hydrogen bonding, which would help reduce the chance that the riboswitch would misfire by binding to the wrong metabolite. Perhaps some organisms indeed have evolved to produce natural products that trick Cbl riboswitches, but none have been discovered so far. Also, there is only one superfamily of Cbl riboswitches known, and so this single major solution to the problem of sensing Cbl is quite successful. Therefore, the findings to date suggest that this mostly 'hand-in-glove' binding strategy used by Cbl riboswitches is more than adequate to survive through eons of evolution.

Structures of the more common Cbl-riboswitch type reveal that they form two hydrogen bonds between the adenosyl moiety of AdoCbl and an adenosine of the aptamer J6-3 region (**Fig. 1**, left). These interactions rely on the precise conformation of J6-3, which is assisted by the extensive substructures carried by this aptamer type. These interactions establish ligand specificity for AdoCbl, and therefore representatives of this type are called AdoCbl riboswitches.

Johnson *et al.*[6] also present both biochemical and genetic evidence indicating that the variant riboswitch type favors binding aquocobalamin (AqCbl), wherein the adenosyl moiety of the coenzyme is replaced by water. The authors reason that certain organisms exposed to high light intensities, such as marine bacteria, would have an advantage by sensing the more stable AqCbl derivative. Although the J6-3 sequence contains the same consensus sequence in both riboswitch types, AqCbl riboswitches carry distinct accessory substructures that alter the conformation of the adenosine moiety in this region that otherwise forms hydrogen bonds to AdoCbl. A shift in the location of this adenosine causes a steric clash with Cbl derivatives carrying large substituents (**Fig. 1**, right), and therefore AqCbl is selectively bound by these variants.

With these structures in hand, those interested in developing riboswitch-targeting antibiotics have a form of heads-up display to aid them in designing novel ligands. Also, these structures make it even more enjoyable to speculate about what B_{12}-using ribozymes from the RNA world might have looked like. As noted above, the structures of the binding pockets of AdoCbl and AqCbl riboswitches differ in the J6-3 region, which sits astride the hemisphere of coenzyme B_{12} that is the reactive part of the molecule. Thus, in a hypothetical ribozyme ancestor, substrates for methylation, reduction or various rearrangement reactions promoted by the coenzyme[10] would need to approach the corrin ring through the space currently occupied by J6-3. Interestingly, it is this precise region that can adopt different conformations when selectively binding AdoCbl or AqCbl. Moreover, the large accessory substructures in certain AdoCbl variants are in the vicinity of J6-3, which demonstrates that a great diversity of RNA sequences and shapes could be accommodated near the reactive face of the coenzyme. Could it be that ancient ribozymes used coenzyme B_{12} by forming a binding pocket much like this modern riboswitch? Such ribozymes might even have catalyzed the reduction reaction necessary to convert RNA nucleotides into their DNA counterparts, by holding substrates near the region where J6-3 is today. Other ribozymes containing this coenzyme-binding core might have been formed by swapping out these accessory substructures with others to selectively catalyze coenzyme B_{12}–dependent reactions on other substrates.

It seems likely that all such speculative ribozymes have long since become extinct. Perhaps, however, researchers should keep a lookout for additional strange Cbl-riboswitch variants with novel architectures near the J6-3 region. Such RNAs might function as riboswitches with distinctive ligand recognition capabilities, or perhaps they could represent modern examples of coenzyme B_{12}–dependent ribozymes.

ACKNOWLEDGMENTS

RNA research in the author's laboratory is supported by the US National Institutes of Health and by the Howard Hughes Medical Institute.

COMPETING FINANCIAL INTERESTS

The author declares competing financial interests: details are available at http://www.nature.com/doifinder/10.1038/nsmb.2453.

1. Nahvi, A. *et al. Chem. Biol.* **9**, 1043–1049 (2002).
2. Winkler, W., Nahvi, A. & Breaker, R.R. *Nature* **419**, 952–956 (2002).
3. Mironov, A.S. *et al. Cell* **111**, 747–756 (2002).
4. Winkler, W.C., Cohen-Chalamish, S. & Breaker, R.R. *Proc. Natl. Acad. Sci. USA* **99**, 15908–15913 (2002).
5. Peselis, A. & Serganov, A. *Nat. Struct. Mol. Biol.* **19**, 1182–1184 (2012).
6. Johnson, J.E. Jr., Reyes, F.E., Polaski, J.T. & Batey, R.T. *Nature* published online, doi:10.1038/nature11607 (14 October 2012).
7. Breaker, R.R. *Mol. Cell* **43**, 867–879 (2011).
8. Nahvi, A., Barrick, J.E. & Breaker, R.R. *Nucleic Acids Res.* **32**, 143–150 (2004).
9. Roth, A. *et al. Nat. Struct. Mol. Biol.* **14**, 308–317 (2007).
10. Golding, B.T. & Buckel, W. in *Comprehensive Biological Catalysis* Vol. 3 (ed. Sinott, M.) 239–259 (Academic Press Ltd., 1998).

