence reporter, we next set out to delineate the regulatory capacity of ProQ by mutational studies. To enable high-throughput identification of ProQ mutants impaired in mediating gene regulation, we used our previously developed method for saturation mutagenesis coupled to phenotypic cell sorting (58). Briefly, we introduced mutations into the proQ ORF by error-prone PCR and cloned mutant proQ sequences in-frame with dTomato on the IPTG-inducible plasmid (Figure 2A). The mutant library was transformed into Δ proQ Salmonella cells carrying the P fizE - gfp reporter, yielding approximately 67 000 transformants. This number is predicted to cover each possible single nucleotide mutation in the proQ sequence (58). Single cell analysis showed that the majority of cells transformed with the mutant library gave dTomato fluorescence levels similar to the non-mutated control (Figure 2B), indicating stable expression of most dTomato-ProQ mutants. Among these cells, half resulted in high GFP fluorescence levels similar to cells expressing non-mutated dTomato-ProQ, while the other half phenocopied the low GFP levels of Δ proQ cells harboring an empty vector. These results demonstrate that the mutant library contained both ProQ mutants that retained and lost the protein’s regulatory capacity.

The mutant library was physically sorted into two pools based on dTomato and GFP levels using fluorescence-activated cell sorting (FACS). Gates for GFP fluorescence were set according to levels in cells harboring the P fizE - gfp reporter combined with either an empty control vector (low GFP) or the non-mutated dTomato-ProQ plasmid (high GFP), yielding approximately a 5-fold difference in GFP levels between gates. The gate for dTomato fluorescence was set according to fluorescence from the non-mutated dTomato-ProQ protein (Figure 2C). Pool 1 contained cells expressing high dTomato and high GFP, representing ProQ mutants that retained their ability to activate P fizE - gfp. Pool 2 contained cells expressing high dTomato and low GFP, representing ProQ mutants impaired in P fizE activation. As a control, cells carrying a non-mutated dTomato-ProQ were sorted for high dTomato and high GFP. The sorting was performed on three independent cultures, each of which gave approximately 2 million cells per pool. Successful sorting was verified by independent single cell analysis (Figure 2C and Supplementary Figure S2).

**Sequencing data analysis**

After FACS sorting, sequences covering the mutated proQ ORF were PCR-amplified from each pool using barcoded primers, pooled together and subjected to high-throughput long-read sequencing. After high-throughput sequencing and read processing, we obtained between 6700 and 17 100 reads of correct length for each of the different pools.
Figure 2. Generation and sorting of a ProQ mutant library. (A) Overview of the experimental setup. (B) Flow cytometry analysis of Salmonella ΔproQ strains carrying a PflIE-gfp reporter construct together with the empty control vector (gray), the non-mutated dTomato-ProQ expression plasmid (pdTomato-ProQ, black) or the mutant dTomato-ProQ plasmid library (light purple: all cells, dark purple: cells with high dTomato expression levels). Cells were grown in LB-medium supplemented with IPTG (500 μM final concentration) to induce expression of dTomato-ProQ fusions. dTomato and GFP fluorescence were measured for single cells. Gates were set according to the non-mutated control (indicated by dotted line). (C) Salmonella cells harboring the ProQ mutant library were first sorted into different pools based on dTomato and GFP fluorescence levels. Pool 1, sorted for high dTomato and high GFP, represents ProQ mutants with intact function. Pool 2, sorted for high dTomato and low GFP, represents ProQ mutants with impaired function. Successful sorting was verified by independent single cell analysis. dTomato and GFP fluorescence were measured as indicated. Representative data are shown for one out of three replicates for each pool.

Quantifying the number of reads with mutated proQ nucleotide sequences for each pool showed that 96% (±1.0%) of reads for the non-mutated ProQ control had wild-type sequence (Figure 3A). In libraries prepared from pool 1 samples (mutant library, high GFP), 19.0% (±0.1%) of sequencing reads had wild-type sequence, in comparison to 6.2% (±0.4%) for libraries prepared from pool 2 samples (mutant library, low GFP). Thus, for pool 2 samples, representing ProQ mutants with impaired function, approximately 94% of sequencing reads had at least one nucleotide substitution.

