

COMMENTARY

Recognition of a pseudo-symmetric RNA tetranucleotide by Csx3, a new member of the CRISPR associated Rossmann fold superfamily

Ece Topuzlu and C. Martin Lawrence

Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA

ABSTRACT

The CRISPR/Cas adaptive immune system shows extreme diversity in the number of CRISPR/Cas types and subtypes, and in the multitude of CRISPR associated protein families of which they are composed. Despite this diversity, the roles of many Cas protein families are now defined with regard to spacer acquisition, crRNA biogenesis, and DNA or RNA surveillance and targeting. However, a number of unclassified CRISPR-Cas proteins remain. Such proteins have traditionally been designated as CRISPR subtype x (Csx). Here we revisit the structural analysis of one such protein, Csx3, and show that this homodimeric protein utilizes a Rossmann fold for the recognition of an RNA tetranucleotide. Tertiary and quaternary structural similarities of Csx3 to CRISPR/Cas proteins Csx1 and Csa3 are identified and suggest Csx3 is a new member of the CRISPR Associated Rossmann Fold (CARF) superfamily. The structure of the Csx3/RNA complex illustrates one way CARF domain proteins may recognize pseudo-symmetric polynucleotides.

Abbreviations: CRISPR, Clustered Regularly Interspaced Palindromic Repeats; Cas, CRISPR associated; Csx, unclassified CRISPR associated subtype; Csa, CRISPR associated subtype A; Csm, CRISPR associated subtype M; CARF, CRISPR associated Rossmann fold; RAMP, Repeat Associated Mysterious Protein; SCOP, Structural Classification of Proteins; RRM, RNA Recognition Motif.

ARTICLE HISTORY

Received 4 November 2015
Accepted 7 December 2015

KEYWORDS

CARF; Cas; CRISPR;
CRISPR/Cas; Csa3; Csm6;
Csx1; Csx3

Yan et al. recently published “Crystal structures of CRISPR-associated Csx3 reveal a manganese-dependent deadenylation exoribonuclease” in RNA Biology.¹ In this research article the authors present compelling evidence that Csx3 possesses manganese dependent RNase activity and also present crystal structures showing that the manganese and RNA binding sites (Fig. 1A) lie at opposite ends of the Csx3 homodimer. In all, their work represents a significant advance in our understanding of Csx3, providing critical new information relevant to its potential roles in CRISPR/Cas.

They also describe the fold of Csx3 (Fig. 1B) as a “ferredoxin-like” fold. At first glance, it is not surprising that Csx3 might also harbor a ferredoxin-like fold. The ferredoxin-like fold (Fig. 1C) is commonly observed in proteins involved in CRISPR-Cas, where it may also be referred to as an RRM (RNA Recognition Motif) or RAMP (Repeat Associated Mysterious Protein) domain. RAMP domains are found in the Cas5, Cas6, Cas7 and Cmr3 families and RAMP-like domains are also found in Cas2 and Cas10.^{2,3} However, the ferredoxin-like fold is defined in the SCOP database⁴ (and elsewhere) as a $\beta\alpha\beta\beta\alpha\beta$ or $(\beta\alpha\beta)_2$ fold, resulting in a 4 stranded antiparallel β sheet arranged as $\beta_2\beta_3\beta_1\beta_4$. The two helices, present as right-handed crossovers between the β strands, lie on the same face of the β -sheet where they run antiparallel to each other (Fig. 1C). In this context, the Csx3 structure does indeed show a β -sheet flanked on one side by 2 α -helices. But in contrast to the ferredoxin-like fold, Csx3 lacks a 4-stranded antiparallel β

sheet, the helices run antiparallel to each other, and the order and connectivity between the β -strands is substantially different (Fig. 1B).

The Csx3 fold can, however, be described as a $\beta\beta\alpha\beta\alpha\beta\beta(\alpha)$ structure resulting in a 6-stranded mixed β -sheet in which the β -strands run sequentially across the sheet from β_1 - β_6 (right to left in Fig. 1B). In addition, as a result of right-handed cross-over connections provided by helices α_1 and α_2 , β_2 - β_4 form a parallel $\beta\alpha\beta\alpha\beta$ super-secondary structure (underlined above) in which the helices run parallel to each other. Notably, this $\beta\alpha\beta\alpha\beta$ sub-structure shows significant similarity to the well-known super-secondary structural motif known as the Rossmann fold that is found in so many nucleotide binding proteins.⁵ In addition, in the RNA bound structure of Csx3 (PDB ID: 3WZI), a third C-terminal α helix is found on the opposite face of the sheet, consistent with the observation that parallel β -sheets and the parallel portions of mixed sheets are frequently flanked on both faces by protecting α helices.⁶