Intimate liaison with SR proteins brings exon junction complexes to unexpected places

Oliver Mühlemann

Exon junction complexes (EJCs) are deposited on mRNAs during splicing and are key regulators of the post-transcriptional fate of messenger ribonucleoprotein particles (mRNPs). Two recent papers reporting on the transcriptome-wide mapping of EJC-binding sites in human cells reveal an unexpected heterogeneity of EJC distribution on mRNAs and a tight network of EJC–SR protein interactions contributing to the formation of a higher-order, compacted mRNP structure.

EJCs are protein complexes, discovered more than ten years ago, that were found to assemble on mRNAs during the process of splicing[1]. The tetrameric core of the EJC consists of the DEAD-box RNA helicase eIF4AIII, the heterodimer Y14–MAGOH and MLN51 (also known as Barentsz), and it appears to bind RNA in a sequence-independent manner[2–4]. The EJC core interacts dynamically with additional proteins, which reflects its role in diverse cellular functions such as mRNA export, localization, translation and degradation[5,6]. Elegant *in vitro* studies using nuclear extracts and reporter transcripts demonstrated that EJCs are deposited 24 nucleotides (nt) upstream of exon-exon junctions during splicing[1], and very recently, the PRP19-associated protein CWC22 was shown to escort eIF4AIII to the spliceosome and

Oliver Mühlemann is at the Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland.
e-mail: oliver.muehlemann@dcb.unibe.ch

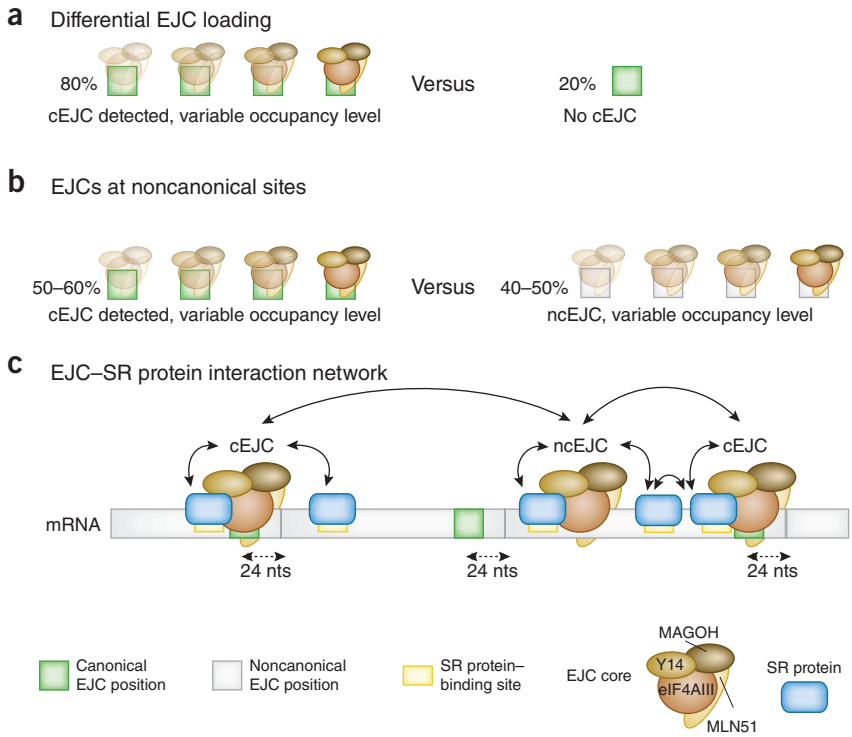

Figure 1 EJCs are differentially loaded onto mRNA, both at the canonical –24-nt positions as well as at noncanonical positions, and form stable megadalton-sized complexes by interacting with SR proteins and with themselves. (**a**) EJCs could be detected at 80% of the canonical EJC binding sites 24 nt upstream of exon-exon junctions (green boxes), whereas 20% of these sites appear to be devoid of EJCs. The variable relative signal strengths (that is, the number of sequencing reads in immunoprecipitates relative to transcript abundance) indicates variable EJC-occupancy frequencies at individual EJC-binding sites, depicted by different color-saturation levels of the EJCs. (**b**) Of all identified EJC-binding sites, 40–50% map to noncanonical positions on mRNA (ncEJCs, white boxes), again with apparently variable occupancy frequencies. (**c**) Binding sites for SR proteins (yellow boxes) correlate with EJC localization and EJCs immunoprecipitated as large, stable, multimeric complexes containing SR proteins (illustrated by arrows), indicating mutual enhancement of RNA binding and a role for SR proteins and EJCs in the formation of compact mRNPs.

promote EJC assembly on the mRNA[7,8]. However, one crucial question remained, namely whether every intron removal event results in the loading of an EJC or whether EJC assembly is a differential, maybe even regulated, process in metazoan cells.