We limited the downstream analysis to mutants with a single amino acid substitution. PCA revealed formation of dense clusters for the three replicates from each pool, which were well separated from samples from other pool, indicating high reproducibility of the data (Figure 3B). In pool 1, single amino acid substitutions were distributed across the entire ProQ sequence, covering 100% of all residues (Figure 3C). In pool 2, mutations were also found in 100% of all residues but were mainly present in the N- and C-terminal domains. To find amino acid changes that were significantly associated with either retained or impaired ProQ function, DESeq2 analysis was applied to pool 1 and pool 2 samples. In total, 66 amino acid substitutions showed a statistically significant (FDR-adjusted \( P \leq 0.1 \)) association with either pool 1 or pool 2 (Figure 3D and Supplementary Table S4). As expected, significantly enriched mutations that led to nonsense codons were exclusively found in pool 2 and were excluded from further analysis. Together, 30 distinct nonsynonymous substitutions, spanning 24 residues, were significantly enriched in pool 2 (Supplementary Table S4), indicating that these mutations disrupted the ProQ regulatory function in vivo. These substitutions were located in both the ProQ NTD and CTD, while pool 2 completely lacked significant mutations in the linker region. In
Figure 3. Sequencing data analysis. (A) Distribution of mutations. Percentage of reads obtained with 0, 1, 2, 3 or $\geq 4$ nucleotide changes for sorted pools of the mutated library (pool 1, pool 2) and non-mutated control (pdTomato-ProQ). Mean values and SD are given for three replicates. (B) Principal component analysis of data clustering for sorted mutant libraries (pool 1, pool 2). PC1 and PC2 represents 50.7% and 17% of the variance in the data, respectively. Three replicates are shown for each sorted pool. (C) Distribution of single amino acid substitutions in samples from pools of the mutated library (pool 1, pool 2) and non-mutated control (pdTomato-ProQ). On top, a ProQ domain representation with the N-terminal domain (NTD, orange), the linker region (gray) and the C-terminal domain (CTD, yellow). (D) Substitution profile based on DESeq2 analysis of differences in mutation patterns between pool 1, sorted for high dTomato and high GFP, representing ProQ mutants with intact function, and pool 2, sorted for high dTomato and low GFP, representing ProQ mutants with impaired function. A colored bar indicates a unique amino acid substitution detected for the corresponding amino acid position. The colors indicate whether a specific amino acid substitution was over-represented (red) or underrepresented (blue) in pool 2 compared to pool 1. Only those substitutions that differed significantly (FDR-adjusted $P$-value $\leq 0.1$) between the pools are shown. The ProQ N-terminal domain, linker region and C-terminal domain are indicated (NTD, linker, CTD). Each column indicates by which amino acid the amino acid position was replaced to. Asterisk (*) indicate stop codon. All positions of mutations that were significantly enriched can be found in Supplementary Table S4.
contrast, five substitutions mapping to the linker were significantly enriched in pool 1 samples, indicating that, in comparison to the globular NTD and CTD, changes in the linker are generally less detrimental to ProQ function.

**Loss-of-function mutations in the ProQ NTD correlate with impaired RNA-binding activity**

The majority of amino acid substitutions that were significantly enriched for loss-of-function were located in the ProQ NTD and covered 21 residues (Figure 3D and Supplementary Table S4). Most of these residues are well conserved in proteins within the ProQ/FinO family (Supplementary Figure S3) and thus likely to be functionally important. The mutated residues include both surface-exposed and buried residues, based on their location in the NMR structure of the *E. coli* ProQ NTD (Supplementary Figure S4A) (30), which is >99% identical to the Salmonella ProQ NTD. The majority of mutated buried residues are hydrophobic, indicating that the mutations may compromise the core structure of the NTD. Mutated residues on the NTD surface included hydrophobic (L63P, L71P, V74D), basic (R20P, R80H), acidic (D82E) residues and one glycine residue (G85C). The mutated surface residues form distinct clusters on the concave- and convex sides of the domain, respectively (Figure 4A and Supplementary Figure S4A), in line with a previously proposed model for *E. coli* ProQ–RNA interaction (31).