This structural similarity to the Rossmann fold, in turn, suggests that the RNA might be expected to bind at the C-terminal end (bottom arrowheads) of strands β_2 , β_3 and β_4 , and that one of the phosphate groups in the RNA could be accommodated by the helix dipole moment and the unsaturated main-chain NH group at the amino-terminal end of the central α -helix.^{7,8} This prediction is borne out by the structure of Csx3 in complex with RNA, where the phosphate group of nucleotide N2 in subunit A is found at the N-terminal end of helix α_2 , and the

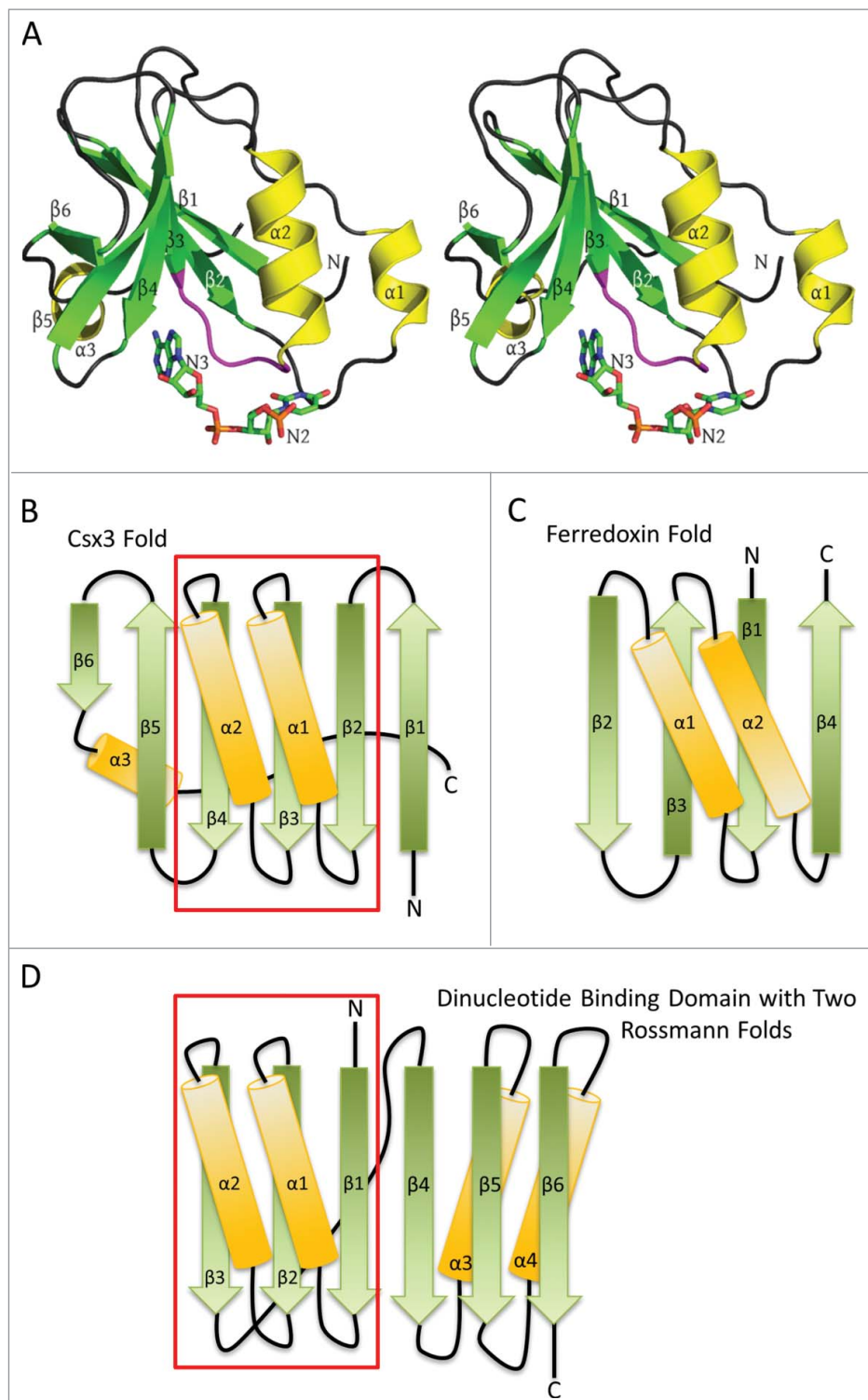


Figure 1. RNA recognition by *AfCxs3*. **A**) Stereo ribbon diagram of *AfCxs3* Chain A bound to a ssRNA (PDB ID: 3WZI). Beta strands, α helices and the conserved $\beta 3$ - $\alpha 2$ loop (Ser44-Ile49) are shown in green, yellow and magenta, respectively. Secondary structures are labeled in order of their occurrence in the amino acid sequence. The orientation is roughly equivalent to that of Fig. 1 in Yan et al.,¹ who modeled nucleotides N2 and N3 into subunit A which are shown as “sticks” and labeled accordingly. **B**) A topology or fold diagram of Csx3. Beta strands and α helices are colored according to panel A. Gradients are added to show the N-to-C directionality of the individual secondary structures, morphing from dark to lighter colors. The boxed region indicates the Rossmann fold present in Csx3. **C**) The similarly depicted RRM or ferredoxin-like fold. This fold is substantially different from the Csx3 fold in panels A and B, where Csx3 lacks the 4-stranded antiparallel β -sheet and the antiparallel helices of the RRM fold. **D**) Fold of the classic dinucleotide binding domain composed of 2 successive Rossmann folds. The first Rossmann fold is contained within the boxed region and is aligned with the Rossmann fold present in Csx3 in order to highlight the similarities between the 2 folds. Helix $\alpha 3$ of Csx3 runs along the back side of the mixed β sheet, contacting strands $\beta 4$, $\beta 3$ and $\beta 2$, but has been slightly repositioned in panel B in order to accentuate the similarities between the Rossmann folds present in the classic dinucleotide binding domain and that in Csx3.

central $\beta 3$ - $\alpha 2$ loop plays a central role in recognizing the 2 nucleotides N2 and N3, with pockets accommodating each of the bases completed by the $\beta 2$ - $\alpha 1$ and $\beta 4$ - $\beta 5$ loops, respectively (Fig. 1A).