To address this question, Saulière *et al.*[9] used cross-linking and immunoprecipitation coupled to high-throughput sequencing (CLIP-seq) to determine eIF4AIII-binding sites in the HeLa cell transcriptome, whereas Singh *et al.*[10] purified EJCs together with their tightly bound RNA fragments from HEK cells by RNA immunoprecipitation in tandem (RIPiT) and subsequently identified associated proteins by mass spectrometry and RNase-protected RNA fragments by deep sequencing. Even though the two labs used different experimental approaches, their answers are remarkably similar, as their data revealed EJCs at the canonical –24 positions in about 80% of all exon-exon junctions (hereafter called cEJCs). In other words, one in five exon-exon junctions in the human transcriptome appears to lack a cEJC (**Fig. 1a**).

Together with the previously reported differential EJC assembly in *Drosophila melanogaster* cells[11], it becomes evident that EJCs are not just constitutive marks of exon-exon junctions and that their selective assembly on mRNAs might provide yet another opportunity for post-transcriptional gene regulation. Thus, the search for tissue-specific or developmentally regulated examples of differential EJC loading can herewith begin. Because CLIP-seq and the RIPiT methods interrogate RNA populations, and neither yields quantitative results, there is no compelling information about the frequency of EJC occupancy at specific exon-exon junctions. However, given that EJCs are not deposited ubiquitously, it is reasonable to assume that different exon-exon junctions assemble EJCs with varying frequencies and probably span the entire spectrum from 'weak' exon-exon junctions that very rarely lead to loading of a stable EJC, to 'medium' junctions that sometimes acquire an EJC, all the way to 'strong' junctions at which EJCs assemble during essentially every splicing event (**Fig. 1a**,**b**). This assumption is supported by both data sets[9,10]: the reads from the protected RNA fragments in the EJC immunoprecipitations show a great heterogeneity even on highly expressed transcripts with a quite uniform mRNA sequencing (mRNA-seq) read distribution[9].

The discovery of differential EJC occupancy has potentially profound consequences for the cellular processes regulated by EJCs. For example, alterations in EJC occupancy may contribute to the modulation of mRNA translation efficiency. In the case of nonsense-mediated decay (NMD), for which the presence of an EJC >30 nucleotides downstream of the termination codon has been characterized as a strong enhancer[12], differential EJC loading renders the reliable prediction of NMD-targeted transcripts on the basis of sequence information and intron-exon positions virtually impossible. Hence, it could potentially explain some examples of NMD triggered by termination codons in the last exon[13,14]. Furthermore, differential EJC loading may also contribute to the observed variability of NMD efficacy in different cellular contexts[15].

Differential EJC occupancy also raises the question about the determinants for EJC assembly. Bioinformatics analysis revealed no difference in 5′- and 3′-splice-site strengths between EJC-occupied and EJC-free canonical sites, and there was only a slight tendency for cEJCs to occur preferentially on shorter-than-average exons and upstream of longer-than-average introns. In contrast, the only feature that showed a strong inverse correlation with cEJC occupancy was the propensity of the canonical position to form a secondary structure, consistent with a preference of eIF4AIII to bind unstructured RNA[2,3,16].

Most surprisingly, however, only about half (50–60%)[9,10] of all mapped EJCs localized to the canonical positions 24 nt upstream of exon-exon junctions (cEJCs; **Fig. 1b**). The other half were found at different positions throughout the mRNA, mostly in the coding sequence but also present to a low extent in 5′ and 3′ UTRs. The fact that intron-less mRNAs are essentially devoid of these noncanonical EJCs (ncEJCs) indicates that ncEJCs in general also assemble in a splicing-dependent manner on mRNA, although splicing-independent assembly of a small number of EJCs cannot be formally ruled out. A portion of ncEJCs appears to represent