To validate the NTD mutation data, we selected enriched mutations from the dataset, re-introduced them in the dTomato-ProQ plasmid, and assayed them individually. For this analysis, we selected loss-of-function mutations located either in the hydrophobic core (L34Q, L57R, L83R) or at the protein surface (R80H, D82E, G85C) (Supplementary Figure S4A), as well as one mutation (A18V) enriched in pool 1. The generated mutants were tested for activation of the transcriptional PfliE-gfp reporter using flow cytometry (Figure 4B and Supplementary Figure S4B). All seven ProQ mutants displayed dTomato levels comparable to the parental non-mutated control, indicating a strong loss-of-function phenotype. Next, we assayed protein levels by western blot using an anti-ProQ antibody (Figure 5B). Surprisingly, although all six loss-of-function mutants were expressed, their levels were substantially lower than the non-mutated control and the A18V mutation. The GFP levels for these six mutants were similar to that of a proQ deletion strain, indicating a strong loss-of-function phenotype. Next, we assayed protein levels by western blot using an anti-ProQ antibody (Figure 5B). Surprisingly, although all six loss-of-function mutants were expressed, their levels were substantially lower than the non-mutated control and the A18V mutation. Since wild-type and mutated ProQ proteins were expressed from the same promoter, we hypothesized that lower expression levels may stem from reduced protein stability. To test this, we selected a mutant with a substitution in the conserved R80 residue, shown to be crucial for RNA binding (30,31). Expression of the R80H mutant or wild-type ProQ was induced by IPTG until the cells reached exponential phase, after which *de novo* protein synthesis was arrested by changing to inducer-free medium. Protein levels were analyzed by western blot using an anti-ProQ antibody (Figure 5C). While levels of wild-type ProQ stayed stable for 1 h after media shift, the R80H mutant protein was hardly detectable after 30 min. This result indicates a clearcut relationship between RNA-binding activity and protein stability and suggests that association with RNA stabilizes the ProQ protein.

**ProQ CTD residues critical for ProQ functionality**

Compared to the NTD, the function of the CTD is less well understood, although a role in RNA-binding and RNA...
strand exchange activity has been suggested (30,32,39). Interestingly, removing the CTD from ProQ does not impair binding to the majority of RNA ligands tested so far, neither in vivo nor in vitro (31,32). However, whether the CTD is required for ProQ’s ability to mediate gene regulation is not known. Our screen identified three mutations in CTD residues that were significantly enriched for impaired activation of the PfilE-gfp reporter (Figure 3D and Supplementary Table S4). Mapping these residues onto the NMR structure of the ProQ CTD showed that all residues were exposed on the protein surface (Figure 6A). Two out of three residues are also moderately or strongly conserved among homologous proteins (Supplementary Figure S5).

Re-introduction of the significantly enriched CTD mutations into the pdTomato-ProQ plasmid followed by PfilE-gfp reporter activity measurements confirmed that all three mutations rendered ProQ ineffective in activating the reporter (Figure 6B). We also re-introduced the CTD mutations into the plasmid expressing native ProQ. GFP measurements again showed lower GFP levels compared to the non-mutated ProQ control (Figure 6C). However, in contrast to the NTD mutations that rendered ProQ unstable, the CTD mutant proteins were expressed at the same (G216D) or even higher (G185V, T200P) levels than wild-type ProQ (Figure 6D). This indicates that the tested CTD mutations impair the regulatory activity of ProQ without affecting protein stability and points at a key function for the CTD in ProQ-dependent gene regulation. To dissect how the identified CTD residues contribute to ProQ-dependent regulation of different targets, the CTD mutants were tested for stabilization of three known ProQ target RNAs, cspE and cspD mRNAs, as well as the sRNA RaiZ. RNA was extracted from cells expressing the three CTD mutants (G185V, T200P, G216D) and subjected to northern blot analysis (Figure 6E). As expected (13,40), the levels of all three targets increased upon complementation of ΔproQ with plasmid-expressed wild-type ProQ. Somewhat surprisingly, expression of each of the CTD mutants lead to strongly elevated levels of all three RNA targets. Taken together, these results indicate that residues within the CTD render ProQ unable to activate flagellar gene expression and overstabilizes direct RNA targets.

