



Conditional Protein Splicing Switch in Hyperthermophiles through an Intein-Extein Partnership

Christopher W. Lennon,^{a,b} Matthew Stanger,^{a,b} Nilesh K. Banavali,^{c,d} Marlene Belfort^{a,b,c}

^aDepartment of Biological Sciences, University at Albany, Albany, New York, USA

^bRNA Institute, University at Albany, Albany, New York, USA

^cDepartment of Biomedical Sciences, School of Public Health, University at Albany, Albany, New York, USA

^dDivision of Genetics, Laboratory of Computational and Structural Biology, Wadsworth Center, NYS Department of Health, Albany, New York, USA

ABSTRACT Inteins are intervening proteins that undergo an autocatalytic splicing reaction that ligates flanking host protein sequences termed exteins. Some intein-containing proteins have evolved to couple splicing to environmental signals; this represents a new form of posttranslational regulation. Of particular interest is RadA from the archaeon *Pyrococcus horikoshii*, for which long-range intein-extein interactions block splicing, requiring temperature and single-stranded DNA (ssDNA) substrate to splice rapidly and accurately. Here, we report that splicing of the intein-containing RadA from another archaeon, *Thermococcus sibiricus*, is activated by significantly lower temperatures than is *P. horikoshii* RadA, consistent with differences in their growth environments. Investigation into variations between *T. sibiricus* and *P. horikoshii* RadA inteins led to the discovery that a nonconserved region (NCR) of the intein, a flexible loop where a homing endonuclease previously resided, is critical to splicing. Deletion of the NCR leads to a substantial loss in the rate and accuracy of *P. horikoshii* RadA splicing only within native exteins. The influence of the NCR deletion can be largely overcome by ssDNA, demonstrating that the splicing-competent conformation can be achieved. We present a model whereby the NCR is a flexible hinge which acts as a switch by controlling distant intein-extein interactions that inhibit active site assembly. These results speak to the repurposing of the vestigial endonuclease loop to control an intein-extein partnership, which ultimately allows exquisite adaptation of protein splicing upon changes in the environment.

IMPORTANCE Inteins are mobile genetic elements that interrupt coding sequences (exteins) and are removed by protein splicing. They are abundant elements in microbes, and recent work has demonstrated that protein splicing can be controlled by environmental cues, including the substrate of the intein-containing protein. Here, we describe an intein-extein collaboration that controls temperature-induced splicing of RadA from two archaea and how variation in this intein-extein partnership results in fine-tuning of splicing to closely match the environment. Specifically, we found that a small sequence difference between the two inteins, a flexible loop that likely once housed a homing endonuclease used for intein mobility, acts as a switch to control intein-extein interactions that block splicing. Our results argue strongly that some inteins have evolved away from a purely parasitic lifestyle to control the activity of host proteins, representing a new form of posttranslational regulation that is potentially widespread in the microbial world.

KEYWORDS Posttranslational regulation, recombinase, hyperthermophile, mobile genetic elements

Received 13 December 2017 **Accepted** 18 December 2017 **Published** 30 January 2018

Citation Lennon CW, Stanger M, Banavali NK, Belfort M. 2018. Conditional protein splicing switch in hyperthermophiles through an intein-extein partnership. *mBio* 9:e02304-17. <https://doi.org/10.1128/mBio.02304-17>.

Editor Jeff F. Miller, UCLA School of Medicine

Copyright © 2018 Lennon et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marlene Belfort, mbelfort@albany.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Uri Gophna, Tel Aviv University; Julio Camarero, University of Southern California.

Inteins, or **intervening proteins**, are mobile genetic elements that are translated within host proteins and removed through protein splicing (1–3). In this process, the intein autocatalytically removes itself by breaking two peptide bonds and joining the flanking sequences (called N- and C-exteins) via a peptide bond, leaving the host protein (ligated exteins) without any trace of where the intein once resided (Fig. 1A).

Inteins are widespread in nature, and they are found in all three domains of life as well as in phages and viruses (4, 5). Inteins are particularly abundant in bacteria and archaea, as they are present in nearly one-quarter and one-half of reference sequences, respectively (4). Not only are inteins widespread, but also they often are found within proteins important or essential to organism fitness. Inteins cluster to particular functional classes of proteins, with over 70% found in ATP-binding proteins and almost two-thirds found in proteins involved in DNA replication, recombination, and repair (RRR). Remarkably, inteins have even been independently acquired by evolutionarily distinct bacterial and archaeal proteins with equivalent functions (4). Inteins are generally found in sites critical to host protein function, such as in active sites or protein-protein interaction interfaces.

Inteins have long been considered parasitic elements that provide no benefit to the host organism (6). Recent work has refuted this notion by demonstrating that the splicing of several inteins can be regulated by a variety of environmental cues relevant to the survival of the host organisms or crucial to host protein function (7–13). These conditions include pH, redox, reactive oxygen-nitrogen species, salt, temperature, and even host protein substrate. This environmental control suggests that in some cases, inteins have evolved to benefit the host through posttranslational regulation of protein function. Furthermore, bioengineers can develop inteins that will splice in response to a variety of stimuli to control ligation or cleavage reactions (14–16), begging the question of how nature may have employed similar tactics as a mechanism of post-translational control in intein-containing proteins.

Our group has discovered a landmark example of conditional protein splicing in the RadA recombinase from the deep sea hyperthermophilic archaeon *Pyrococcus horikoshii*, a member of the highly conserved RecA/Rad51 homologous recombinase family (17). What makes this example striking is that splicing in native exteins requires high temperatures (~75°C) to proceed, while splicing in foreign exteins occurs rapidly at low temperature (~25°C) (9). Based on structure modeling and mutagenesis, we determined that splicing inhibition was due to intein-C-extein interactions between residues well-separated in sequence that must be broken for splicing to proceed (9). Our findings on *P. horikoshii* RadA became even more remarkable when we discovered that splicing can be activated by the substrate of the host protein, single-stranded DNA (ssDNA). In the first example of such regulation, ssDNA, the cell's signal that RadA recombinase activity is needed, dramatically accelerates and improves the accuracy of *P. horikoshii* RadA splicing at all temperatures tested (13).