displaced EJCs whose assembly at the canonical position was precluded owing to physical constraints (for example, a secondary structure[16]; discussed above). However, for the majority of ncEJCs, there is no obvious explanation for their unusual positioning. Singh *et al.*[10] report a high correlation between average cEJC signals and ncEJC density across mRNAs, which supports their model of cooperative EJC assembly to form a higher-order mRNP structure (discussed below). Consistent with this correlation, the vast majority of mRNAs analyzed by Saulière *et al.*[9] harbor a mix of both cEJCs and ncEJCs. Strikingly, the search for a consensus-sequence motif in EJC-binding sites revealed that ncEJCs (shown in both studies[9,10]) as well as cEJCs (only detected by Saulière *et al.*[9], who included flanking sequence in their analysis) are preferentially associated with unstructured sequence motifs resembling known binding sites for several serine-arginine–rich proteins (SR proteins), for example GAAGA, C-rich and CG-rich motifs (**Fig. 1c**)[17]. Careful inspection of the CLIP-seq reads showed that the GAAGA motif resides, on average, ~10 nt upstream of the eIF4AIII cross-links, which suggests that EJCs frequently occur in close proximity to SR-binding sites and that EJC assembly and/or stability may be enhanced by SR proteins. Consistent with this view, SR proteins (SRSF1 and SRSF7) co-immunopurified with eIF4AIII from RNase-treated extracts[9].

Mass spectrometry analysis of the EJC proteome corroborated the apparent tight connection between EJCs and SR proteins. Among the ~70 proteins that were more than ten-fold enriched in the RNase-treated anti-Flag immunoprecipitations from cells expressing Flag-eIF4AIII or Flag-MAGOH, the entire SR protein family is prominently represented (except for SRSF5), with several members being stoichiometric (SRSF9, SRSF10 and SRSF11) or even superstoichiometric (SRSF1, SRSF3 and SRSF7) relative to the EJC core factors[10]. Together with the evidence for multimerization of endogenous EJCs[10], this high abundance of SR proteins in the EJC interactome further supports the evidence for the formation of megadalton-sized EJC–SR protein complexes that serve to tightly pack mRNA into compact mRNP structures (discussed below).

In addition to SR proteins, and consistent with EJC deposition occurring concomitantly or after the second step of splicing[18–20], late-stage spliceosome components were also detected. Likewise, the identification of several TREX components underscores the previously reported role of EJCs in mRNA export. Notable absentees in this EJC proteome analysis are the NMD factor UPF3, which interacts with the EJC core[21], and SKAR, which was shown to bind the EJC and recruit activated S6K1 kinase, leading to increased translation efficiency[22].

As alluded to above, one of the main conclusions of the study by Singh *et al.*[10] is that EJCs and SR proteins cooperatively bind mRNA, forming a tight protein-protein and protein-RNA interaction network that promotes mRNA packaging and compaction of the mRNP (**Fig. 1c**). The consistent observation of 30- to 150-nt RNase-resistant RNA fragments after RIPiT that persisted even under extreme nuclease conditions and were more abundant than the short RNA footprints originating from monomeric EJCs was a first indication for EJCs forming large and stable complexes *in vivo*. The strong enrichment of SR protein–binding sites in the long RNA footprints suggests that some of the protected RNA fragments designated as ncEJCs by Singh *et al.*[10] actually represent SR protein– rather than EJC-binding sites (the two studies use the term 'ncEJC' differently, with the ncEJCs identified by Saulière *et al.*[9] exclusively referring to eIF4AIII–cross-linked RNA fragments). The observed co-occurence of cEJCs and ncEJCs on neighboring exons further supports the hypothesis of stable physical interactions between adjacent EJCs. Finally, to investigate the functional role of the EJC–SR protein interactions, Singh *et al.*[10] determined the cross-linking efficiencies of various mRNA-binding proteins to mRNA upon knockdown of eIF4AIII and found that reduced eIF4AIII levels resulted in less cross-linked SRSF1 and SRSF3, whereas other mRNA-binding proteins were not affected (interestingly for NMD aficionados, UPF1 interaction with mRNA was not affected by eIF4AIII depletion). This result indicates that EJC assembly facilitates recruitment and/or stable association of SR proteins with mRNA and vice versa, reflecting functional cooperation between SR proteins and EJCs. Indeed, the tight physical interactions between EJCs and SR proteins can explain numerous functional parallels between SR proteins and EJCs: both EJCs and multiple SR proteins promote nucleo-cytoplasmic mRNA export[23,24], SRSF1 and EJCs both enhance NMD[25,26] and stimulate translation[27–30], and knockdown of SRSF1 or EJC components leads to genomic instability[31,32].