The ProQ CTD is required for activation of gene expression in vivo

The data presented in Figure 6 show that mutations in the CTD are detrimental for ProQ functionality in vivo. Yet, ProQ lacking the CTD is not impaired in binding to the majority of RNA ligands tested so far (31,32). To
systematically test the contribution of the NTD and CTD to ProQ's regulatory function, we cloned truncated ProQ variants in the IPTG-inducible plasmid (Figure 7A) and transformed them into ΔproQ cells harboring the PfliE-gfp reporter construct. In order to monitor expression of truncated ProQ proteins, all constructs were tagged with a C-terminal 3xFLAG peptide. Western blot analysis using an anti-FLAG antibody showed that all truncated ProQ proteins were expressed, with the full-length and the NTD + linker variants at slightly higher levels than the NTD and linker + CTD variants (Figure 7B). Consistent with results presented in Figure 1C, full-length ProQ strongly activated the PfliE-gfp reporter (Figure 7C). However, truncating ProQ by removing either the NTD or the CTD resulted in low promoter activities of the reporter gene that were indistinguishable from that of the strain lacking proQ entirely. Since the CTD is dispensable for binding to many RNA ligands (31,32), but required for gene regulation (Figure 7C), we tested whether truncated ProQ variants could compete with full-length ProQ expressed from its native locus. To this end, wild-type Salmonella cells harboring the PfliE-gfp reporter were transformed with plasmids overexpressing truncated ProQ variants and assayed for GFP fluorescence. Overexpression of full-length ProQ increased GFP expression beyond the activation achieved by chromosomally expressed ProQ, but neither the NTD nor the CTD alone had any effect on GFP expression (Figure 7D and Supplementary Figure S6). This suggests that although the NTD alone is binding competent, it is not able to compete with full-length ProQ, even when overexpressed.

The impact of the ProQ CTD on stabilization of RNA targets in vivo

The cspE mRNA has emerged as a model RNA ligand for analyzing ProQ binding and regulation. Binding of ProQ to cspE mRNA has been analyzed in vivo by crosslinking (13,18) and a three-hybrid system (31), as well as in vitro by electromobility shift assays (32) and hydrogen deuterium exchange (HDX) experiments (30). The interaction between ProQ and cspE was shown to rely solely on the NTD; ProQ lacking the CTD binds as well to cspE mRNA as does the full-length protein in vivo (31) and in vitro (32). In accordance with these results, co-immunoprecipitation experiments verified that truncating ProQ by removal of the CTD + linker conferred binding to the cspE mRNA (Supplementary Figure S7A). Since removing the CTD decreased ProQ's ability to activate the PfliE-gfp reporter (Figure 7C), we wondered whether removing the CTD would affect ProQ's ability to stabilize RNA targets. To test this, RNA harvested from cells expressing the truncated ProQ variants were subjected to northern blot analysis using probes against cspE mRNA and RaiZ. The results presented in Figure 7E show that while full-length ProQ strongly elevated the steady-state levels of both RNAs, all truncated versions failed to affect cspE mRNA levels, while the NTD alone conferred stability to RaiZ. To test whether the reduced cspE mRNA levels upon truncation of ProQ were due to a reduction in RNA stability, cells were subjected to transcriptional arrest by rifampicin, and RNA was isolated 2, 4 and 8 min after rifampicin addition, followed by northern blot analysis (Supplementary Figure S7B). Complementation of a ΔproQ strain with full-length ProQ, but not the ProQ NTD, restored cspE stability. The reduced mRNA stability in the NTD-expressing strain was dependent on RNase II, as introducing an nrb deletion into the strain restored stability. This suggests a function for the ProQ CTD in protecting the cspE mRNA against RNase II mediated-exoribonuclease activity. Thus, even though the NTD is sufficient for RNA binding, the CTD is required for mRNA stabilization.