Nature has provided an interesting ortholog of *P. horikoshii* RadA in *Thermococcus sibiricus*. *T. sibiricus* is an archaeon that is similar in lifestyle to *P. horikoshii* but, whereas *P. horikoshii* grows between 70 and 102°C (18), *T. sibiricus* survives at significantly lower temperatures, between 40 and 88°C (19). While the difference in growth temperatures between these two archaea is substantial, *P. horikoshii* and *T. sibiricus* RadA proteins are highly similar at the amino acid level, including an identical intein insertion site within the P-loop of the RadA ATPase domain. Both *P. horikoshii* and *T. sibiricus* RadA naturally possess mini-inteins that lack the homing endonuclease domain, which is often present and presumed to be required for spread into non-intein-containing alleles through homing (4, 5, 20).

Here, we compare the splicing of intein-containing RadA proteins from *T. sibiricus* and *P. horikoshii*. We found that, whereas splicing is accelerated at a high temperature only in the native exteins for both RadA inteins, the temperature required is different and corresponds closely to those where the organisms are found in nature. Through analysis of a region poorly conserved between the *T. sibiricus* and *P. horikoshii* inteins, we describe the first substitution that disrupts splicing in native, but not foreign, extein

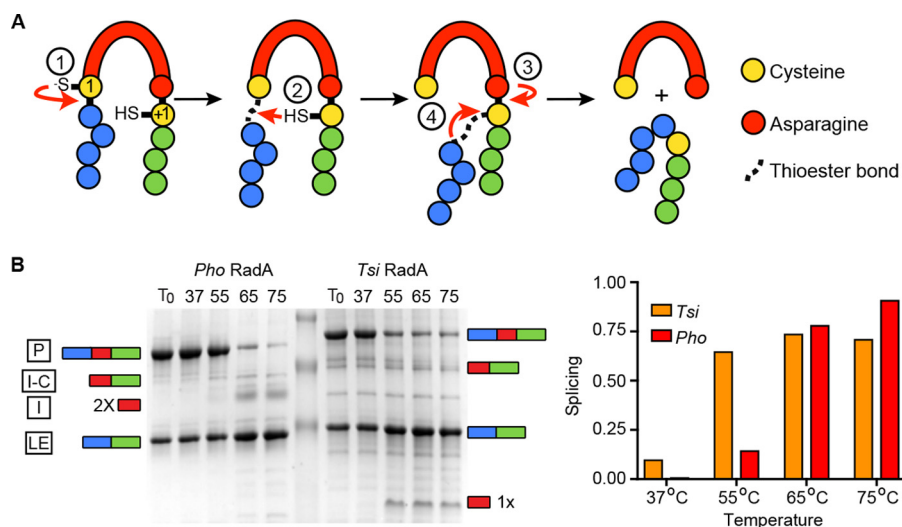


FIG 1 Splicing mechanism and temperature requirements of *P. horikoshii* and *T. sibericus* RadA proteins for splicing in native exteins. (A) Schematic of class I protein splicing mechanism, where both catalytic nucleophiles are cysteines. Details of each reaction step are described in the text. (B) *P. horikoshii* and *T. sibericus* RadA splicing requires an elevated temperature in native exteins. (Left) Representative data are shown for samples incubated for 30 min at the indicated temperatures. Samples were separated by SDS-PAGE and stained with Coomassie, and levels of each species were determined using densitometry. A cartoon representation of on-pathway intein splicing with native exteins to the side shows precursor (P), ligated exteins (LE), and intein (I) and off-pathway intein-C-extein product (I-C). The *P. horikoshii* RadA intein runs as a dimer (13), whereas the *T. sibericus* RadA intein resolves as a monomer. 2× indicates an intein dimer, and 1× indicates an intein monomer. (Right) Bar graph showing quantification of the samples shown in the schematic on the left.

contexts, and this defect can be rescued by substrate ssDNA. We propose that this region, a flexible loop that previously housed a homing endonuclease and that is located between conserved splicing domains, behaves as a hinge that helps disrupt an inhibitory intein-extein partnership, effectively acting as a switch to turn on splicing.

RESULTS

***P. horikoshii* and *T. sibericus* RadA splicing is differentially regulated by temperature.** The canonical class I intein splicing mechanism proceeds in four steps (Fig. 1A). In the first step, the first residue of the intein (known as the 1 position; either a cysteine or serine) performs a nucleophilic attack on the preceding peptide bond, forming a (thio)ester linkage. Next, the first residue of the C-extein (known as the +1 position; either a cysteine, serine, or threonine) performs a nucleophilic attack on the (thio)ester resulting from step 1 and forms a branched intermediate. Third, an asparagine found at the end of the intein cyclizes to liberate the intein. Finally, the (thio)ester linking the N- and C-exteins rearranges to form a peptide bond. Off-pathway cleavage reactions can occur as well, in which either the N- or C-extein is released prior to extein ligation, resulting in products of no known utility to the cell.

Previously we found that when we removed the first 115 residues corresponding to the N-terminal domain (NTD) of *P. horikoshii* RadA, leaving the ATPase and intein intact, splicing acceleration by temperature and ssDNA was unaffected (13). We therefore used Δ NTD constructs of *P. horikoshii* and *T. sibericus* RadA that were expressed and purified from *Escherichia coli* whenever splicing was examined in the native extein context; we refer to these constructs as wild type. First, we examined whether *T. sibericus* RadA intein splicing was stimulated by increasing temperature in the context of native exteins, as is the case for *P. horikoshii* RadA (9), as this would indicate conservation of this mode of regulation in nature. Following overexpression and His tag purification of *T. sibericus* and *P. horikoshii* RadA under identical conditions, a significant amount of unspliced RadA precursor was isolated in both cases. We next incubated purified *T. sibericus* and *P. horikoshii* RadA proteins at temperatures ranging from 37°C to 75°C and