Singh *et al.*[10] speculate that these EJC–SR protein interactions could constitute the major driving force for mRNP compaction, which might be necessary for proper mRNP transport and translation. To date, only little is known about mRNP architecture, but consistent with the proposed model, the Balbiani-ring mRNPs of *Chironomus tentans* salivary glands adopt a highly compacted ring-like structure and contain several SR proteins[33]. Of course, much more work is required to fully elucidate the structure-function relationships in mRNPs, which undoubtedly have key roles in determining the fate of mRNA and therewith in the control of gene expression, but the two papers discussed here provide a solid and valuable foundation for further research in this direction.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Le Hir, H., Izaurralde, E., Maquat, L.E. & Moore, M.J. *EMBO J.* **19**, 6860–6869 (2000).
2. Andersen, C.B. *et al. Science* **313**, 1968–1972 (2006).
3. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. *Cell* **126**, 713–725 (2006).
4. Ballut, L. *et al. Nat. Struct. Mol. Biol.* **12**, 861–869 (2005).
5. Le Hir, H. & Seraphin, B. *Cell* **133**, 213–216 (2008).
6. Tange, T.Ø., Nott, A. & Moore, M.J. *Curr. Opin. Cell Biol.* **16**, 279–284 (2004).
7. Barbosa, I. *et al. Nat. Struct. Mol. Biol.* **19**, 983–990 (2012).
8. Steckelberg, A.L., Boehm, V., Gromadzka, A.M. & Gehring, N.H. *Cell Rep.* **2**, 454–461 (2012).
9. Saulière, J. *et al. Nat. Struct. Mol. Biol.* **19**, 1124–1131 (2012).
10. Singh, G. *et al. Cell.* **151**, 750–764 (2012).
11. Saulière, J. *et al. Nat. Struct. Mol. Biol.* **17**, 1269–1271 (2010).
12. Nicholson, P. *et al. Cell. Mol. Life Sci.* **67**, 677–700 (2010).
13. Bühler, M., Steiner, S., Mohn, F., Paillusson, A. & Mühlemann, O. *Nat. Struct. Mol. Biol.* **13**, 462–464 (2006).
14. Zhang, J., Sun, X., Qian, Y. & Maquat, L.E. *RNA* **4**, 801–815 (1998).
15. Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W. & Kulozik, A.E. *Nucleic Acids Res.* **35**, 4542–4551 (2007).
16. Mishler, D.M., Christ, A.B. & Steitz, J.A. *RNA* **14**, 2657–2670 (2008).
17. Long, J.C. & Caceres, J.F. *Biochem. J.* **417**, 15–27 (2009).
18. Ideue, T., Sasaki, Y.T., Hagiwara, M. & Hirose, T. *Genes Dev.* **21**, 1993–1998 (2007).
19. Reichert, V.L., Le Hir, H., Jurica, M.S. & Moore, M.J. *Genes Dev.* **16**, 2778–2791 (2002).
20. Zhang, Z. & Krainer, A.R. *Proc. Natl. Acad. Sci. USA* **104**, 11574–11579 (2007).
21. Buchwald, G. *et al. Proc. Natl. Acad. Sci. USA* **107**, 10050–10055 (2010).
22. Ma, X.M., Yoon, S.O., Richardson, C.J., Julich, K. & Blenis, J. *Cell* **133**, 303–313 (2008).
23. Huang, Y. & Steitz, J.A. *Mol. Cell* **7**, 899–905 (2001).
24. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M.J. *EMBO J.* **20**, 4987–4997 (2001).
25. Sato, H., Hosoda, N. & Maquat, L.E. *Mol. Cell* **29**, 255–262 (2008).
26. Zhang, Z. & Krainer, A.R. *Mol. Cell* **16**, 597–607 (2004).
27. Gudikote, J.P., Imam, J.S., Garcia, R.F. & Wilkinson, M.F. *Nat. Struct. Mol. Biol.* **12**, 801–809 (2005).
28. Michlewski, G., Sanford, J.R. & Caceres, J.F. *Mol. Cell* **30**, 179–189 (2008).
29. Nott, A., Le Hir, H. & Moore, M.J. *Genes Dev.* **18**, 210–222 (2004).
30. Wiegand, H.L., Lu, S. & Cullen, B.R. *Proc. Natl. Acad. Sci. USA* **100**, 11327–11332 (2003).
31. Li, X. & Manley, J.L. *Cell* **122**, 365–378 (2005).
32. Silver, D.L. *et al. Nat. Neurosci.* **13**, 551–558 (2010).
33. Björk, P. *et al. J. Cell Biol.* **184**, 555–568 (2009).