DISCUSSION

The recent discovery of ProQ as a global RNA-binding protein in Salmonella and E. coli has spurred a revived interest in ProQ biology. However, many aspects still remain poorly understood, including how ProQ regulates its RNA targets. In this work, we have systematically charted the
functional landscape of ProQ and revealed residues essential for its gene regulatory function. This was possible by exploiting the fact that ProQ activates flagellar gene expression, allowing the construction of a fluorescent reporter for ProQ-dependent regulation (Figure 1). By combining saturation mutagenesis with phenotypic cell sorting and high-throughput sequencing (Figure 2), mutations in all amino acid residues within ProQ were assessed for effects on regulatory activity (Figure 3). We show here that mutations in the globular NTD and CTD domains, but not the linker region, are detrimental for ProQ-dependent gene regulation in vivo (Figures 3–6). As judged from comparison to a recent mutational study, residues in the NTD appear to be critical for both RNA-binding and regulation (Supplementary Figure S4C) (31). Strikingly, NTD mutations impairing both RNA-binding and regulatory activity render ProQ unstable in vivo (Figure 5), which suggests that ProQ's association with RNA enhances protein stability. While the major RNA-binding activity of ProQ seem to be specific to the NTD (30–32), our data indicate an important role for the CTD in gene regulation. Mutations in conserved CTD residues (Figure 6), or removal of the entire CTD (Figure 7), abolish ProQ-dependent gene regulation. Taken together, this indicates that ProQ's capacity to bind and regulate RNA targets requires both the NTD and CTD.

Global transcriptomic analyses have consistently associated deletion of proQ with reduced expression of motility genes, suggesting that ProQ controls flagellar gene expression (14, 41). Our work shows that ProQ indirectly activates transcription from the class II fliE promoter (Figure 1C).
The transcriptional reporter PflfE-gfp was exploited here for high-throughout screening of ProQ mutants with reduced regulatory activity. One important question is how ProQ activates the fliE promoter. In Salmonella and E. coli, the fliE promoter is directly activated by FlhDC, the master transcriptional regulator of the flagellar network (Figure 1B) (51–54). We speculate that ProQ either acts upstream or downstream of flhDC transcription or regulates the flhDC mRNA by a post-transcriptional mechanism. In fact, the flhDC mRNA has previously been identified as a direct ProQ ligand in cross-linking experiments (13). Beyond the scope of the current study, we are currently investigating the mechanism behind ProQ-dependent activation of flagellar genes, which will be reported elsewhere.

On the molecular level, we have identified >20 NTD residues that are critical for the regulatory activity of ProQ (Figures 3–5 and Supplementary Table S4). Our data complement the findings of Pandey et al. (31) and provide a direct link between RNA-binding activity and regulation (Supplementary Figure S4C and Supplementary Table S4). Conserved residues on both the concave and convex sides of the NTD appear to be functionally important, including surface-exposed acidic, basic and hydrophobic residues, and core-located hydrophobic residues (Figure 4; Supplementary Figures S3 and S4). In fact, several of these NTD residues have been suggested to mediate RNA-binding and/or regulation in other ProQ/FinO proteins (19,22,28). For example, random mutagenesis of Legionella pneumophila RocC revealed that several residues in the ProQ/FinO domain (ProQ numbering: L31, F21, L63 and L71) affected its ability to repress competence and to stabilize sRNA (19).