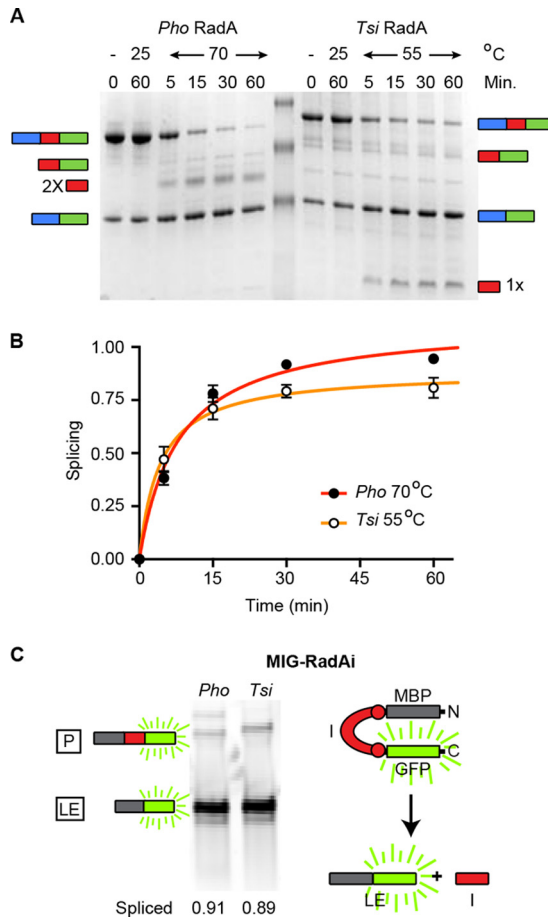


FIG 2 *P. horikoshii* and *T. sibericus* RadA splicing at different temperatures. (A) Splicing time course of *T. sibericus* RadA at 55°C and *P. horikoshii* RadA at 70°C at the indicated times. Additionally, each protein was incubated at 25°C for 1 h. Samples in panels A and B were separated, analyzed, and identified as described for Fig. 1. (B) Scatter plot of splicing results shown in panel A. Points are averages of three independent experiments, with error bars showing standard deviations. Error bars smaller than the size of the point are not shown. A small portion of the *T. sibericus* RadA precursor does not splice, even after prolonged incubation at an elevated temperature, which is likely a result of a catalytically inactive alternative conformation following purification. The initial rates of splicing are indicated in the text. (C) *P. horikoshii* and *T. sibericus* RadA splicing in the foreign MIG extein context. Samples were incubated at 25°C for 1 h following expression. Included is a cartoon representation of the MIG reporter. Samples were separated by SDS-PAGE, and in-gel GFP fluorescence was measured.

measured splicing after 30 min. Whereas *T. sibericus* RadA splicing was also activated by increasing temperatures, the initiation of splicing occurred at a lower temperature than for *P. horikoshii* RadA (Fig. 1B). (Throughout figures *P. horikoshii* is referred to as Pho and *T. sibericus* is referred to as Tsi.)

To establish the precise difference in temperature of splicing between *T. sibericus* and *P. horikoshii* RadA proteins, we determined the temperatures at which the rates of splicing were approximately the same for the two proteins, which was 70°C and 55°C for *P. horikoshii* RadA and *T. sibericus* RadA, respectively. We thus incubated *P. horikoshii* RadA at 70°C and *T. sibericus* RadA at 55°C and measured splicing at various times over 1 h (Fig. 2A). We also incubated each protein at 25°C for the duration of the time course as a control to ensure that temperature was responsible for activation of splicing. The initial rates of splicing were approximately equal under these conditions ($7.7 \times 10^{-2}/\text{min}$ for *P. horikoshii* and $9.4 \times 10^{-2}/\text{min}$ for *T. sibericus*) (Fig. 2B), demonstrating at least a 15°C shift in splicing activation temperature. This 15°C difference agrees closely with the optimal (20°C) and maximal (14°C) growth temperature difference between *T. sibericus* and *P. horikoshii* (18, 19).

***T. sibiricus* RadA does not require elevated temperature for splicing in foreign exteins.** We next asked if increased temperature was needed to stimulate *T. sibiricus* RadA splicing because of something inherent to the intein itself, or whether this was dependent on its being in the native extein context, as is the case for *P. horikoshii* RadA (9). We used our foreign extein splicing reporter, MIG (MBP-intein-GFP), which contains the maltose binding protein (MBP) followed by the intein (flanked on each side by 10 native extein residues), followed by green fluorescent protein (GFP). After expression in *E. coli*, *T. sibiricus* and *P. horikoshii* RadA inteins in the MIG reporter were reincubated for 1 h at 25°C, and the precursor and products were measured based on in-gel fluorescence of C-extein GFP. Both inteins spliced rapidly and accurately without incubation at higher temperatures (Fig. 2C), indicating that splicing can proceed at low temperatures but that processing is blocked in the native extein context.

Tuning of thermoregulated splicing by intein–C-extein interactions. We posit that the difference between the temperature required to initiate splicing of *P. horikoshii* and that for *T. sibiricus* RadA in native exteins (Fig. 1B and 2A) results from variable intein–C-extein interactions. We previously assessed these interactions by modeling the *P. horikoshii* RadA precursor protein, based on available structures, followed by extensive confirmatory mutagenesis (9). Several residues in the C-extein at the predicted intein interface were replaced with alanine, resulting in disruption of thermoregulated splicing (9). One of these residues of the *P. horikoshii* RadA C-extein, E364, was predicted to interact with the H312 side-chain of the intein (Fig. 3A) (9). Interestingly, the residue corresponding to *P. horikoshii* RadA E364 in *T. sibiricus* RadA is Q361, which is similar in size but lacks a side-chain negative charge. Following modeling of the *T. sibiricus* RadA precursor, there was no predicted side chain interaction between corresponding *T. sibiricus* RadA residues (H309 and Q361 average, -0.1 kcal/mol), compared to similarly minimized *P. horikoshii* RadA precursor models (H312 and E364 average, -2.7 kcal/mol; see Materials and Methods for construction of RadA precursor models). When examining the intein-extein side-chain interactions with energies greater than -1 kcal/mol, only this predicted interaction was variable between *T. sibiricus* and *P. horikoshii* RadA (Fig. 3A). We previously replaced *P. horikoshii* RadA E364 with alanine, measured splicing in response to temperature, and observed that this mutation allowed splicing to proceed ~ 5 -fold faster than with the wild type at 55°C (9).

***P. horikoshii* and *T. sibiricus* RadA are divergent in the endonuclease loop of the intein.** While the E364A substitution led to a substantial increase in *P. horikoshii* RadA splicing, it was slower than *T. sibiricus* RadA at 55°C (9), indicating that C-extein variation between *P. horikoshii* and *T. sibiricus* RadA alone may not fully explain the observed difference in thermoregulated splicing. We hypothesize that something inherent to the intein must also play a role in temperature-induced splicing, possibly by controlling intein-extein interactions. We therefore focused on sequence differences to better understand the observed variation in thermoregulated splicing. One region within the inteins of *P. horikoshii* and *T. sibiricus* RadA stands out as highly variable, both in sequence and length, and we call this the nonconserved region (residues 273 to 283 of *P. horikoshii* RadA and residues 274 to 280 of *T. sibiricus* RadA). Both *P. horikoshii* and *T. sibiricus* RadA house mini-inteins, and the NCR is located between conserved splicing blocks, where the homing endonuclease is located in full-length inteins. Using our previously generated model of the *P. horikoshii* RadA precursor (9), we found that the NCR is distant from both catalytic intein splicing residues and the intein–C-extein interaction interface (Fig. 3B). While there is considerable sequence divergence in the NCR, *T. sibiricus* and *P. horikoshii* RadA as a whole are highly similar. Comparison of the amino acid sequences of the *T. sibiricus* and *P. horikoshii* RadA precursors, including ATPase domains and inteins, shows that they are 75.8% identical and 92.7% similar based on Clustal Omega analysis (21) (Fig. 3C). Not surprisingly, given the high level of conservation between the recombinases RadA, eukaryotic Rad51, and bacterial RecA (22), the extein sequence is more conserved than is the intein. Extein