A Rossmann fold is also found in CRISPR associated proteins Csx1^{9,10} and Csa3,⁹ where they are present within modified dinucleotide binding domains. Like Csx3, each of these proteins are homodimers, such that the Rossmann-folds in each subunit give rise to a putative ligand binding site that is 2-fold symmetric. Bioinformatics efforts suggest that similar domains are also present in the Csm6 family, and other non-Cas proteins associated with antiviral defense or a more general stress response.¹¹ These “CRISPR Associated Rossmann Folds,” or CARF domains are frequently fused to potential effector domains that can include winged helix-turn-helix (wHTH) DNA binding domains, DNases of the restriction endonuclease fold and RNases of the RelE and HEPN families.^{9,11-13} Collectively, these observations have led to the suggestion that the CRISPR associated Rossmann fold serves to provide ligand binding sites for nucleotides or modified nucleotides that may signal invading nucleic acid, and in cases where they are bound by a homodimer, the ligand may be 2-fold or pseudo-2-fold symmetric.^{9,11}

While Csx3 was not identified in the bioinformatics study as a member of the CARF domain superfamily, the tertiary structure certainly bears some resemblance. Further, at the quaternary level, the presence of the nucleotide binding domains of Csx3 arranged within the context a symmetric homodimer is very reminiscent of the arrangement of the proposed ligand binding domains Csa3 and Csx1.⁹ For these reasons, we suggest that Csx3 is a more distant member of the CARF superfamily. In this context, the structure of Csx3 with bound RNA determined by Yan et al. takes on additional importance, as the structure of a CARF domain protein with bound ligand has not been reported before.

The electron density for the bound RNA in Csx3 is found to span the dimer interface, and because of this, it results in a 2-fold symmetric ring of density that encircles the dimer axis. This density is inconsistent with binding of separate dinucleotides in each of the subunits, and instead indicates covalent linkage between the RNA bound in each of the separate subunits. Given the content of the crystals, Yan et al. quite reasonably chose to interpret this as a bound RNA tetra-nucleotide that follows a C-shaped path as it loops around the 2-fold axis of symmetry in the Csx3 homodimer, in which the stoichiometry is one tetra-nucleotide bound to 2 subunits of Csx3. This ligand is presumably present in either of 2 equivalent orientations throughout the crystal, such that the average of the 2 structures results in the appearance of the closed circular ring of electron density. Yan et al. built one such C-shaped tetra-nucleotide into this density to illustrate how it can adopt this pseudo-2-fold symmetric structure as it binds the Csx3 homodimer. We agree that the interpretation of the electron density by Yan et al. is the most reasonable one. However, we also note that because the electron density for the RNA in the Csx3 crystal structure encircles the dimer axis, it would also be consistent with a cyclic tetra-nucleotide, and the possibility that Csx3 or CARF domain proteins such as Csa3 or Csx1 might bind a cyclic polynucleotide has not been ruled out.