ProQ is an abundant, constitutively expressed protein (14,42). However, whether ProQ levels and/or activity can
be regulated remains unknown. Interestingly, our data show that mutations in the NTD RNA-binding surface affect the intracellular levels of ProQ (Figure 5). It appears that binding to RNA stabilizes the ProQ protein, while disruption of RNA-binding through mutations renders ProQ unstable. We predict that ProQ-bound RNA ligands protect otherwise susceptible amino acid residues within the protein against proteolytic degradation. This seems to be a distinctive feature for native ProQ (Figure 5); fusing ProQ to other proteins such as the red fluorescent protein dTomato (Figure 4), or the alpha-subunit of RNA-polymerase (31), enhances protein stability and masks the destabilizing effect of NTD mutations (Figure 4). One important question is which protein is responsible for degradation of mutant ProQ proteins. In a recent pre-print presenting a saturation mutagenesis study of Salmonella ProQ, the authors suggest that the Lon protease is responsible for removal of RNA-binding deficient mutants of both ProQ and Hfq (El Mouali, Y., Ponath, F., Scharrer, V., Wenner, N., Hinton, J. C. D. and Vogel, J. (2021) Scanning mutagenesis of RNA-binding protein ProQ reveals a quality control role for the Lon protease. bioRxiv doi: https://doi.org/10.1101/2021.07.12.452043, 12 July 2021, pre-print: not peer-reviewed). Thus, the observed instability of RNA-binding deficient ProQ NTD variants presented here (Figure 4), most likely depends on Lon. We anticipate that future studies will shine light on how ProQ protein levels are regulated and clarify the relationship between RNA-binding and protein stability.

While the ProQ NTD has been the subject of several recent studies (30–32), the function(s) of the CTD has remained unclear. Our experiments reveal that mutations in three CTD residues (G185, T200P, G216) completely abolish the ability of ProQ to activate the PflIE-gfp reporter (Figure 6). The same mutations lead to strongly elevated levels of three known ProQRNA ligands, which likely reflect over-expression of the CTD domain (30,31). It thus appears that residues within the CTD may contribute to or impact RNA binding but only for specific RNA ligands. Along this line, it was recently proposed that while the NTD is required for binding to all ligands, the CTD may be involved in RNA-binding only to specific ligands (32). In a similar manner, N- or C-terminal extensions present in other ProQ/Fio proteins are suggested to widen target recognition for each protein member (19–21). The increased stability of RNA ligands associated with the CTD mutations identified in this study (Figure 6) may suggest that these mutant proteins harbor a CTD with increased RNA-binding activity.

Perhaps the most striking result to emerge from our data is that the CTD is required for the regulatory activity of ProQ (Figure 7 and Supplementary Figure S7B). As exemplified by the model ligand cspE, the CTD is dispensable for mRNA-binding (32) (Supplementary Figure S7A) but necessary for full RNA stabilization (Figure 7E and Supplementary Figure S7B). This was true for mRNA stabilization of cspE but not sRNA stabilization of RaiZ (Figure 7E). Thus, it appears that the ProQ CTD is required for ProQ-dependent regulation for a set of RNA ligands but not for all. Given that ProQ protects the cspE mRNA from RNase II-mediated degradation in Salmonella, we speculate that the ProQ CTD is involved in blocking RNase activity. How does this occur mechanistically? As judged from co-immunoprecipitation experiments, ProQ does not seem to interact with RNase II in vivo (13), pointing against direct contact between the CTD and RNase II as a mechanism of action. It is possible that the CTD may sterically block access of RNase II to the cspE mRNA (13) or bind to other proteins that may block RNAse activity, directly or indirectly. Further work will be required to identify any potential connection between the ProQ CTD and RNA stabilization. A related issue is whether ProQ is dependent on, and/or interacts with, other proteins. In an attempt to shine light on this we induced the plasmid-encoded NTD or CTD in a strain encoding full-length ProQ on the chromosome. The experiment shown in Figure 7D shows that the single domains are not able to interfere with ProQ function (which should result in reduced fluorescence from the reporter), suggesting that the domains alone do not compete for binding to trans-acting factors. However, more experiments are certainly needed for addressing this issue.