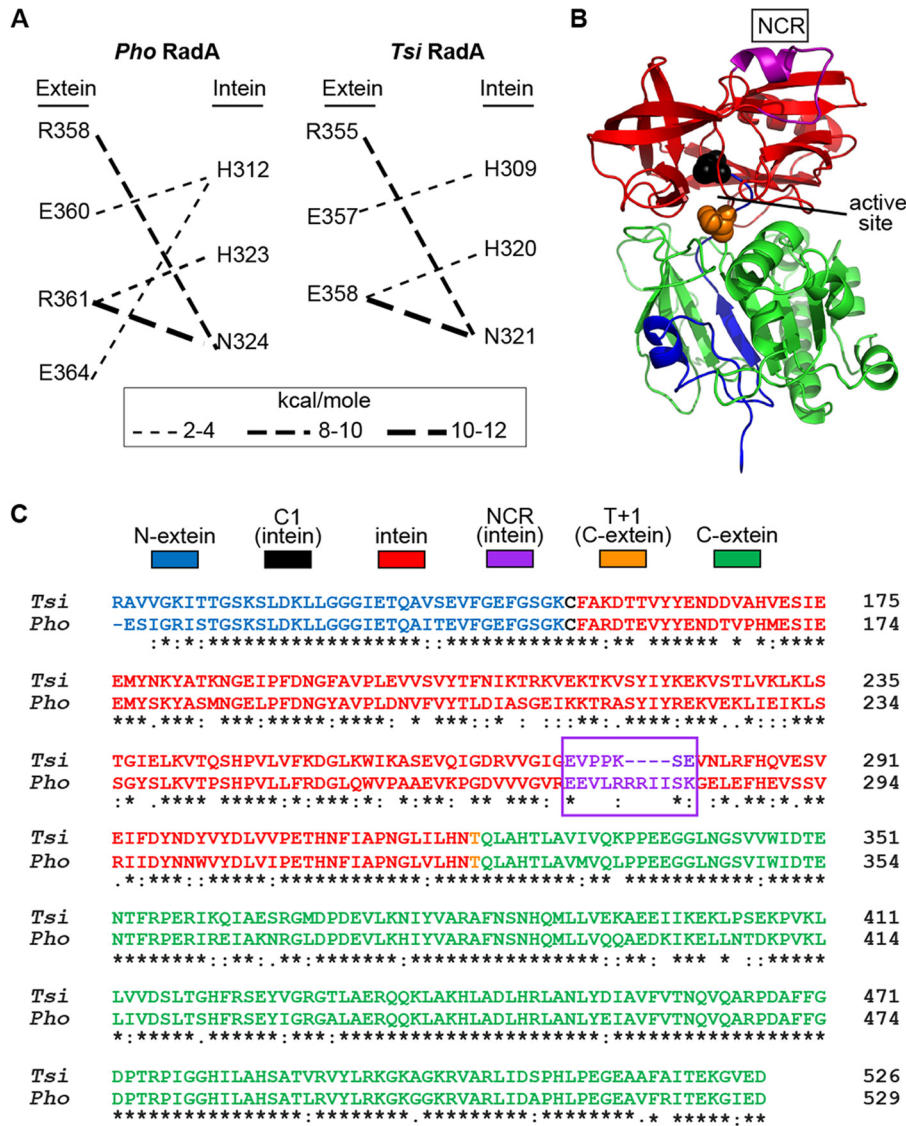


FIG 3 Predictions for intein-extein interactions, model of the *P. horikoshii* RadA precursor, and alignment of the *P. horikoshii* and *T. sibericus* RadA proteins. (A) Predicted intein-extein side chain interactions greater than 1 kcal/mol. A E364-H312 side chain interaction is predicted in the *P. horikoshii* RadA precursor models, but not between the corresponding Q361-H309 of *T. sibericus* RadA. (B) Molecular model of the *P. horikoshii* RadA precursor residues 116 to 529. Pymol was used to generate the image from a previously rendered model (9). The sequence and model are colored according to the color key of panel C. (C) Residues 116 to 529 of *P. horikoshii* RadA and residues 116 to 526 of *T. sibericus* RadA were aligned using Clustal Omega. An asterisk indicates absolute conservation, a colon indicates strong conservation, a period indicates weak conservation, and an empty space indicates no conservation.

residues are 85.5% identical and 97.1% similar, whereas the inteिन residues are 69.8% identical and 84.3% similar.

The *P. horikoshii* inteिन NCR is dispensable for splicing in foreign exteिनs but critical in native exteिनs. The structure of the *P. horikoshii* RadA inteिन was previously solved both by both X-ray crystallography and nuclear magnetic resonance in the absence of native exteिनs (23). Interestingly, residues 272 to 285 of *P. horikoshii* RadA were disordered (residues 120 to 133 of the inteिन), prompting those authors to delete inteिन residues 273 to 282, mutate residue 283 from lysine to asparagine, and resolve the crystal structure of this inteिन. The structure of the minimized *P. horikoshii* RadA inteिन was very similar to that of the full-length inteिन, but it lacked the disordered loop and allowed the inteिन active site to be visualized at improved resolution (23). Splicing