The bioinformatics work of Makarova et al. provides yet another connection between Csx3 and the previously identified proteins harboring the CARF domain. Specifically, they looked for protein families associated with genes encoding CARF domains, and found that genes for Csx3 are enriched in the gene neighborhoods of the CARF superfamily.¹¹ This suggests Csx3 and the previously identified CARF domain proteins may be functionally linked. One potential connection is the shared CARF domain common to each of these proteins.

In conclusion, we believe the structure of Csx3 with bound nucleotides by Yan et al. identifies an additional member of the CARF superfamily, and provides our first look at how symmetric or pseudo-symmetric polynucleotides can be accommodated by the symmetric interface common to CARF domain homodimers. Further, as the number of proteins containing the CARF domain continues to grow, this domain is clearly emerging as a major functional domain in the protein machinery of CRISPR/Cas. It remains to be determined whether the role of these CARF domain proteins is to regulate CRISPR/Cas, or as Makarova et al. suggest, form an additional arm of CRISPR/Cas, or both.^{9,11} However, it is increasingly clear that these domains are of significant importance in CRISPR/Cas, and the observation of Yan et al. that Csx3 is a manganese-dependent deadenylation exoribonuclease capable of binding a pseudo-symmetric RNA tetranucleotide is a significant step toward understanding the roles of these CARF domain proteins in CRISPR/Cas.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by the National Science Foundation, grant MCB-1413534 to CML.

References

1. Yan X, Guo W, Yuan YA. Crystal structures of CRISPR-associated Csx3 reveal a manganese-dependent deadenylation exoribonuclease. *RNA Biol* 2015; 12:749-60; PMID:26106927; <http://dx.doi.org/10.1080/15476286.2015.1051300>
2. Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. *Biochem J* 2013; 453:155-66; PMID:23805973; <http://dx.doi.org/10.1042/BJ20130316>
3. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 2013; 82:237-66; PMID:23495939; <http://dx.doi.org/10.1146/annurev-biochem-072911-172315>
4. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* 2008; 36:D419-25; PMID:18000004; <http://dx.doi.org/10.1093/nar/gkm993>
5. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. *J Mol Biol* 1973; 76:241-56; PMID:4737475; [http://dx.doi.org/10.1016/0022-2836\(73\)90388-4](http://dx.doi.org/10.1016/0022-2836(73)90388-4)
6. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protein Chem* 1981; 34:167-339; PMID:7020376; [http://dx.doi.org/10.1016/S0065-3233\(08\)60520-3](http://dx.doi.org/10.1016/S0065-3233(08)60520-3)

7. Hol WG. Effects of the alpha-helix dipole upon the functioning and structure of proteins and peptides. *Adv Biophys* 1985; 19:133-65; PMID:2424281; [http://dx.doi.org/10.1016/0065-227X\(85\)90053-X](http://dx.doi.org/10.1016/0065-227X(85)90053-X)
8. Wierenga RK, De Maeyer MCH, Hol WGJ. Interaction of pyrophosphate moieties with alpha-helices in dinucleotide binding proteins. *Biochemistry* 1985; 24:1346-57; <http://dx.doi.org/10.1021/bi00327a012>
9. Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copie V, Young MJ, Tainer JA, Lawrence CM. The Structure of the CRISPR-Associated Protein Csa3 Provides Insight into the Regulation of the CRISPR/Cas System. *J Mol Biol* 2011; 405:939-55; PMID:21093452; <http://dx.doi.org/10.1016/j.jmb.2010.11.019>
10. Kim YK, Kim YG, Oh BH. Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of *Pyrococcus furiosus*. *Proteins* 2013; 81:261-70; PMID:22987782; <http://dx.doi.org/10.1002/prot.24183>
11. Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front Genet* 2014; 5:102; PMID:24817877; <http://dx.doi.org/10.3389/fgene.2014.00102>
12. Haft D, Selengut J, Mongodin E, Nelson K. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *Plos Computational Biology* 2005; 1:474-83; <http://dx.doi.org/10.1371/journal.pcbi.0010060>
13. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* 2006; 1:7; PMID:16545108; <http://dx.doi.org/10.1186/1745-6150-1-7>