In summary, we have identified functionally important residues within Salmonella ProQ, and revealed the significance of its CTD in gene regulation. We expect that future analyses will expand the individual role(s) of this domain, and elucidate how it cooperates with the NTD to orchestrate the biological role of the global RBP ProQ.

DATA AVAILABILITY

Sequencing data is available as a National Center for Biotechnology Information (NCBI) Bioproject under accession number PRJNA713973. Python scripts used for sequencing read filtering and subsequent quantification of mutations are available in the GitHub repository and can be accessed via https://github.com/kjellinjonas/proQ_mutagenesis.

FCS files for flow cytometry and FACS data sets have been deposited in the FlowRepository database and can be accessed via http://flowrepository.org/id/FR-FCM-Z3JW.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to E.G.H. Wagner for critically reading the manuscript. We acknowledge support of the Microbial Single Cell Genomics Facility (SiCell) for providing assistance.
in single cell sorting, and the National Genomic Infrastructure (NGI)/Uppsala Genome Center and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure. We thank J. Vogel for sharing the anti-ProQ antibody, S. Koskineni for sharing plasmid pCH450, and M. Hoekzema for sharing strain MH235.

**FUNDING**

Swedish Research Council [2016-03656]; Swedish Foundation for Strategic Research [ICA16-0021]; Uppsala Antibiotic Center; Åke Wiberg Foundation; RFI/VR and Science for Life Laboratory, Sweden. Funding for open access charge: Stiftelsen för Strategisk Forsknings. *Conflict of interest.* None declared.

**REFERENCES**

1. Wagner, E.G.H. and Romby, P. (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.*, **90**, 133–208.

2. Holmqvist, E. and Vogel, J. (2018) RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.*, **16**, 601–615.

3. Corley, M., Burns, M.C. and Yeo, G.W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell.*, **78**, 9–29.

4. Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: Modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.

5. Baltz, A.G., Munschaever, M., Schwanzäuser, B., Vasiile, A., Murakawa, Y., Schuler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M. et al. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell.*, **46**, 674–680.

6. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. et al. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.

7. Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., Schwarzl, T., Curk, T., Foehr, S., Huber, W. et al. (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.*, **6**, 10127.

8. Asencio, C., Chatterjee, A. and Hentze, M.W. (2018) Silica-based solid-phase extraction of cross-linked nucleic acid–bound proteins. *Life Sci. Alliance.*, **1**, e201800088.

9. Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.

10. Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.M., Ramakrishna, M., Harvey, R.F., Dezi, V. et al. (2019) Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.*, **37**, 169–178.

11. Shchepachev, V., Bresson, S., Spanos, C., Petalski, E., Fischer, L., Rappoloser, J. and Tollervey, D. (2019) Defining the RNA interactome by total RNA–associated protein purification. *Mol. Syst. Biol.*, **15**, e8689.

12. Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Grimm, S., Selbach, M. et al. (2019) Purification of cross-linked RNA-protein complexes by phenol-toluene extraction. *Nat. Commun.*, **10**, 990.

13. Holmqvist, E., Li, L., Bischler, T., Barquist, L. and Vogel, J. (2018) Global maps of ProQ binding in vivo reveal target recognition via RNA structure and stability control at mRNA 3′ ends. *Mol. Cell.*, **70**, 971–982.

14. Smirnov, A., Förster, K.U., Holmqvist, E., Otto, A., Günstner, R., Becher, D., Reinhardt, R. and Vogel, J. (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 11591–11596.

15. Hör, J., Gorskii, S.A. and Vogel, J. (2018) Bacterial RNA biology on a genome scale. *Mol. Cell.*, **70**, 785–799.