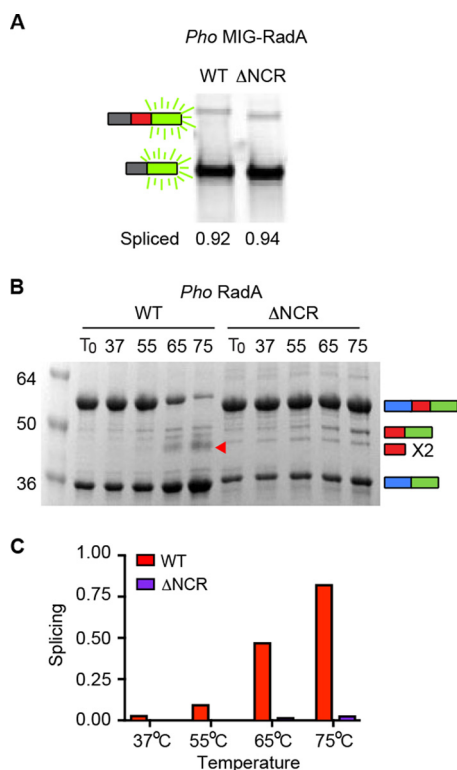


FIG 4 The *P. horikoshii* RadA NCR is only critical to splicing in native exteins. (A) *P. horikoshii* wild-type and Δ NCR RadA protein splicing in the foreign MIG extein context are indistinguishable. Samples were incubated for 1 h at 25°C following expression. (B) *P. horikoshii* RadA Δ NCR splicing is defective in the native extein context. Samples were incubated for 30 min at the temperatures indicated. Samples in panels A and B are representative examples and were separated, analyzed, and identified as described for Fig. 1. The red arrow indicates free intein, observed only in WT splicing. (C) Bar graph illustrating quantification of samples of panel B.

of the minimized intein was indistinguishable from the wild type in a foreign extein context (different from our MIG splicing reporter), prompting those authors to conclude that this region was dispensable for splicing and simply the remnant of the once-present homing endonuclease (23).

To determine if the NCR influenced splicing in our foreign extein MIG reporter, we deleted residues 273 to 283 of the *P. horikoshii* RadA within the intein (Δ NCR, residues 121 to 131 of the intein) and compared its splicing to the wild-type *P. horikoshii* RadA intein based on in-gel fluorescence from the C-extein-GFP (9). After expression and incubation at 25°C for 1 h, both wild-type and the Δ NCR protein were >90% spliced in MIG (Fig. 4A). This result fit perfectly with previous findings on splicing in the foreign extein context (23).

To investigate any possible role of the NCR in the native extein context, we deleted this region of the *P. horikoshii* RadA intein within the native exteins and incubated the *P. horikoshii* RadA wild type and *P. horikoshii* RadA Δ NCR protein at temperatures ranging from 37°C to 75°C for 30 min. Strikingly, splicing of the *P. horikoshii* RadA Δ NCR protein was strongly inhibited at all temperatures tested, even at 75°C, at which nearly all *P. horikoshii* RadA wild-type was spliced after 30 min (Fig. 4B and C). Even over a 3-h time course at 75°C, little accurate splicing was observed for the *P. horikoshii* RadA Δ NCR protein; rather, off-pathway N-terminal cleavage and disulfide bond formation between precursors at C1 were observed (see Fig. S1 and S2 in the supplemental material). Together, these results demonstrated the first example of an intein mutant that was conditionally defective for splicing in native but not foreign exteins.

Splicing of the *P. horikoshii* RadA intein NCR deletion is rescued by ssDNA. As described above, a natural RadA substrate, ssDNA, accelerates wild-type *P. horikoshii*

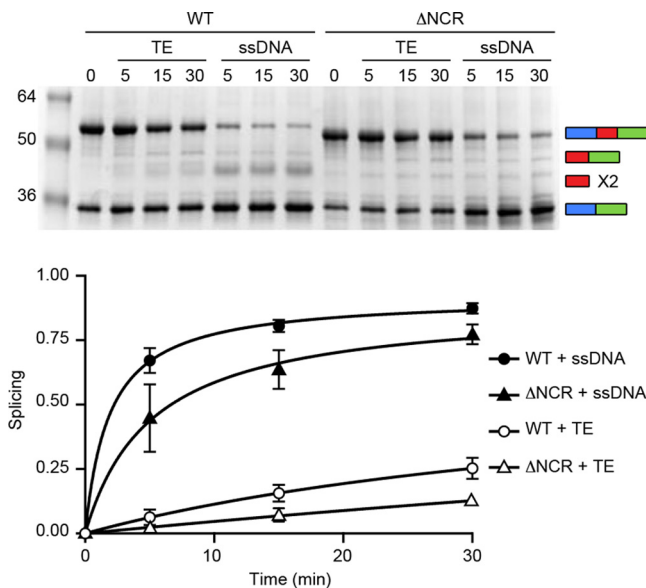


FIG 5 The *P. horikoshii* RadA Δ NCR splicing defect can be rescued by an ssDNA substrate. Samples were incubated at 63°C for the indicated times. (Top) Gel; (bottom) corresponding scatter plot. While splicing is largely on-pathway for the *P. horikoshii* RadA Δ NCR in the presence of ssDNA, as indicated by the increase in ligated exteins, we did not observe free intein dimer, which we attribute to poor resolution on the SDS-PAGE. Each point is the average of three independent experiments, with standard deviations shown by the error bars. Error bars smaller than the size of the point are not shown. Samples were separated, analyzed, and identified as described for Fig. 1.

RadA splicing by nearly 50-fold, as well as dramatically improving the accuracy of splicing (13). We therefore asked if ssDNA could alleviate the splicing defect caused by the NCR deletion in *P. horikoshii* RadA. At 63°C, both wild-type and the Δ NCR *P. horikoshii* RadA spliced little over a 30-min time course in the absence of ssDNA, especially in the case of the Δ NCR *P. horikoshii* RadA protein, for which virtually no new ligated exteins were generated. Strikingly, the splicing rate and accuracy were largely restored by addition of substrate ssDNA (Fig. 5), demonstrating that the Δ NCR *P. horikoshii* RadA can achieve a catalytically competent conformation. Furthermore, ssDNA binding is sufficient to overcome NCR control of thermoregulated splicing in the native extein context, indicating that this mutant is specifically defective for temperature-activated splicing.

DISCUSSION

Splicing regulation of intein-containing proteins provides an attractive posttranslational mechanism to inhibit protein activity until conditions are ideal for function. Here, we have shown that while both *P. horikoshii* and *T. sibiricus* RadAs require an elevated temperature to splice in the native extein context, the temperature needed corresponds closely to ambient temperatures in the natural environments of *T. sibiricus* and *P. horikoshii*. We propose that the tailoring of splicing to a particular temperature is regulated by a variable intein-extein partnership. This natural example of conditional protein splicing, where formation of a functional host protein is exquisitely tuned to the evolutionary niche of the organism, demonstrates the power of inteins to impart regulation on protein activity to the potential benefit of the host organism. However, it is possible that conditional protein splicing evolved as a mechanism to minimize damage from nonspecific and off-pathway splicing. In either case, conditional protein splicing could be beneficial to the host.

Given the position of the NCR remote from the active site in what is a vestigial endonuclease loop in the structure of the intein (Fig. 3B), it is likely that this region influences the activity of distant residues. Structural work has shown that the *P. horikoshii* RadA intein NCR is highly flexible (23), leading us to posit that the NCR

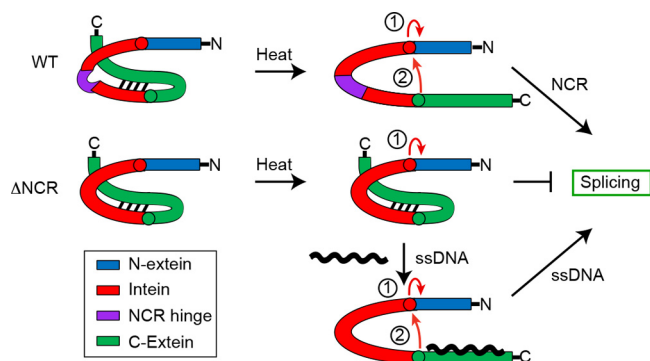


FIG 6 Model describing a role for the intein NCR in control of intein-extein interactions. Color coding of intein, N-extein, and C-extein is indicated in the key. Interactions between the intein and C-extein block splicing of the wild-type and Δ NCR RadA at low temperatures are shown. High temperatures can break these interactions in a NCR-dependent manner, whereby the NCR acts as a flexible hinge to switch on splicing. In contrast, Δ NCR maintains inhibitory intein-C-extein interactions. ssDNA, which accelerates and improves the accuracy of *P. horikoshii* RadA splicing, compensates for the absence of the NCR, interacting with the C-extein to break interactions with the intein, and promoting the catalytically competent state. The first two steps of the splicing reaction are indicated by red arrows (described in the text and the legend for Fig. 1A). We propose that step 2 of the reaction is blocked in the Δ NCR RadA protein in the absence of ssDNA.

functions to allow flexibility between conserved splicing domains, much like a hinge, to promote splicing. In support of this role for the NCR, computational studies with the *Mycobacterium tuberculosis* RecA intein found that the flexibility in the endonuclease loop region is higher in the active vs. inactive intein and that linker length influences dynamics (24).

To explain why the *P. horikoshii* RadA NCR is needed for splicing in the native extein context but not in two foreign extein reporters (Fig. 4A) (23), we considered what is unique to splicing in native exteins, namely, an interaction between the intein and C-extein that must be broken to allow for splicing to proceed (9). In the absence of the NCR, we propose that this interaction is stable even at an increased temperature due to the lack of intein flexibility conferred by the NCR, blocking the availability of other intein residues for catalysis (Fig. 6) (9). Specifically, key residues at the C-terminal splice site would be unavailable due to proximal interactions with the C-extein (Fig. 3A and 6) (9). This model is supported by our observation that the Δ NCR *P. horikoshii* RadA C-1 is still highly reactive, resulting in increased off-pathway splicing compared to the wild type, namely, N-terminal cleavage (Fig. 4B; Fig. S1) and disulfide bond formation at C-1 between adjacent RadA precursors (Fig. S2). In the presence of ssDNA, C-extein binding to the substrate overrides the intein-C-extein interactions and releases the intein to achieve a catalytically competent state (Fig. 6). The slow, off-pathway splicing is also alleviated (Fig. 5; Fig. S1 and S2). Interestingly, we saw some evidence of *P. horikoshii* RadA Δ NCR splicing to form ligated exteins following expression in *E. coli* (Fig. 4 and 5; Fig. S1 and S2), in contrast to our observations *in vitro* in the absence of ssDNA, suggesting the presence of elements in the cellular environment, possibly the low levels of ssDNA when there is fast growth, that can promote splicing.

We thus demonstrated that the NCR of the *P. horikoshii* RadA intein is crucial to the rate and accuracy of splicing in native exteins but is dispensable in the foreign extein context. To our knowledge, this is the first such intein-extein connection described. This finding is of importance beyond the implications for RadA splicing regulation and demonstrates the close relationship between inteins and their native exteins, supporting a complex coevolutionary scenario which can serve to control splicing in response to environmental signals (25, 26).

This is not the first work to demonstrate that intein substitutions in the endonuclease loop can have a profound effect on splicing despite the distance from the active site. In a selection effort to generate hyperactive variants of the RecA intein from *M.*

tuberculosis, our group previously found that the most common site of mutation was in this loop, and we hypothesized that substitutions led to a “ripple effect” on the environment of the catalytic intein residues (27). The endonuclease loop has also been used as a site for insertion of foreign proteins that can be engineered to splice in response to conformational changes triggered by small-molecule binding (15, 16). Further support was recently garnered by Iwai and colleagues, who demonstrated via a combination of structural and biochemical techniques the importance of endonuclease loop dynamics for the splicing of the TFIIB inteins from *Methanococcus jannaschii* and *Methanocaldococcus vulcanius* M7 (28).

Beyond the mechanistic implications of *P. horikoshii* RadA Δ NCR splicing rescue by ssDNA, these findings could be useful in biotechnology. Virtually no on-pathway splicing of the *P. horikoshii* RadA Δ NCR occurred in the absence of substrate ssDNA *in vitro* at all temperatures tested (Fig. 4B and 5; Fig. S2), but splicing is rapid and accurate when incubated in the presence of ssDNA (Fig. 5). This system could be engineered into an on/off splicing switch or sensor to generate ligated exetins tightly controlled by the addition of ssDNA or DNA damage.

Because the NCR is outside the highly conserved blocks of the intein required for splicing, there exists greater plasticity for adaptation within this region for fine-tuning of splicing in a specific extein context while maintaining conservation of the splicing blocks critical to catalysis. Thus, an intein region that was considered to be irrelevant due to lack of conservation, that was located in the truncated endonuclease region not required for splicing, has been repurposed via evolution to act as a switch to promote splicing in the native extein context. Our results as a whole demonstrate the contribution of both intein-extein interactions and a nonconserved intein region to conditionally control the progression of splicing in response to environmental cues in native exteins. These findings speak strongly to the complex evolutionary relationship between these mobile elements and the host organisms they invade, and our findings also demonstrate the potential for nature to utilize protein splicing as a means of posttranslational regulation.

MATERIALS AND METHODS

Plasmid construction. *T. sibericus* RadA (Δ 1-115 codons) was PCR amplified using primers IDT4699 (5'-GGTGGTGGATCCGAGGGCGGTGGTGGGAAG-3') and IDT4670 (5'-GGTGGTCTCGAGTCAATCTTCGACACCTTCTCCG-3') from *T. sibericus* genomic DNA (DSM 12597; Leibniz-Institut DSMZ) and cloned into pET45b by using BamHI/XhoI restriction sites in-frame with an N-terminal His tag. Plasmid pACYC-*T. sibericus* MIG-RadAi was constructed by amplifying the intein sequence plus 10 flanking codons on each side (codons 144 to 331 of full-length *T. sibericus* RadA) and then cloned into SphI/Clal sites from the previously constructed pACYC-*M. tuberculosis* MIG-SufBi (10). The *P. horikoshii* NCR (codons 273 to 283 of full-length *P. horikoshii* RadA) was deleted from pET45b-*P. horikoshii* RadA and *P. horikoshii* MIG-RadAi by using Q5 mutagenesis (New England Biolabs) with primers IDT5040 (5'-CCTAACACCAACACAAC-3') and IDT5041 (5'-GAAGAACTTGAATCCATG-3'). Plasmid pET45b-*P. horikoshii* RadA (Δ 1-115 codons) and pACYC-*P. horikoshii* MIG-RadAi were described previously (9, 13).

Protein expression and purification. *P. horikoshii* RadA, *P. horikoshii* RadA Δ NCR, and *T. sibericus* RadA were expressed from pET45b constructs in BL21(DE3) cells (Novagen). Cells were grown to mid-log phase (optical density at 600 nm of ~0.5) at 37°C and then switched to 16°C, and protein expression was induced by addition of 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for ~16 h. Proteins were purified by immobilized metal affinity chromatography as previously described (13). MIG-*P. horikoshii* RadAi, MIG-*P. horikoshii* RadAi Δ NCR, and MIG-*T. sibericus* RadAi were expressed from pACYC constructs in BL21(DE3) cells. Cells were grown to mid-log phase at 37°C, then switched to 30°C, and protein expression was induced for 1 h by addition of 0.5 mM IPTG. Cells were next pelleted by centrifugation at 4,000 \times g for 10 min, lysed by sonication, and then 100 μ g/ml spectinomycin was added to halt protein synthesis.

Splicing assays. Splicing assays were performed as previously described (13), with temperatures, times, and additions of ssDNA (M13mp18; New England Biolabs) as noted in the figures and their legends. Calculation of splicing in native exteins was done as previously described (13), except for comparison of *P. horikoshii* RadA to *P. horikoshii* RadA Δ NCR, for which the intein band was omitted from analysis due to lack of identification in the *P. horikoshii* RadA Δ NCR samples. For MIG-RadAi splicing assays, crude lysate was incubated for 1 h at 25°C following expression and lysis (above) and analyzed by in-gel fluorescence as described elsewhere (10). The extent of splicing was calculated as previously described (13). Samples in all cases were separated on 8-to-16% SDS-PAGE gels. In all splicing assays, densitometry using ImageJ (<http://imagej.nih.gov/ij/>) was used to calculate product levels. Data were plotted and curves were fit using Prism7 (GraphPad). Further information regarding our experimental procedures is provided in Text S1.

Sequence alignment. Clustal Omega was used to align the protein sequences of *P. horikoshii* and *T. sibiricus* RadA proteins. Specifically, codons 116 to 529 of *P. horikoshii* RadA (O58001.1) and codons 116 to 526 of *T. sibiricus* RadA (ACS91039.1) were aligned with sequences analyzed from UniProt.org.

Structural modeling. Ten *P. horikoshii* RadA wild-type precursor models reported in a previous study (9) were used as structural templates for generating 10 corresponding homology models for the *T. sibiricus* RadA wild-type precursors. These precursor models contained 482 amino acid residues that spanned the range from Ser45 to Asp526 in wild-type *T. sibiricus* RadA. The homology models were generated using the Modeller software (29) with the best model chosen using the DOPE (30) and GA341 (31) energy functions. Further minimization was performed identically for all 10 *T. sibiricus* and *P. horikoshii* models by using the program CHARMM, version c35b3 (32, 33) with the CHARMM36 force field for proteins (34). All protein nonhydrogen atoms were harmonically restrained with a force constant of 1 kcal/mol/Å², and the model was minimized using 5,000 steps-of-steepest-descent (SD) minimization with an energy change tolerance of 0.001 kcal/mol and SHAKE constraints (35) on all hydrogen atoms. Nonbonded interactions were calculated using 999.0-Å cutoff distances for van der Waals and electrostatic interactions, to ensure that all possible interactions were fully included. The restraint force constant was then reduced to 0.5 kcal/mol/Å², and the same minimization was repeated. Finally, the restraint was removed altogether, and an unrestrained minimization was performed to obtain each final optimized precursor model. The interaction energy between residue backbone or side chain atoms was calculated by designating the atoms types C, N, O, CA, HA, and HN as backbone atoms and all other atoms as side chain atoms.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.02304-17>.

TEXT S1, DOCX file, 0.03 MB.

FIG S1, TIF file, 0.7 MB.

FIG S2, TIF file, 0.2 MB.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants GM39422 and GM44844 to M.B. and F32GM121000 to C.W.L.

The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

We declare we have no conflicts of interest regarding any work presented in the manuscript.

We thank Rebecca McCarthy for assistance in preparation of the manuscript.

C.W.L. and M.B. conceived the study. C.W.L. and M.S. performed experiments. N.K.B. performed the computational modeling. C.W.L. and M.B. wrote the manuscript, with contributions from N.K.B. M.S. rendered the figures. M.B. supervised the project. All authors were involved in discussion and editing of the manuscript.

REFERENCES

- Mills KV, Johnson MA, Perler FB. 2014. Protein splicing: how inteins escape from precursor proteins. *J Biol Chem* 289:14498–14505. <https://doi.org/10.1074/jbc.R113.540310>.
- Paulus H. 2001. Inteins as enzymes. *Bioorg Chem* 29:119–129. <https://doi.org/10.1006/bioo.2001.1203>.
- Shah NH, Muir TW. 2014. Inteins: nature's gift to protein chemists. *Chem Sci* 5:446–461. <https://doi.org/10.1039/C3SC52951G>.
- Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, Belfort M. 2016. Intein clustering suggests functional importance in different domains of life. *Mol Biol Evol* 33:783–799. <https://doi.org/10.1093/molbev/msv271>.
- Kelley DS, Lennon CW, SEA-PHAGES, Belfort M, Novikova O. 2016. Mycobacteriophages as incubators for intein dissemination and evolution. *mBio* 7:e01537-16. <https://doi.org/10.1128/mBio.01537-16>.
- Naor A, Altman-Price N, Soucy SM, Green AG, Mitiagin Y, Turgeman-Grott I, Davidovich N, Gogarten JP, Gophna U. 2016. Impact of a homing intein on recombination frequency and organismal fitness. *Proc Natl Acad Sci U S A* 113:E4654–E4661. <https://doi.org/10.1073/pnas.1606416113>.
- Mills KV, Lew BM, Jiang S, Paulus H. 1998. Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc Natl Acad Sci U S A* 95:3543–3548. <https://doi.org/10.1073/pnas.95.7.3543>.
- Callahan BP, Topilina NI, Stanger MJ, Van Roey P, Belfort M. 2011. Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications. *Nat Struct Mol Biol* 18: 630–633. <https://doi.org/10.1038/nsmb.2041>.
- Topilina NI, Novikova O, Stanger M, Banavali NK, Belfort M. 2015. Post-translational environmental switch of RadA activity by extein-intein interactions in protein splicing. *Nucleic Acids Res* 43:6631–6648. <https://doi.org/10.1093/nar/gkv612>.
- Topilina NI, Green CM, Jayachandran P, Kelley DS, Stanger MJ, Piazza CL, Nayak S, Belfort M. 2015. SufB intein of Mycobacterium tuberculosis as a sensor for oxidative and nitrosative stresses. *Proc Natl Acad Sci U S A* 112:10348–10353. <https://doi.org/10.1073/pnas.1512777112>.
- Reitter JN, Cousin CE, Nicastrì MC, Jaramillo MV, Mills KV. 2016. Salt-dependent conditional protein splicing of an intein from Halobacterium salinarum. *Biochemistry* 55:1279–1282. <https://doi.org/10.1021/acs.biochem.6b00128>.
- Ciragan A, Aranko AS, Tascon I, Iwai H. 2016. Salt-inducible protein splicing in cis and trans by inteins from extremely halophilic archaea as a novel protein-engineering tool. *J Mol Biol* 428:4573–4588. <https://doi.org/10.1016/j.jmb.2016.10.006>.
- Lennon CW, Stanger M, Belfort M. 2016. Protein splicing of a recombinase intein induced by ssDNA and DNA damage. *Genes Dev* 30: 2663–2668. <https://doi.org/10.1101/gad.289280.116>.
- Wood DW, Camarero JA. 2014. Intein applications: from protein purifi-

- cation and labeling to metabolic control methods. *J Biol Chem* 289: 14512–14519. <https://doi.org/10.1074/jbc.R114.552653>.
15. Buskirk AR, Ong YC, Gartner ZJ, Liu DR. 2004. Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc Natl Acad Sci U S A* 101:10505–10510. <https://doi.org/10.1073/pnas.0402762101>.
 16. Skretas G, Wood DW. 2005. Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci* 14:523–532. <https://doi.org/10.1110/ps.04996905>.
 17. Seitz EM, Brockman JP, Sandler SJ, Clark AJ, Kowalczykowski SC. 1998. RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev* 12:1248–1253. <https://doi.org/10.1101/gad.129.1248>.
 18. González JM, Masuchi Y, Robb FT, Ammerman JW, Maeder DL, Yanagibayashi M, Tamaoka J, Kato C. 1998. *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* 2:123–130. <https://doi.org/10.1007/s007920050051>.
 19. Miroshnichenko ML, Hippe H, Stackebrandt E, Kostrikina NA, Chernyh NA, Jeanthon C, Nazina TN, Belyaev SS, Bonch-Osmolovskaya EA. 2001. Isolation and characterization of *Thermococcus sibiricus* sp. nov. from a Western Siberia high-temperature oil reservoir. *Extremophiles* 5:85–91. <https://doi.org/10.1007/s007920100175>.
 20. Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysis of intron/intein mobility. *Nucleic Acids Res* 29:3757–3774. <https://doi.org/10.1093/nar/29.18.3757>.
 21. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
 22. Lin Z, Kong H, Nei M, Ma H. 2006. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc Natl Acad Sci U S A* 103:10328–10333. <https://doi.org/10.1073/pnas.0604232103>.
 23. Oeemig JS, Zhou D, Kajander T, Wlodawer A, Iwaï H. 2012. NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J Mol Biol* 421:85–99. <https://doi.org/10.1016/j.jmb.2012.04.029>.
 24. Cronin M, Coolbaugh MJ, Nellis D, Zhu J, Wood DW, Nussinov R, Ma B. 2015. Dynamics differentiate between active and inactive inteins. *Eur J Med Chem* 91:51–62. <https://doi.org/10.1016/j.ejmech.2014.07.094>.
 25. Lennon CW, Belfort M. 2017. Inteins. *Curr Biol* 27:R204–R206. <https://doi.org/10.1016/j.cub.2017.01.016>.
 26. Belfort M. 2017. Mobile self-splicing introns and inteins as environmental sensors. *Curr Opin Microbiol* 38:51–58. <https://doi.org/10.1016/j.mib.2017.04.003>.
 27. Hiraga K, Soga I, Dansereau JT, Pereira B, Derbyshire V, Du Z, Wang C, Van Roey P, Belfort G, Belfort M. 2009. Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *J Mol Biol* 393:1106–1117. <https://doi.org/10.1016/j.jmb.2009.08.074>.
 28. Iwaï H, Mikula KM, Oeemig JS, Zhou D, Li M, Wlodawer A. 2017. Structural basis for the persistence of homing endonucleases in transcription factor IIB inteins. *J Mol Biol* 429:3942–3956. <https://doi.org/10.1016/j.jmb.2017.10.016>.
 29. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
 30. Shen MY, Sali A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524. <https://doi.org/10.1110/ps.062416606>.
 31. Melo F, Sánchez R, Sali A. 2002. Statistical potentials for fold assessment. *Protein Sci* 11:430–448. <https://doi.org/10.1002/pro.110430>.
 32. Brooks BR, Brooks CL, III, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. 2009. CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614. <https://doi.org/10.1002/jcc.21287>.
 33. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macro-molecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217. <https://doi.org/10.1002/jcc.540040211>.
 34. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J Chem Theory Comput* 8:3257–3273. <https://doi.org/10.1021/ct300400x>.
 35. Ryckaert JP, Ciccotti G, Berendsen HJC. 1977. Numerical Integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys* 23:327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).